


2011

An Empirical Investigation of Decision Aids to Improve Auditor Effectiveness in Analytical Review

Robert N. Marley

University of South Florida, rmarley@coba.usf.edu

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>

 Part of the [Accounting Commons](#), and the [American Studies Commons](#)

Scholar Commons Citation

Marley, Robert N., "An Empirical Investigation of Decision Aids to Improve Auditor Effectiveness in Analytical Review" (2011).
Graduate Theses and Dissertations.
<http://scholarcommons.usf.edu/etd/3232>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

An Empirical Investigation of Decision Aids to Improve
Auditor Effectiveness in Analytical Review

by

Robert N. Marley, II

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
School of Accountancy
College of Business
University of South Florida

Major Professor: Uday S. Murthy, Ph.D.
Stephanie Bryant, Ph.D.
Terry Engle, Ph.D.
Rosann Collins, Ph.D.

Date of Approval:
June 27, 2011

Keywords: Experimental, analytical procedures, cognitive load theory, working memory,
analytical review paradox

Copyright ©2011, Robert N. Marley, II

ACKNOWLEDGMENTS

I do not believe that a doctoral student could have a better dissertation committee than mine. I am greatly indebted to my dissertation chairman, Professor Uday S. Murthy, for his seemingly unending supply of time, wisdom, and guidance. The grace and academic prowess displayed by Dr. Murthy is the standard I seek to one day achieve. I am deeply grateful to Professor Stephanie Bryant for her rigorous and thorough feedback. Although I have no idea how Professor Bryant found the time to serve on my dissertation committee while being the director of USF's school of accountancy, I have learned tremendously from her during this journey and I greatly appreciate Dr. Bryant's willingness to share her accounting information systems knowledge with me. I cannot thank Professor Terry Engle enough for his tireless guidance with respect to helping me to incorporate the relevant audit standards into my dissertation, for guiding me to clearly articulate the practical implications of my work, and for forcing me to think critically about unquestionably accepting the methods used in prior research. Finally, I must express my deep gratitude to Professor Rosann Collins for volunteering to serve as my "non-accounting" dissertation committee member and for patiently taking the time to explain the subtleties of many of the cognitive theories used in the academic literature. Whenever I met with Dr. Collins she always guided me towards see accounting in a new light, giving me the gift of a new perspective. Among the many things I have learned from Dr. Collins, my favorite lesson is to never fight the Berkeley law librarians.

TABLE OF CONTENTS

LIST OF TABLES	iv
LIST OF FIGURES	vi
ABSTRACT.....	vii
CHAPTER ONE: INTRODUCTION.....	1
CHAPTER TWO: BACKGROUND AND LITERATURE REVIEW	11
Analytical Review Context.....	11
The Role of Analytical Review in Audit Planning	11
Using Analytical Review to Support Financial Statement Assertions.....	12
The Role of Analytical Review in the Overall Review Stage	13
The Value of Analytical Review.....	14
The Analytical Review Task Setting	16
Decision Aids.....	18
Decision Aids Used by Audit Practitioners in Analytical Review	19
Why Decision Aids Must be Introduced into New Contexts with Caution.....	21
Cognitive Load Theory	23
Schemata and Their Relationship to Cognitive Load	24
Extending Cognitive Load Theory to Analytical Review.....	25
The Three Types of Cognitive Load.....	26
The Hypothesis Generation Stage.....	29
Risks Associated With Client Explanations	30
Hypothesis Self-Generation	33
The Hypothesis Evaluation Stage	34
The Hypothesis Selection Stage.....	37
Hypothesis Development	39
The No-Aid Intervention.....	39
The Activity Relationship Diagram Intervention	40
Hypothesis Generation Stage Predictions	44
Hypothesis Evaluation Stage Predictions	45
Hypothesis Selection Stage Predictions.....	47
The Pattern-Consideration Aid Intervention.....	47
Hypothesis Generation Stage Predictions.....	49
Hypothesis Evaluation Stage Predictions	50
Hypothesis Selection Stage Predictions.....	52

Combined-Aid Intervention Predictions	52
Planned Mediation Analysis	57
CHAPTER THREE: METHOD	59
Experimental Design.....	59
Case Design Considerations	60
Task.....	61
Case Materials.....	62
Participants.....	63
Analytical Review Software Application	64
Experimental Procedure.....	66
The Analytical Review Task and Decision Aid Manipulations.....	68
Dependent Measures	74
Data Coding	75
Hypotheses Coding	76
CHAPTER FOUR: RESULTS	78
Analysis of Participant Demographic Information and Responses	78
Pre-Test Knowledge Assessment.....	79
Participants Reporting No Decision Aid Reliance.....	81
Participant Descriptive Statistics	82
Participant Note-Taking Descriptive Statistics	85
Decision-Aid Reliance and Usefulness Descriptive Statistics.....	88
Participant Proposed Hypotheses.....	90
Assumption Testing	90
Univariate Analysis of Variance (ANOVA).....	90
Logistic Regression (LOGIT).....	91
Hypotheses Testing.....	92
Planned Mediation Analysis	106
Measuring Cognitive Load	107
Where I Measured Cognitive Load During the Task.....	108
Determining Whether a Mediating Relationship Exists	109
Activity Relationship Diagram Mediation Results	110
Pattern-Consideration Aid Mediation Results	112
Combined-Aid Mediation Results	113
Summary of Mediation Analysis Results and Discussion of the Implications.....	113
Post Hoc Analysis	114
The ARD and the PCA as Memory Aids.....	115
The ARD as a Memory Aid.....	116
The PCA as a Memory Aid.....	116
Overconfidence	118
Effective Hypothesis Generation	121
Summary of Post Hoc Findings	123

CHAPTER FIVE: SUMMARY AND CONCLUSIONS	124
Summary of Key Findings	124
Contributions.....	126
Limitations	128
Future Research Opportunities Arising from this Study’s Results.....	129
Future Research Opportunities Extending this Study’s Decision Aids	131
REFERENCES	135
APPENDIX A: CASE MATERIALS.....	145
APPENDIX B: ONLINE PRE-TEST.....	151
APPENDIX C: TRAINING MATERIALS.....	155
APPENDIX D: ONLINE POST-EXPERIMENT QUESTIONNAIRE	166
APPENDIX E: ORAL SCRIPT DELIVERED TO PARTICIPANTS DURING THE MAIN TASK.....	176
APPENDIX F: WRITTEN INSTRUCTIONS GIVEN TO PARICIPANTS IN STAGE THREE.....	180

LIST OF TABLES

TABLE 2.1:	Hypotheses by Decision Aid Intervention and Stage	56
TABLE 4.1:	Descriptive Statistics on Participant Exclusion.....	83
TABLE 4.2:	Participant Descriptive Statistic	86
TABLE 4.3:	Ex-Post Note-Taking Analysis Among No-Aid Intervention Participants	88
TABLE 4.4:	Decision Aid Reliance and Usefulness Descriptive Statistics.....	89
TABLE 4.5:	Number of Plausible Hypotheses Proposed by Intervention.....	92
TABLE 4.6:	Hypothesis 1b: Hypothesis Generation Effectiveness (ARD vs. No-Aid).	94
TABLE 4.7:	Hypothesis 2: Hypotheses Evaluation and Effectiveness (ARD vs. No-Aid).....	95
TABLE 4.8:	Hypothesis 3: Hypothesis Selection Effectiveness (ARD vs. No- ARD).	96
TABLE 4.9:	Hypothesis 4b: Hypothesis Generation Effectiveness (PCA vs. No-Aid).	98
TABLE 4.10:	Hypothesis 5: Hypothesis Evaluation Effectiveness (PCA vs. No-Aid).	99
TABLE 4.11:	Hypothesis 6: Hypothesis Selection Effectiveness (PCA vs. No- Aid).....	100
TABLE 4.12:	Descriptive Statistics: Hypothesis Generation, Evaluation, and Selection	102
TABLE 4.13:	Hypothesis Evaluation Effectiveness (Combined-Aid vs. No-Aid).....	103
TABLE 4.14:	Hypothesis 9: Hypothesis Selection Effectiveness (Combined- Aid vs. No-Aid).....	104

TABLE 4.15: Summary of Hypotheses and Results	105
TABLE 4.16: The Four Steps and Required Outcomes Necessary to Provide Evidence that Cognitive Load Mediates Decision Aid Use and Analytical Review Effectiveness.....	110
TABLE 4.17: Post Hoc Overconfidence Descriptive Statistics	120
TABLE 4.18: Post Hoc Analysis: Hypothesis Generation Effectiveness	122

LIST OF FIGURES

FIGURE 1:	Illustration of All Stages Associated With Analytical Review	3
FIGURE 2a:	The activity relationship diagram. (Part 1 of 2)	42
FIGURE 2b:	The activity relationship diagram. (Part 2 of 2).	43
FIGURE 3:	Mediation Models.....	58
FIGURE4:	Research Design	59
FIGURE 5:	Screenshot of the Analytical Review Application’s Main Screen.	65
FIGURE 6:	Experimental Procedure	67
FIGURE 7:	Screenshot Example of Participant-Identified Cues Displayed by the PCA During Hypothesis Generation	70
FIGURE 8:	Screenshot Example of the Hypothesis Evaluation and Selection Screen	71
FIGURE 9:	Screenshot Example of the Hypothesis Evaluation Checklist.....	72
FIGURE 10:	Screenshot Example of the Hypothesis Selection Screen	73
FIGURE A1:	Case Materials: Financial Information Provided to All Participants	150
FIGURE C1:	Training Case Materials: Financial Information Presented to All Participants During Training	161

ABSTRACT

There is considerable evidence in the audit literature that even though auditors usually identify the relevant information needed to propose and select the correct cause of an unexpected fluctuation, they frequently do not propose the correct cause, and even when they do propose the correct cause, they often fail to select it. I suggest that working memory limitations might be a factor contributing to this analytical review paradox. Consequently, this study investigates whether two new decision aids, designed from Cognitive Load Theory, reduce auditors' cognitive load during analytical review, freeing cognitive resources for problem solving, and ultimately leading to improved auditor analytical review effectiveness. My first decision aid, an activity relationship diagram (ARD), gives the auditor a graphical depiction of common accounting relationships. My second decision aid, a pattern-consideration aid (PCA), automatically recalls and textually displays the auditor-identified relevant information cues. In an experimental setting, I find that auditors who rely on either decision aid significantly improve their analytical review effectiveness compared to auditors who conduct analytical review unaided. However, contrary to my predictions, auditors who rely on both decision aids do not outperform auditors who rely on only one decision aid. Although I find empirical evidence that cognitive load is negatively related to analytical review effectiveness, I do not find evidence that my decision aids reduce cognitive load.

1.0 INTRODUCTION

When performing financial statement audits, external auditors around the world employ analytical review because it is consistent with their desire to conduct the audit from a holistic, risk-based approach (Trompeter and Wright 2010) and because conducting analytical review is presumptively mandatory under audit standards (AICPA 2010a; IAASB 2010a).¹ Analytical review is used during audit planning “...to assist in planning the nature, timing, and extent of auditing procedures that will be used to obtain audit evidence for specific account balances or classes of transactions” (AICPA 2010a, paragraph 6), during fieldwork to evaluate and support financial statement assertions, and during overall review to evaluate the reasonableness of the financial statements (AICPA 2010a, paragraph 9 and paragraph 23, respectively). Thus, auditors use analytical review as an attention-directing device to identify and assess risk and to obtain audit evidence. This study examines auditors’ analytical review effectiveness during audit planning and in audit fieldwork to evaluate financial statement assertions.

Analytical review is the “...diagnostic process of identifying, investigating, and resolving unexpected fluctuations.” (Koonce 1993, p. 57). An unexpected fluctuation arises when there is a significant difference between a client’s reported balance and the

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.

auditor's expectation for that balance.² The auditor formulates the expectation based on many factors. For example, the expectation should be based on the auditor's understanding of the client and the industry in which the client operates (AICPA 2010a). To identify unexpected fluctuations, the auditor applies *analytical procedures* which "...involve comparisons of recorded amounts, or ratios developed from recorded amounts, to expectations developed by the auditor." (AICPA 2010a, paragraph 5). Although the terms *analytical review* and *analytical procedures* are often used interchangeably in the audit literature, analytical procedures refer only to the specific tests an auditor performs while conducting analytical review, such as the calculation of a ratio (Koonce 1993). Thus, analytical procedures are used during the analytical review process. Although audit standards use the term *analytical procedures*, the spirit behind the standards' use clearly indicates auditors should conduct *analytical review* to achieve the audit objectives.³

After an unexpected fluctuation has been identified the auditor engages in three stages to investigate and resolve the unexpected fluctuation. First, the auditor generates hypotheses that can potentially explain why the fluctuation occurred. The process of obtaining hypotheses fuels an information search, though the hypotheses obtained often set the boundaries of the search (Asare and Wright 2001). Next, the auditor evaluates the

² The auditor's expectation is not necessarily expressed solely as a point-estimate, as auditors frequently establish thresholds, or upper and lower boundaries of tolerable expectation deviation. When a client-reported balance exceeds the threshold an unexpected fluctuation exists.

³ For example, ISA 520 defines analytical procedures as "...evaluations of financial information through analysis of plausible relationships among both financial and non-financial data. Analytical procedures also encompass such investigation as is necessary of identified fluctuations or relationships that are inconsistent with other relevant information or that differ from expected values by a significant amount." (IAASB 2010a, paragraph 4). The ASB is in the process of revising AU 329 such that the current redraft defines analytical procedures in exactly the same terms as ISA 520. If adopted, the ASB's proposal will be effective for audits ending on or after December 15, 2012.

merits of each proposed hypothesis against the relevant information. After evaluating each hypothesis, the auditor may decide none correctly explain the cause of the unexpected fluctuation; if so, he or she returns to the hypothesis generation stage. Alternatively, the auditor may select a hypothesis as the cause of the unexpected fluctuation, ending the analytical review process.⁴ The analytical review literature classifies these three stages as hypothesis generation, hypothesis evaluation, and hypothesis selection, respectively. These three stages are the focus of this study; Figure 1 presents an illustration of all the stages associated with conducting analytical review and denotes the stages investigated within this study.

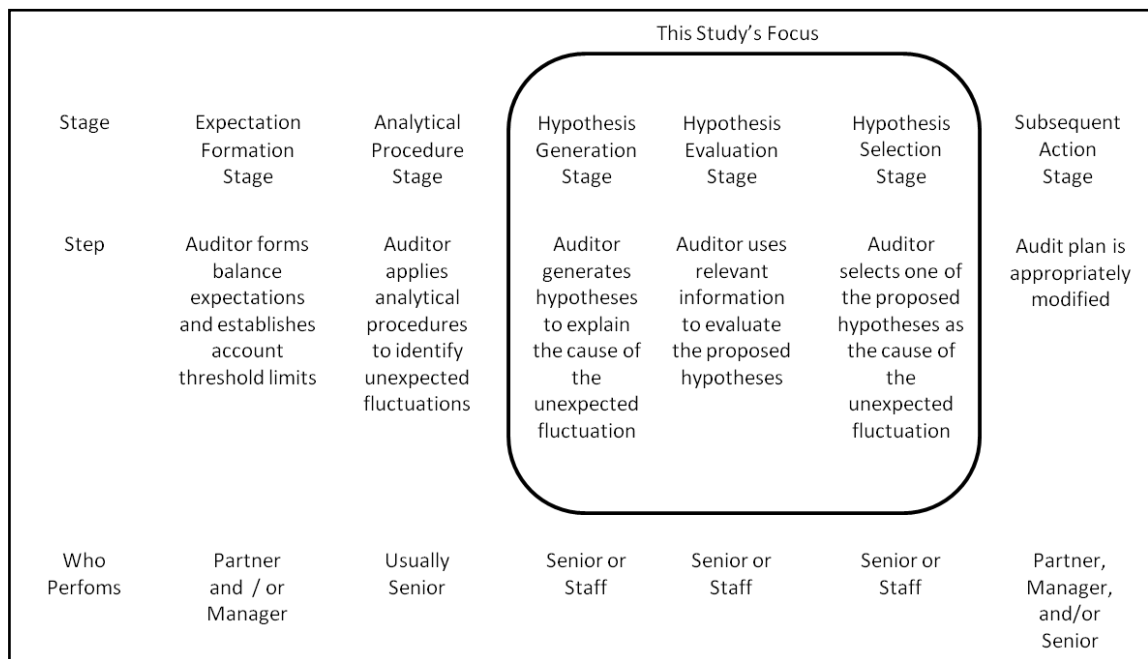


FIGURE 1. Illustration of All Stages Associated With Analytical Review.

⁴ In practice, an unexpected fluctuation can be comprised of one or many causes. Thus, to fully explain an unexpected fluctuation the auditor needs to obtain a hypothesis that identifies all causes. Within this study, the unexpected fluctuation I employ arises from only one cause. While consistent with prior analytical review research, the use of only one correct cause is an abstraction from practice in situations where an unexpected fluctuation is comprised of multiple causes. This abstraction is employed for interpretation tractability. A detailed discussion of this abstraction is presented in the method section.

I investigate these three stages because there is considerable evidence in the literature that it is difficult for auditors to propose, evaluate, and select the correct cause of an unexpected fluctuation, giving rise to an analytical review paradox: Even though auditors usually identify the relevant information needed to propose and select the correct cause of an unexpected fluctuation, they frequently do not propose the correct cause, and even when they do propose the correct cause, they often fail to select it (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).⁵ A theoretical explanation for this problem is that the cognitive requirements of the task exceed the auditor's available cognitive resources. Specifically, I suggest that although auditors can identify the pieces of information required to propose the correct cause of the unexpected fluctuation, the cognitive load placed upon the auditor's working memory can exceed its capacity.⁶ As a result, I suggest working memory limitations frequently cause the auditor to conduct ineffective analytical review.

This study investigates the analytical review effectiveness of inexperienced auditors during hypothesis generation, hypothesis evaluation, and hypothesis selection. Although standards allow inexperienced auditors to perform analytical review, the existing literature focuses almost exclusively on experienced auditors even though it seems likely that inexperienced auditors should have greater difficulty performing analytical review. Thus, it seems reasonable to expect that inexperienced auditors may

⁵ The research cited here examined the analytical review effectiveness of experienced auditors that were asked to perform an analytical review task in an experimental setting.

⁶ Cognitive load is defined as the burden placed upon working memory when conducting a task (Sweller 1988). Working memory is the amount of memory used by an individual to retain and process current information (Baddeley 1992).

perform analytical review less effectively than experienced auditors. Given that staff auditors now conduct analytical review 48 percent of the time (Trompeter and Wright 2010), there is a need for research that examines the analytical review effectiveness of inexperienced auditors to complement and extend prior research that has examined the analytical review effectiveness of experienced auditors.⁷

I apply Cognitive Load Theory (Sweller 1988) to examine whether inexperienced auditors' judgment and decision making performance can be improved by reducing the cognitive load placed upon them during analytical review. Using Cognitive Load Theory as the foundation for the creation of two new decision aids, I examine whether the use of these aids can help to reduce auditors' cognitive load during the analytical review, leading to improved analytical review effectiveness. The first decision aid that I designed and developed is an activity relationship diagram (ARD), gives the auditor a graphical depiction of common accounting relationships. This aid should help an auditor recall common accounting relationships during the analytical review task, reducing the auditor's cognitive load by eliminating the need to expend cognitive resources recalling and maintaining common accounting relationships in working memory. By graphically presenting common accounting relationships to auditors, the aid should heighten the salience of these relationships, which should help the auditor to generate, evaluate, and select the correct cause of the unexpected fluctuation. The second decision aid that I designed and developed is a pattern-consideration aid (PCA), automatically recalls and textually displays the auditor-identified relevant information cues. The PCA should help

⁷ Since this study is concerned with investigating the analytical review effectiveness of inexperienced auditors (i.e., staff auditors) tasked with performing analytical review, hereafter the term "auditor" in this study refers to inexperienced auditors and not those with relatively high levels of audit experience such as senior and manager level auditors.

an auditor to accurately recall and simultaneously consider the relevant information cues during the analytical review task, reducing the auditor's cognitive load by eliminating the need to expend cognitive resources recalling and maintaining the relevant cues in working memory. By textually presenting the auditor-identified relevant cues, the PCA should heighten the salience of the cues, helping the auditor to consider the cues while conducting hypothesis generation, ultimately helping the auditor to generate the correct hypothesis. The PCA should also improve hypothesis evaluation and hypothesis selection by permitting the auditor to identify and select the hypothesis that is best supported by the auditor-identified relevant cues.

This study is important because an auditor's failure to correctly generate, evaluate, and select the actual cause of an unexpected fluctuation can negatively impact both audit effectiveness and audit efficiency. For example, failing to correctly attribute an unexpected increase in gross margin to accounting error can lead to misstated financial statements. Audit failure can result if the auditor fails to identify misstatements that individually, or in the aggregate, materially misrepresent a company's financial position. At a minimum, conducting analytical review ineffectively has audit efficiency consequences. Specifically, if an auditor fails to select the correct cause of an unexpected fluctuation during analytical review, subsequent fieldwork by way of audit procedures may be unproductively channeled to relatively low risk audit areas, resulting in needlessly high audit costs. Although the cause of the unexpected fluctuation may ultimately be uncovered during fieldwork, such detection would be accompanied by high audit costs, which could have been avoided if the initial analytical review had been effective. Thus, an auditor may be required to perform analytical review a second time to

explain the same unexpected fluctuation. Consequently, there are negative consequences associated with ineffective analytical review. Within this study, I use the term *analytical review effectiveness* to refer to auditor performance in three stages: hypothesis generation, hypothesis evaluation, and hypothesis selection.

Two recent trends suggest that accounting firms remain at risk for conducting ineffective analytical review: First, analytical review constitutes an increasingly large part of the audit engagement budget. According to Trompeter and Wright (2010), analytical review constitutes approximately 25 percent of an audit engagement's budgeted hours (as of mid-2005) compared to 21 percent of the engagement hours during the period 1998 to 2003. This trend is consistent with the findings of Ameen and Strawser (1994), who present evidence that approximately 15 percent of an audit engagement's budget was allocated to analytical review during the early 1990's. Thus, recent accounting scandals and legislation notwithstanding, analytical review continues to play an increasingly larger role in financial statement audits. Second, audit firms are progressively employing staff auditors to conduct analytical review; staff auditors now conduct analytical review 48 percent of the time (Trompeter and Wright 2010) compared to ten percent of the time as reported by Hirst and Koonce's (1996).⁸ Thus, analytical review is increasingly being performed by auditors with less audit experience than in the past.

Accounting firms have expended significant resources developing *audit aids* to improve auditor judgment and decision making (O'Donnell and Perkins 2011). "Industry

⁸ Trompeter and Wright (2010) note that although less experienced staff (i.e., staff auditors) now frequently *conduct* analytical review, experienced auditors (i.e. audit seniors and above) generally *design* the analytical review tests. Thus, inexperienced auditors are often the ones who propose, evaluate, and select the cause of the unexpected fluctuation, which are the three stages of analytical review this study examines.

packs” are an example of an audit aid used by each of the Big 4 public accounting firms to improve auditor judgment and decision making. An industry pack automatically tailors each client’s audit plan to directly address industry-specific risks. Industry packs also prompt the auditor to consider management’s integrity when formulating the audit plan (Dowling and Leech 2007). However, accounting research finds that *decision aids* are “rarely used” during analytical review (Trompeter and Wright 2010, p. 690).⁹ It is therefore not surprising that relatively few accounting studies have investigated decision aids within the context of analytical review (O’Donnell and Perkins 2011). A potential explanation for the absence of decision aids during analytical review may be that experienced decision makers tend to ignore or under-rely on decision aids because they do not want to surrender their own judgment to the aid (Rose 2002). Since analytical review was largely conducted by experienced auditors in the past, decision aids may have been considered unwanted or unnecessary. Given that audit firms are increasingly using inexperienced auditors to conduct analytical review, and since it is reasonable to expect that inexperienced auditors are just as prone to the analytical review paradox as experienced auditors, this study is important because it tests the effectiveness of two decision aids that can be used to help inexperienced auditors perform analytical review. The use of decision aids within this context should be especially helpful because research finds inexperienced users to be more likely to place reliance upon decision aids compared to experienced users (Rose 2002).

⁹ Additional indirect evidence is provided by Dowling and Leech (2007), who examine the audit support systems and decision aids of the Big Four and find that while the audit systems of some firms have rudimentary analytical review aids such as ratio calculators, none of the Big Four firms used decision aids to perform the analytical review steps examined in this study. As the audit practices of the Big Four are highly standardized, the Dowling and Leech (2007) study provides evidence that the decision aids utilized in this study are not already widely used in practice.

Prior audit decision aid research demonstrates that while decision aids can improve task performance (Blocher et al. 1983; McDaniel and Kinney 1995; Mueller and Anderson 2002), decision aids can also introduce new biases (Pincus 1989; Glover et al. 1997; Kowalczyk and Wolfe 1998). Thus, when introducing a decision aid into a new context it is important to investigate what impact it has upon existing judgment and decision making processes. For these reasons, the findings of prior decision aid research cannot automatically be extended to the setting in this study, i.e., analytical review being conducted by inexperienced auditors.

In a between-subjects experiment, I manipulate one factor, decision aids, at four levels. The four levels of the decision aid factor are: (1) a no-aid intervention, (2) an activity relationship diagram intervention, (3) a pattern-consideration aid intervention, and (4) a combined-aid intervention, where participants are provided with both an activity relationship diagram and a pattern-consideration aid.¹⁰ Participants are masters of accountancy (MAcc) students and accounting seniors who serve as proxies for inexperienced auditors. Applying Cognitive Load Theory, I predict auditors' analytical review effectiveness can be improved by reducing the cognitive load placed upon them. Thus, I examine what impact the two decision aids utilized in this study have on the cognitive load experienced by auditors during analytical review and I examine the impact that reducing cognitive load has on analytical review effectiveness. I suggest that cognitive load mediates the relationship between decision aid use and analytical review effectiveness and I conduct mediation analysis to test this assertion.

¹⁰I do not suggest that my two decision aids are the normative standard. Instead, I examine whether the two decision aids utilized within this study are a step towards improving auditors judgment and decision making during analytical review. Therefore, unlike proof-of-concept research, this study does not attempt to demonstrate the feasibility of a new method or principle.

This study responds to Bonner's (1999) call for research to mitigate known judgment and decision making deficiencies. Bonner categorizes judgment and decision making deficiencies in terms of task, person, and environmental factors. Thus, I address her call by introducing two new decision aids that are designed from theory to help attenuate known auditor (i.e., person) judgment and decision making deficiencies.

The remainder of this dissertation is organized as follows. In the Background and Literature Review section, I present a literature review. In the Hypotheses Development section, I discuss this study's decision aids in depth and develop and propose my hypotheses. In the Method section, I discuss the experimental design, the task, and the participants employed. In the Results section, I discuss the statistical tests associated with testing each hypothesis and present the results of hypothesis testing. Finally, the Summary and Conclusion section discusses the implications of my findings, highlights the study's contributions, recognizes the study's limitations, and presents future research opportunities.

2.0 BACKGROUND AND LITERATURE REVIEW

2.1 Analytical Review Context

AU 329, *Analytical Procedures*, describes the three ways analytical review may be used during the audit: to help plan the audit, as a substantive test in support of specific financial statement assertions, and as a form of reasonableness testing during the final review stage of the engagement (AICPA 2010a).¹¹ Although there are numerous analytical review techniques available (i.e. use of accounting ratios, trend analysis, etc) the universal rationale for using any technique is to provide assurance that financial statement assertions are not materially misstated.

2.1.1 The Role of Analytical Review in Audit Planning.

The traditional role of analytical review within the United States has been to help formulate the audit plan (Hirst and Koonce 1996). Per AU 329, “The purpose of applying analytical procedures in planning the audit is to assist in planning the nature, timing, and extent of auditing procedures that will be used to obtain audit evidence for specific account balances or classes of transactions.” (AICPA 2010a, paragraph 6). In other words, the auditor should conduct planning stage analytical review to identify areas of the financial statements with the highest potential for misstatement by identifying those

¹¹ This promulgation was adopted by the PCAOB as part of the body of interim auditing standards. However, the PCAOB has recently made some changes to the interim standard. For example, PCAOB Audit Standard 14, *Evaluating Audit Results*, now establishes the requirements regarding the performance of analytical procedures as part of the overall review stage of the audit.

accounts that vary significantly from auditor-developed expectations. Planning stage analytical review is typically used to calibrate the scope and intensity of substantive testing planned for a specific financial statement area; substantive testing tends to be heavier in areas where the risk of misstatement is deemed to be higher (Trompeter and Wright 2010). Consequently, the use of analytical review during audit planning is not only consistent with risk-assessment audit standards, but also amenable to practitioners' desire to formulate a risk-based, efficient audit plan. In the past, planning stage analytical review was conducted by only experienced auditors (Hirst and Koonce 1996; Abdolmohammadi 1999). However, a recent study by Trompeter and Wright (2010) finds that public accounting firms are increasingly using less experienced auditors to perform, though not design, analytical review. For example, Trompeter and Wright (2010) find that while seniors and managers design the analytical procedures to be used 80 percent of the time, staff auditors perform analytical review approximately 48 percent of the time.

2.1.2 Using Analytical Review to Support Financial Statement Assertions.

Although AU 329 does not require auditors to use analytical review as a substantive test to support financial statement assertions, it legitimizes the use of analytical review for this purpose, subject to four considerations. Per AU 329, "The expected effectiveness and efficiency of an analytical procedure in identifying potential misstatements depends on, among other things, (a) the nature of the assertion, (b) the plausibility and predictability of the relationship, (c) the availability and reliability of the data used to develop the expectation, and (d) the precision of the expectation" (AICPA 2010a, paragraph 11). The goal of using analytical review to support financial statement

assertions is to determine the fairness of one or more financial statement account balances. In contrast to audit planning, where the auditor typically uses highly aggregated financial information, analytical review used to support financial statement assertions is often applied using disaggregated data (Hirst and Koonce 1996). Thus, auditors using analytical review to test assertions may use data from a division or product line as opposed to the firm level (Hirst and Koonce 1996).

International Standard on Auditing 520, *Analytical Procedures* (ISA 520), legitimizes the use of analytical review as a substantive test to support financial statement assertions (IAASB 2010a). Similar to US audit standards, international auditing standards require the auditor to consider several factors before using analytical review as a substantive test to support financial statement assertions. Specifically, the auditor should determine the suitability of the analytical procedure employed for testing the assertion, evaluate the reliability of the data from which the auditor's expectation is developed, develop an expectation sufficiently precise to identify misstatements, and determine the amount of deviation from auditor expectations that is acceptable without further investigation (IAASB 2010a).

2.1.3 The Role of Analytical Review in the Overall Review Stage.

Performing analytical review during the overall review stage of the audit is presumptively mandatory under both US and international audit standards. AU 329 states that the auditor should employ analytical review during the overall review stage to "...assist the auditor in assessing the conclusions reached and in the evaluation of the overall financial statement presentation." (AICPA 2010a, paragraph 23). Similarly, ISA 520 states "The auditor shall design and perform analytical procedures near the end of the

audit that assist the auditor when forming an overall conclusion as to whether the financial statements are consistent with the auditor's understanding of the entity.” (IAASB 2010a, paragraph 6). Thus, standard setting bodies around the world recognize that conducting analytical review during the final stage of the audit serves as an overall reasonableness test to evaluate the financial statements as a whole. Consequently, the standards imply that analytical review detects unusual or inconsistent patterns that may go unnoticed during fieldwork.¹²

2.2 The Value of Analytical Review

When analytical review is applied correctly, empirical evidence demonstrates it can be effective in detecting misstatements. For example, Hylas and Ashton (1982) present evidence that when financial statement errors are identified during an engagement, analytical review was the tool that identified them 40 percent of the time. Loebbecke and Steinbart (1987) find that analytical review can be effective in detecting errors and Kinney (1987) applies a model to demonstrate how analytical review can effectively identify material misstatements when reasonably disaggregated data (e.g., by store, department, product line, etc) are used.

Within the practitioner literature, Joseph T. Wells, a well-known fraud examiner, discusses how the use of analytical review would have enabled ZZZZ Best's auditor to easily identify client fraud risks (Wells 2001). Further, Wells suggests that analytical review is an effective tool for identifying fraud because client personnel are generally unable to manipulate all the information necessary to produce normal or expected account relationships or account balances (Wells 2007). In summary, regardless of

¹² Although audit standards permit staff auditors to conduct analytical review during overall review stage, this study does not examine this use of analytical review because an informal conversation with two Big Four managers suggests that staff auditors do not perform analytical review during the overall review stage.

whether a material financial misstatement is caused by error or fraud, the audit standards, academic literature, and practitioner literature all suggest analytical review is a useful tool for identifying material misstatements.

Recognizing the value of analytical review procedures, standard setting bodies around the world have promulgated standards prescribing the use of analytical review in a variety of risk assessment contexts. For example, within the United States both the PCAOB and the ASB have adopted a set of risk assessment standards that require auditors to gain an in-depth understanding of the client, client operating environment, and client internal control. The purpose of requiring this in-depth understanding is to help the auditor assess the risk of material misstatement.¹³ AU 314, *Understanding the Entity and Its Environment and Assessing the Risks of Material Misstatement*, contains a presumptively mandatory requirement that the auditor use analytical review to help develop this understanding (AICPA 2010b). Similarly, PCAOB Auditing Standard No. 12, *Identifying and Assessing Risks of Material Misstatement*, specifies that the auditor should apply analytical review designed to enhance the auditor's understanding of the client's business, to identify significant transactions occurring since year-end, and to test revenue accounts to identify unusual or unexpected relationships (PCAOB 2010). In a fraud risk-assessment context, AU 316, *Consideration of Fraud in a Financial Statement Audit*, suggests that the auditor should consider results of analytical review procedures performed during audit planning as part of the information consulted to identify risks of material misstatement due to fraud (AICPA 2010c). Finally, International Standards on

¹³Statements on Auditing Standards No. 104 – 111 are the ASB's risk assessment standards (AICPA 2007). Similarly, Auditing Standards 8-15 are the PCAOB's risk assessment standards (PCAOB 2010). Both ASB and PCAOB risk assessment standards are effective now. The SEC recently approved the PCAOB's risk assessment standards on December 23, 2010 (SEC 2010).

Auditing also prescribe the use of analytical review to help identify and assess risk. For example, International Standard on Auditing 315, *Identifying and Assessing the Risk of Material Misstatement Through Understanding the Entity and Its Environment* (ISA 315), makes it presumptively mandatory for the auditor to apply analytical review to identify and assess risks of material misstatement at the financial statement and assertion level (IAASB 2010b). In sum, the fact that global standard setting bodies continue to expand the recommended applications of analytical review to achieve a wide variety of audit objectives highlights the importance of analytical review as a valuable audit tool.

2.3 The Analytical Review Task Setting

As opposed to contexts where auditors typically work in groups, such as fraud-risk brainstorming (Lynch, Murthy, and Engle 2009), analytical review is generally conducted individually.¹⁴ Thus, an auditor must utilize his or her accounting-domain and client-specific knowledge for effective performance. For example, an auditor conducting analytical review must understand how each financial statement account relates to another (accounting domain knowledge), understand the client's business operations (client-specific knowledge), then synthesize his or her domain- and context-specific knowledge to make inferences about how client activities should affect financial reporting.

Once an unexpected fluctuation is identified, the audit literature categorizes analytical review into three distinct, though inter-related, and sometimes iterative stages: hypothesis generation (and information search), hypothesis evaluation, and hypothesis selection (Koonce 1993; Asare and Wright 1997; Asare and Wright 2001). These stages

¹⁴This does not imply that analytical review is conducted by an individual in a vacuum. Since auditors generally conduct an audit as part of a team, discussions between auditors can and do occur. However, audit standards do not require analytical review to be conducted by more than one individual.

end when the auditor identifies, evaluates, and selects a hypothesis as the cause of the unexpected fluctuation. In this study, I investigate the analytical review effectiveness of inexperienced auditors during these three stages. Although Koonce (1993) and Asare and Wright (2001) classify the subsequent action taken by an auditor after hypothesis selection as part of the hypothesis selection stage, I suggest that since inexperienced auditors are increasingly performing analytical review it is more appropriate to recognize the subsequent action taken by an auditor in a new “subsequent action stage” because inexperienced auditors are unlikely to be the ones to decide what subsequent action to take (Abdolmohammadi 1999; Trompeter and Wright 2010). Figure 1 illustrates the stages of analytical review and denotes the three stages examined within this study. Since analytical review can be cognitively demanding, this study explores the use of decision aids to mitigate the effects of excessive cognitive load. Applying Cognitive Load Theory, I suggest the use of decision aids during analytical review should reduce the cognitive load placed upon the auditor, leading to more effective task performance. Thus, I expect cognitive load to mediate the relationship between decision aid use and task effectiveness.

To explain why I predict the use of decision aids will lead to more effective analytical review, I next provide an examination of the decision aid literature and Cognitive Load Theory. Then, I provide an in-depth examination of the hypothesis generation stage, the hypothesis evaluation stage, and hypothesis selection stage, identifying known auditor judgment and decision making deficiencies and apply Cognitive Load Theory to explain these deficiencies. Finally, I discuss how the

deficiencies I identify may be mitigated by using decision aids that are designed to reduce cognitive load.

2.4 Decision Aids

Decision aids are tools designed to improve decision quality by addressing specific aspects of the judgment and decision making process (Rose 2002). Decision aids are perceived to have positive effects on the quality of an individual's judgment and decision making. Research in both accounting and psychology examines a wide variety of decision aid designs, ranging from simple instructions such as "be sure to consider inherent client risk when planning the audit" to complex expert systems created to make the decision quality of a novice comparable to that of an expert (Bonner 2008). Decision aids have also been investigated in a myriad of contexts. A considerable body of research supports the notion that decision aids outperform unaided humans; decision aids can improve decision making and decision quality (Kleinmuntz 1990; Benbasat and Nault 1990; Eining, Jones, and Loebbecke 1997), overcome cognitive constraints (Butler 1985), and improve judgment consistency (Ashton 1992). Further, decision-aid users have been found to process information more consistently than non-aid users (Peterson and Pitz 1986; Ashton 1992).

Within the audit literature, decision aids are perceived to have positive effects on the quality of an individual's cognitive processing, memory retrieval, information search, problem representation, hypothesis generation, and evidence evaluation (Bonner 2008). Checklists and brief instructions are two of the most popular decision aids examined. Checklist research finds that providing lists to auditors results in more planned testing (Blocher et al. 1983) and leads to the collection of more information (Pincus 1989).

Checklist length inversely affects auditor probability assessments and explanations not included in a checklist are deemed less likely to occur than included explanations (Johnson and Kaplan 1996). Explicitly instructing auditors to eliminate hypotheses results in the elimination of fewer hypotheses compared to situations where these instructions are not provided (Muller and Anderson 2002). McDaniel and Kinney (1995) find that asking auditors to form their expectations before performing analytical review results in superior performance relative to auditors not asked to do so.

In practice, decision aids are used by accounting firms to improve auditors' judgment and decision making. Each of the Big Four public accounting firms utilize some sort of decision aid (Abdolmohammadi and Usoff 2001; Dowling and Leech 2007). Accounting firms use decision aids to provide structure to the audit process, to help auditors make better decisions, and to reduce the variability of decisions made by auditors across different accounting firm offices (Abdolmohammadi and Usoff 2001; Dowling and Leech 2007). For example, each of the Big Four use a decision aid application that tailors audit programs to achieve country-specific audit objectives, providing structure to the audit process. Another example of an audit structuring aid is the use of audit support software to embed electronic files containing suggested substantive tests, accessible to the auditor during key phases of the audit (Dowling and Leech 2007). Thus, audit firms use decision aids to improve audit effectiveness and audit efficiency.

2.4.1 Decision Aids Used by Audit Practitioners in Analytical Review.

Although accounting firms use decision aids in numerous audit tasks, Trompeter and Wright (2010) find analytical review to be an area where decision aids are “rarely

used” (p. 690). This finding by Trompeter and Wright in 2010 is consistent with results presented 15 years earlier by Hirst and Koonce (1996), who found 85 percent of auditors “never use” decision aids during any stage of analytical review. Dowling and Leech (2007) provide direct evidence that decision aids are not frequently used by auditors during analytical review and provide indirect evidence that decision aids like the ones investigated in this study are not widely used. Dowling and Leech (2007) conduct semi-structured interviews with audit partners and audit managers from the Big Four and one other international public accounting firm to obtain an understanding of the audit support systems and decision aids used in public accounting firms. They find that among these five public accounting firms, 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). Although specifics regarding the functionality of the tools audit firms use to extract and analyze data during analytical review are not presented by Dowling and Leech, 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 they heighten the salience of information needed to effectively perform analytical review, which should reduce the cognitive load placed upon an auditor during the task. Consequently, only one of the accounting firms examined by Dowling and Leech potentially uses a decision aid comparable to the ones examined in this study. Thus, the findings of Trompeter and Wright (2010) provide direct empirical evidence that decision aids are not widely used during analytical review, while Dowling and Leech (2007) provide indirect evidence that the decision aids utilized in this study are not already widely used.

Although decision aids are not widely used by practice during analytical review, empirical evidence suggests that auditors believe decision aids could be useful within this context. In a study that employed an instrument containing over 300 audit tasks across six different audit phases, Abdolmohammadi and Usoff (2001) asked a group of 90 audit partners and managers to evaluate each task's suitability to decision aid development and use. Participants could also indicate that a task was best performed by human processing alone. Broadly speaking, auditors indicated that some type of decision aid (either completely automating the task or using a decision support system) was preferable to unaided human processing during analytical review tasks; 57 percent of respondents felt decision aids would be useful during the planning stage, 46 percent felt aids would be useful to corroborate and support financial statement assertions, and 46 percent believed decision aids would be useful during the final review stage. Further, although not specific to the analytical review context, Dowling and Leech (2007) provide evidence that partners at Big Four accounting firms believe decision aids are useful because they improve audit efficiency, help to control junior staff (i.e. provide structure to the audit process), and promote compliance with accounting standards and the firm's audit methodology. Therefore, drawing upon the audit literature, I suggest that public accounting firms are receptive to using decision aids during analytical review even though the use of aids is not yet widespread.

2.4.2 Why Decision Aids Must be Introduced into New Contexts with Caution.

Although academic research finds many benefits to employing decision aids there are some potential drawbacks associated with decision aid use. For example, decision aids can prolong the decision making process (Mackay et al. 1992). Additionally,

decision aids can actually decrease decision quality when users place inappropriate reliance on them (Glover et al. 1997; Kowalczyk and Wolfe 1998). Further, research demonstrates that introducing a decision aid into an untested setting can have unexpected consequences, especially when the new context requires domain-specific knowledge. For example, although the decision aid literature suggests checklists help to improve judgment and decision making, Pincus (1989) finds auditors using a fraud risk checklist assess fraud risk as lower than non-checklist users when fraud is actually present because checklist users fixate on the fraud risks provided and do not fully consider fraud risks that are present, but not identified in the checklist. Ashton (1990) finds that introducing a decision aid into a setting where competitive monetary incentives are present can reduce users' effectiveness because individuals attempt to outperform the decision aid to obtain higher compensation. Ashton's finding is especially counterintuitive because he asks inexperienced participants to conduct a bond-rating task requiring specialized knowledge the participants do not have. The fact Ashton's participants try to outperform the decision aid without possessing the required specialized knowledge serves as a cautionary tale illustrating why decision aids must be introduced into new contexts with care. Thus, it is not necessarily appropriate to presume the findings of one context will seamlessly generalize to another.

Accordingly, this study contributes to the accounting decision aid literature by examining whether auditor effectiveness during hypothesis generation, hypothesis evaluation, and hypothesis selection can be improved by providing auditors with decision aids. I next present an overview of Cognitive Load Theory, the theory I apply to explain why reducing cognitive load should lead to improved task effectiveness.

2.5 Cognitive Load Theory

Originally proposed by Sweller (1988), Cognitive Load Theory maintains that learning will not be successful if the decision maker faces heavy cognitive load because an individual's working memory (i.e., short-term memory) is limited. Sweller defines cognitive load as the burden placed upon working memory in the conduct of a task. Working 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' working memory is not unlimited, the size of its limits is widely debated. Sweller suggests that working memory can contain seven pieces of information at any one time (plus or minus two pieces) while Miller (1956) suggests working memory may be even more constrained.¹⁵ Regardless of working memory's exact size, there is consensus that working memory is substantially limited (Simon 1974; Penney 1989; Baddeley 1992).

In many respects, the manner in which cognitive load and working memory function are similar to that of a personal computer; working memory is analogous to the amount of random-access memory installed on the computer and cognitive load is analogous to the amount of random-access memory in use at a given time. Just as a computer cannot engage in any incremental processing once its random access memory is full, an individual can only process as much information as he or she has working memory space. Consequently, once an individual's working memory is fully consumed no incremental learning occurs.

¹⁵ Fittingly, research within analytical review suggests that the number of hypotheses explored at any one point in time is usually between four and five with an upper bound of six or seven (Libby 1985; Asare and Wright 1997).

2.5.1 Schemata and Their Relationship to Cognitive Load.

Schemata (also referred to in the literature as mental models and chunks) represent the structure and organization of knowledge within long-term memory (Rose and Wolfe 2000). Schemata are described as “a cognitive construct that permits problem-solvers to recognize a problem as belonging to a specific category requiring particular moves for solution” (Tarmizi and Sweller 1988, p.424). An individual’s knowledge of subject matter is organized into schemata and this organization helps determine how to deal with new information (Sweller 1994). Schemata may contain multiple information components, permitting individuals to hold and organize vast amounts of information (Horcher and Tejay 2009). Individuals with knowledge and experience related to a task are believed to perform the task with greater efficiency and effectiveness because they develop schemata permitting them to retrieve and process relevant information better than novices (Rose and Wolfe 2000). Although a single schemata may contain a vast amount of information each schemata is processed by working memory as a single working memory item (Kirschner 2002). Thus, an individual’s information processing is greatly enhanced when the individual possesses the relevant schemata because a vast amount of information can be maintained and processed as a single working memory unit instead of a separate working memory unit for each piece of information. As a result, possessing the relevant problem-specific schemata generally improves an individual’s problem-solving performance.

When an individual is presented with new information or an unfamiliar task, the individual’s working memory is used to process the information. If the processing requirements of a task do not consume all of the individual’s working memory, the

available working memory will be used to encode the new information into long-term memory, thereby building an individual's schemata. Thus, the amount of working memory available bounds the limits of schemata acquisition (Sweller 1988; Mousavi et al. 1995; Rose and Wolfe 2000). As a result, reducing an individual's cognitive load during a task should help to improve an individual's schemata development.

2.5.2 Extending Cognitive Load Theory to Analytical Review.

Although Cognitive Load Theory was originally applied within a learning context, I extend the theory to task effectiveness because I suggest the same cognitive resource constraints that inhibit learning will also inhibit task effectiveness. In other words, I suggest that once the cognitive demands of the task (i.e., the cognitive load) consume all available working memory, no incremental processing can occur.

Consequently, I suggest that when the cognitive processing requirements of a task exceed an individual's working memory capacity, task effectiveness will suffer because the individual will be unable to attend to all of the elements necessary for successful task completion. My theoretical application is consistent with Brewster (2010), who investigates the link between cognitive load and effectiveness in an inference task and finds that reducing cognitive load results in better task effectiveness among individuals asked to make predictions about how changing business conditions should impact a firm's operating results.

Cognitive load is created by the demands of the problem solving requirements.¹⁶

Recalling, acquiring, processing, and manipulating information are all examples of

¹⁶ The inherent nature of analytical review is complex and this study's task process requirements are also of a complex nature. By complex, I mean the consideration of a variety of factors (i.e. cues, mental processes, etc) is required for successful task completion. Thus, a maintained assumption of this study is that the results may be generalizable only to situations where both task and processing requirements are complex.

problem solving processes that create cognitive load. Thus, task complexity increases with the number of informational cues a decision-maker is required to acquire and evaluate (Speier 2006). As cognitive load is created, working memory is consumed, leaving less working memory available for the acquisition, processing, and storage of information (Rose and Wolfe, 2000). As an individual's available working memory decreases, the cognitive resources available for problem solving decrease, ultimately having a negative effect on task performance (Rose and Wolfe 2000; Rose 2004). When there are not enough cognitive resources available to process information (i.e., when there is too much information to process), an individual often chooses to reduce their cognitive load by attempting to reduce the cognitive demands of the task. For example, Speier (2006) finds individuals are willing to trade a significant degree of decision accuracy for reduced cognitive load. This behavior is consistent with Simon's (1956) concept of *satisficing*, whereby an individual seeks to perform a task in a manner that is good enough, though not optimal, due to human processing constraints. Bonner (2008) finds evidence that auditors attempt to reduce their cognitive load by reducing the cue set considered. Thus, while satisficing and cue set reduction may be useful strategies in some contexts, such as the consideration of factors to help an individual decide which apartment to rent, these strategies are not acceptable within analytical review since generating and selecting the correct answer is most likely when all the relevant information cues are considered by the auditor.

2.5.3 The Three Types of Cognitive Load.

The amount of cognitive load imposed by a task is the additive result of three elements: intrinsic cognitive load, germane cognitive load, and extraneous cognitive load

(Sweller 1994). Intrinsic cognitive load is characterized by the amount of cue interactivity inherent in the task; higher levels of cue interactivity lead to greater cognitive load. A cue is a piece of information that needs to be acquired to successfully complete a task. Cue interactivity refers to the extent to which each cue can be meaningfully acquired without having to consider the relationship between it and other cues (Sweller 1994). Thus, cues interact if they are related in a manner which requires them to be considered simultaneously for successful task completion. Intrinsic cognitive load is constant and un-alterable for a given task because it is inherent to the task itself. In other words, an individual will experience a certain amount of cognitive load from the underlying nature of task itself. Therefore, intrinsic cognitive load is task idiosyncratic. Analytical review is an inherently complex task since there is a great degree of interactivity among the elements of the task (i.e., accounting relationships and cues necessary for successful task completion). Thus, analytical review should impose a relatively high degree of intrinsic cognitive load. Since intrinsic cognitive load is created by the task and since my task is the same across all treatment conditions, the level of intrinsic cognitive load should be constant within this study. Further, the decision aids utilized in this study should not reduce intrinsic cognitive load because they do not change the underlying nature of the task.

Germane cognitive load is the mental effort an individual devotes to processing information and acquiring task schema (Horcher and Tejay 2009). Since the amount of mental effort an individual exerts during a given task varies by individual, a single task can impose varying degrees of germane cognitive load. Consequently, individuals experience varying degrees of germane cognitive load based upon an individual's

idiosyncratic cognitive resources and the mental effort expended by the individual. Therefore, germane cognitive load is decision-maker idiosyncratic. Within this study, although I acknowledge that germane cognitive load varies based upon idiosyncratic factors, I assume the random assignment of participants to treatment conditions will balance germane cognitive load. However, to measure and potentially control for germane cognitive load, I ask participants to self-report the level of mental effort they exerted during the task.

Extraneous cognitive load is characterized as the cognitive load that arises from the method or processes an individual employs to perform a task (Sweller 1994). In other words, extraneous cognitive load is the cognitive load generated by the individual's problem-solving approach (Chandler and Sweller 1991). Extraneous cognitive load is thus controllable in the sense that the method or processes used to conduct a task may lend themselves to alteration in a manner that reduces cognitive load.

The decision aids utilized in this study are specifically designed to reduce the amount of extraneous cognitive load placed upon auditors during the analytical review process by helping auditors with the *process* of accurately recalling and maintaining information needed for successful task performance. Since the support provided by the decision aids is external to working memory, the aids should reduce extraneous cognitive load because auditors will not need to recall and maintain the information in working memory. The ARD should reduce extraneous cognitive load because it helps auditors with the *process* of accurately recalling and maintaining accounting relationships. The PCA should reduce extraneous cognitive load because it helps auditors with the *process* of accurately recalling and maintaining the auditor-identified relevant cues.

I next present an in-depth discussion of the hypothesis generation stage, the hypothesis evaluation stage, and the hypothesis selection stage, highlighting known deficiencies in each stage and apply Cognitive Load Theory to provide an explanation for these deficiencies. After a discussion of the three stages, I introduce each of my interventions (i.e., decision aids) by treatment condition, explain how the design of each intervention is rooted in Cognitive Load Theory, and provide an overview of why each intervention should reduce the level of extraneous cognitive load placed upon auditors during the hypothesis generation, hypothesis evaluation, and hypothesis selection stages. I frame my intervention predictions as hypotheses.

2.6 The Hypothesis Generation Stage

After using analytical review procedures to identify an unexpected fluctuation, the auditor begins an investigation to find its cause. During hypothesis generation, the first stage of the investigation process, the auditor obtains plausible hypotheses to explain why the unexpected fluctuation occurred.¹⁷ The auditor can acquire hypotheses externally or self-generate them using his or her domain and context specific knowledge (Koonce 1993; Asare and Wright 2001).¹⁸ The literature suggests hypothesis generation is the most important stage of analytical review because the failure to generate sufficient hypotheses negatively impacts performance across the other stages (Asare and Wright 1997, Asare and Wright 2003; Green 2004). In practice, the auditor usually asks the client for an explanation once an unexpected fluctuation is identified; 53 percent of

¹⁷ Recognizing that the information search and hypothesis generation is likely to occur at the same time, the analytical review literature classifies both hypothesis generation and information search as part of the same stage, even though the auditor is likely to have acquired some relevant information prior to this stage based upon their knowledge of the client (Koonce 1993; Asare and Wright 2001).

¹⁸ An example of an externally obtained explanation is when the auditor directly asks the client to explain why the fluctuation occurred.

auditors make client inquiry their first step, while 33 percent of auditors make client inquiry their second step (Trompeter and Wright 2010). Thus, client explanation is usually the first source in hypothesis generation.

Although client explanations should be corroborated with additional audit evidence, the extent to which auditors' accept client explanations remains an empirical question. Since client inquiry is generally the first step of the investigation process, academics and public watchdogs have voiced concern that this practice may lead to biased judgments (Trompeter and Wright 2010). For example, The Public Oversight Board's Panel of Audit Effectiveness specifically raised this concern in its annual report, recommending that standards setters "Develop more guidance on when it is appropriate (and when it is inappropriate) for the auditor to rely on management's explanations during the course of the audit and on obtaining additional evidence to corroborate those explanations" (POB 2000, p. 43). More recently, empirical evidence suggests auditors may accept a client's explanation without corroboration nearly one-third of the time (Trompeter and Wright 2010).

2.6.1 Risks Associated With Client Explanations.

Although it may be more efficient to initially approach the client for an explanation (i.e., it may be quicker to get a client explanation as compared to the amount of time required to self-generate one or more explanations), doing so raises audit effectiveness concerns. While auditors are cognizant of source reliability limitations, making sure to evaluate the competence of source of the information when evaluating an explanation's sufficiency (Hirst 1995), accounting research demonstrates auditors may inappropriately fixate on inherited client explanations. Bedard and Biggs (1991) find the

receipt of an inherited client explanation negatively affects subsequent hypothesis self-generation because even experienced auditors fixate on the client's explanation, generating fewer alternative hypotheses than if a client explanation was not first obtained. In an experimental setting, Anderson and Koonce (1995) found that senior auditors accept the client's incorrect explanation as the cause of the unexpected fluctuation nearly 66 percent of the time even though the client's explanation accounted for only 40 percent of unexpected fluctuation's variance.¹⁹ Further, a wealth of studies demonstrate that both inexperienced and experienced auditors generate fewer alternative hypotheses in the presence of an inherited explanation, regardless of the source (Anderson et al. 1992; Church and Schneider 1993; Asare and Wright 1997; Bierstaker et al. 1999; Asare and Wright 2003). In an experimental setting, Green (2004) finds that although auditors' overall analytical review effectiveness did not differ based upon inheriting an incorrect client explanation, auditors who receive an inherited and incorrect client explanation selected it approximately 40 percent of the time, whereas auditors who did not receive the incorrect client explanation never selected it as the cause of the unexpected fluctuation.²⁰ Further, auditors receiving the inherited and incorrect client explanation clustered their substantive testing around it (Green 2004). In sum, there is empirical evidence to support

¹⁹ Within the analytical review literature, the term "explanation inherited from the client" or "inherited client explanation" means the explanation provided to participants by the researcher (ostensibly from the client) before participants are asked to formulate their own hypothesis. This operationalization is necessary within laboratory settings because participants are not able to personally ask the client for an explanation.

²⁰ Green (2004) operationalizes analytical review effectiveness as the auditor's selection of the correct cause of the unexpected fluctuation. Auditors' overall analytical review effectiveness did not differ based upon the presence of an inherited and incorrect hypothesis because 76.2 percent of auditors who received an inherited and incorrect hypothesis failed to select the correct hypothesis compared to 75.0 percent of auditors who did not receive an inherited and incorrect hypothesis.

audit effectiveness concerns stemming from the auditor's decision to approach the client for an explanation before self-generating hypotheses.

I examine hypothesis self-generation due to the audit effectiveness concerns associated with the practice of approaching the client first and because asking the client for an explanation is not likely to be the best course of action in at least two situations: when an accounting error is the cause of the unexpected fluctuation and when management fraud is the cause of the unexpected fluctuation. In the case of accounting errors, the client is not likely to be aware that an error exists since the client would presumably not permit known, uncorrected errors within their financial statements.²¹ In the case of management fraud, client personnel associated with the fraud have a disincentive to disclose the actual cause of the unexpected fluctuation. Since restatements arising from accounting errors are far more prevalent than intentional manipulation (Palmrose et al. 2004; Hennes et al. 2008; Plumlee and Yohn 2010), this study examines a situation where the unexpected fluctuation is the result of an accounting error.²² Since seeking a client explanation is moot in a client error scenario, auditors within my study are asked to self-generate their own hypotheses. Throughout the rest of the study, any references to "hypothesis generation" imply hypothesis self-generation.²³

²¹ It should be recognized that the decision aids utilized in this study could be used to examine the veracity of a client's explanation in situations where the auditor first asks the client for an explanation before self-generating hypotheses. However, as an examination of inherited client explanations is beyond the scope of this study, I leave the question open to future research.

²² These studies suggest intentional manipulation is responsible for financial restatements 21 percent, 25 percent, and 3 percent of the time, respectively. Internal client error is responsible for financial restatements 76 percent and 57 percent of the time (Hennes et al. 2008; Plumlee and Yohn 2010, respectively).

²³ I do not suggest the auditor *must* refrain from approaching the client for an explanation, rather, I suggest it is a more appropriate course of action after the auditor has self-generated his or her own hypotheses because research suggests it is difficult for an auditor to completely disregard an inherited client explanation.

2.6.1.1 Hypothesis Self-Generation.

During the hypothesis generation stage, auditors generally propose multiple hypotheses because an unexpected fluctuation can have many potential causes and it is unlikely the auditor's first proposed hypothesis will identify the correct cause. To effectively generate hypotheses, the auditor needs to develop a mental model that organizes the relevant information into a pattern.²⁴ Then, the auditor draws upon his or her accounting knowledge to make inferences about potential causes of the unexpected fluctuation from the pattern.

Accounting research suggests that auditors have trouble performing hypothesis generation effectively. While auditors are generally able to identify the relevant cues needed to generate the correct hypothesis (Bedard and Biggs 1991; Bedard, Biggs, and Maroney 1998), they have trouble considering the cues in combination and, therefore, frequently fail to generate a hypothesis that addresses all cues. This gives rise to the first part of the analytical review paradox: even though auditors usually identify the relevant information cues needed to propose the correct cause of an unexpected fluctuation, they frequently do not propose the correct cause.

In an experimental setting Bedard, Biggs, and Maroney (1998) find 82 percent of experienced auditors acquire and correctly interpret the relevant cues needed to propose the correct hypothesis. A similar study by Bedard and Biggs (1991) find 86 percent of experienced auditors do so. However, these two studies find only 9 percent and 29 percent of experienced auditors propose the correct hypothesis, respectively.²⁵ In a more

²⁴ Some examples of potentially relevant information are macro economic conditions, industry norms, client operations, how operational activities impact account balances, and how inter-related accounts affect financial reporting.

recent study that utilized a computerized analytical review task, Green and Trotman (2003) find that 61 percent of experienced auditors propose the correct hypothesis. From an effectiveness standpoint, I suggest these results leave room for improvement.

Drawing upon cognitive load theory, I suggest that the inherent complexity of the hypothesis generation stage explains why auditors' performance can be sub-optimal. Hypothesis generation is inherently complex because auditors must draw upon their domain-specific knowledge to recall how the numerous accounts relate to each other, apply context-specific knowledge to successfully identify the relevant cues, and mentally arrange the relevant cues into a meaningful pattern which enables them to propose the correct hypothesis. Research provides evidence working memory limitations reduce individuals' ability to construct proper mental models of complex environments even in situations where individuals possess all the relevant information (Serman 1989; Sweeney and Serman 2000). This occurs because individuals can process and retain only a few items in working memory and an individual's ability to hold and manipulate items in working memory quickly decreases as the inter-relationships among items increases (Sweller et al. 1998; Engle and Kane 2003; Brewster 2010).

2.7 The Hypothesis Evaluation Stage

When the auditor finishes generating hypotheses, the next step is to evaluate each hypothesis to determine if any satisfactorily explain the cause of the unexpected fluctuation (Koonce 1993; Hirst and Koonce 1996). This evaluation process is called the

²⁵ These two studies constitute the known universe of research examining whether auditors are able to identify to the relevant cues during analytical review. While a bit dated, there is no empirical evidence suggesting that auditors' cue identification effectiveness has improved. Further, I argue that since these two studies examine experienced auditors, who are more likely to be able to identify the relevant cues, the fact that inexperienced auditors are increasingly performing analytical review suggests auditors' analytical review effectiveness may actually be decreasing since it is less likely the correct hypothesis will be proposed if all the relevant cues are not identified.

hypothesis evaluation stage. During hypothesis evaluation, the auditor must consider the plausibility of each hypothesis in terms of whether the account relationships embedded within it are consistent with the unexpected fluctuation. The auditor must also consider the completeness of each hypothesis in terms of how much of the unexpected fluctuation's variance it explains. Therefore, the auditor must determine how many relevant cues support each proposed hypothesis. Theoretically, as the number of relevant cues supporting a proposed hypothesis increases, the amount of variance explained by the hypothesis increases. The correct hypothesis should be plausible *and* account for the unexpected fluctuation's variance.²⁶ During the hypothesis evaluation stage, the auditor may decide that none of his or her proposed hypotheses correctly explain the cause of the unexpected fluctuation; if this occurs, the auditor returns to the hypothesis generation stage. Alternatively, the auditor may select one of his or her proposed hypotheses as the correct explanation of the unexpected fluctuation, thus moving to the hypothesis selection stage. The hypothesis evaluation stage has analytical review effectiveness implications because the auditor's final assessment of the likelihood that a given hypothesis is the correct cause of the unexpected fluctuation will guide the auditor's hypothesis selection decision (Asare and Wright 2001).

Research demonstrates that both inexperienced and experienced auditors have difficulty evaluating multiple hypotheses (Bedard and Biggs 1991; Jamal et al. 1995; Anderson and Koonce 1995; Asare and Wright 2001; Asare and Wright 2003; Green and Trotman 2003; Green 2004), leading them to frequently perform hypothesis evaluation

²⁶ Although the correct hypothesis should account for all the variance from expectation, given that the expectation setting process is usually not perfect leads both academics and practitioners to generally hold that a correct hypothesis will account for substantially all the variance from the expectation.

poorly, which ultimately results in the selection of an incorrect hypothesis.²⁷ Using verbal protocol analysis to obtain insight into the hypothesis evaluation process, Bedard and Biggs (1991) find that among auditors who correctly identify the cue pattern needed to evaluate their proposed hypotheses, 57 percent fail to disconfirm hypotheses that are not consistent with the cue pattern. When the auditor's evaluation process does not appropriately disconfirm a hypothesis, the hypothesis remains available to the auditor for selection even though it should be discarded. Green and Trotman (2003) provide further evidence that auditors have difficulty evaluating competing hypotheses during analytical review; in an experiment containing 63 experienced auditors who self-generated the correct hypothesis during the hypothesis generation stage, only 38 percent ultimately selected the correct hypothesis. Anderson and Koonce (1995) also provide evidence that auditors have difficulty evaluating hypotheses; in a between-subjects experiment, experienced auditors were asked to evaluate a client's explanation for the reason gross margin unexpectedly increased. When given a client explanation that accounted for only 40 percent of the unexpected fluctuation, 66 percent of the auditors rated it as likely to be the correct cause.

Applying Cognitive Load Theory, I suggest that auditors lack the cognitive resources required to effectively evaluate the plausibility and completeness of each proposed hypothesis because I posit the cognitive load associated with recalling,

²⁷ An experimental study by Bedard, Biggs, and Maroney (1998) finds the correct hypothesis was self-generated and selected by 1 auditor out of 11. However, combining the performance of individuals and groups, when self-generated the correct hypothesis was selected 50.0 percent of the time. In a study where auditors were given hypotheses and asked to choose the correct one Asare and Wright (2003) finds that 43 percent of auditors select the correct hypothesis. I suggest these numbers are sub-optimal from an audit effectiveness standpoint.

maintaining, and processing the relevant information often exceeds the auditors' available working memory.

2.8 The Hypothesis Selection Stage

The hypothesis selection stage is characterized by the auditor's decision to select one of their proposed hypotheses as the cause of the unexpected fluctuation. The auditor's decision-making performance in this stage is critical to performing analytical review effectively (Asare and Wright 2001). If the auditor accepts an incorrect hypothesis to explain the cause of the unexpected fluctuation four negative audit outcomes may follow. Three of the negative outcomes have audit effectiveness implications: If the auditor accepts an incorrect, non-error explanation when the unexpected fluctuation is actually caused by an error the financial statements may be misstated. If the auditor accepts an incorrect, non-fraud explanation when the unexpected fluctuation is actually caused by fraud the auditor will fail to recognize audit fraud risks and the financials may be misstated. If the auditor accepts an incorrect, non-error explanation when the unexpected fluctuation is actually caused by another non-error cause, the financial statements will not be materially misstated, but the auditor may acquire an incorrect understanding of the client, affecting the overall effectiveness of the audit.

The fourth negative outcome has audit efficiency implications: The auditor may conclude the unexpected fluctuation is caused by an error or fraud when it is not. Although efforts taken by the auditor to substantiate the false explanation should reveal the auditor's mistake, the additional audit effort results in wasted time and wasted audit resources.

The auditor's hypothesis selection effectiveness is dependent upon both the hypothesis generation stage and the hypothesis evaluation stage (Asare and Wright 2001). Since it is not possible to select the correct hypothesis if it is not proposed, the hypothesis generation stage impacts the auditor's hypothesis selection effectiveness. Additionally, because the hypothesis evaluation stage ends once the auditor believes he or she has identified the actual cause of the unexpected fluctuation, the judgments made during the hypothesis evaluation stage directly impact the auditor's hypothesis selection decision.

Analytical review research consistently finds that experienced auditors frequently fail to select the actual cause of the unexpected fluctuation, giving rise to the second half of the analytical review paradox: Even when auditors propose the correct cause of the unexpected fluctuation, they often fail to select it. For example, Bedard, Biggs, and Maroney (1998) find that only nine percent of the experienced auditors who participated in their experiment ultimately selected the correct hypothesis. This finding is consistent with more recent experimental research by Green and Trotman (2003), Asare and Wright (2003), and Green (2004) who find that experienced auditors select the correct hypothesis only 29, 28, and 25 percent of the time, respectively. Additionally, Asare and Wright (2003) investigate what impact giving an auditor the correct hypothesis in a set of plausible hypotheses has on selection effectiveness. Although Asare and Wright's results suggest this improves the auditor's hypothesis selection effectiveness, more than fifty percent of the auditors inheriting the correct hypothesis still failed to select it.

Hypothesis selection is the culmination of an auditor's analytical review efforts. Effective hypothesis selection depends upon the joint outcomes of hypothesis generation (because it is not possible to select the correct hypothesis if it is not proposed) and

hypothesis evaluation (because the hypothesis ultimately selected as the correct one is a direct result of judgments made during hypothesis evaluation). Applying Cognitive Load Theory, I suggest that hypothesis selection is difficult for auditors because the cognitive requirements of the preceding two stages can exceed the auditor's cognitive resources, making it difficult for auditors to select the correct hypothesis. I suggest this explains why auditors tend to perform hypothesis generation poorly, which leads to ineffective hypothesis evaluation, and ultimately leads to poor hypothesis selection.

Now that I have discussed the decision aid literature, Cognitive Load Theory, and presented an in-depth examination of the stages of analytical review investigated within this study, I next synthesize this discussion by introducing my study's two decision aids by treatment condition and illustrate how the design of each decision aid is rooted in theory. I then formally hypothesize how each treatment intervention should reduce cognitive load and lead to better analytical review effectiveness.

2.9 Hypotheses Development

2.9.1 The No-Aid Intervention.

The no-aid intervention reflects the current state of analytical review practice. This condition provides the base from which I compare the relative effectiveness of each decision aid intervention. Auditors in the no-aid intervention condition conduct analytical review without decision aids. Mirroring the analytical review process as currently conducted in practice, auditors in this condition are free to conduct analytical review in whatever individual manner they wish and are not restricted from taking notes or making their own decision aids to help conduct the task.

2.9.2 The Activity Relationship Diagram Intervention.

Since finding the cause of an unexpected fluctuation requires the consideration of multiple interacting accounts, the cognitive load involved with recalling and maintaining these relationships in working memory is likely to be high. The application of a diagram to a context where the task places a large demand upon working memory can be useful since the diagram can serve as a form of external memory, reducing the cognitive load placed upon the user (Hegarty and Steinhoff 1997; Rose 2002). Consequently, a diagram-type decision aid can permit the user to maintain a picture of a whole representation simultaneously (Scaife and Rogers 1996). Accordingly, I have created an “activity relationship diagram” (ARD), a diagram-type decision aid that provides auditors with a graphical depiction of common accounting relationships in a flowchart-type presentation format.²⁸

The ARD provides a diagram that depicts common client operations, how these operations impact the appropriate accounts, and how these accounts are related to other accounts. The aid pictorially displays the duality of accounting by demonstrating how an increase (decrease) in one account results in a concomitant decrease (increase) in related accounts.²⁹ Consequently, the ARD should permit auditors to recognize and recall their domain-specific knowledge of the relationships that exist between financial statement accounts and how operating activities are reflected in the financial statements. A key

²⁸ The use of flowcharts for documenting and presenting information is common-place in auditing and auditors are generally very conversant with the technique. For example, most auditors rely upon flowcharting to document business processes and internal controls as required by the Sarbanes-Oxley Act of 2002 (Romney and Steinbart 2009).

²⁹ The operating activities depicted by the ARD are common in the sense that they are not specific to the operations of the firm presented within my case materials.

aspect of the ARD is that it does not provide any new or problem-specific information.³⁰ For readability, I present the activity relationship diagram in two parts, Figure 2.1 and Figure 2.2.

By making common accounting relationships available to an auditor, the ARD should heighten the salience of these account relationships. This heightened relationship salience should help the auditor to perform the task more effectively because analytical review specifically requires the auditor to draw upon his or her domain-specific knowledge for effective task completion. Further, because it takes fewer cognitive resources to recognize information than recall it (Haist et al. 1992), the ARD should reduce an auditor's cognitive load because it eliminates the need to expend cognitive resources recalling and maintaining domain-specific knowledge (i.e., the account relationships) in working memory. I predict that by reducing the cognitive load placed upon an auditor during analytical review, the ARD will free cognitive resources that the auditor can use for task problem solving, leading to improved analytical review effectiveness compared to auditors who conduct the task without a decision aid.

³⁰ Because the ARD only presents information already known to auditors, this rules out the alternative explanation that ARD intervention participants are more effective at analytical review because they are provided with more information compared to other intervention treatments.

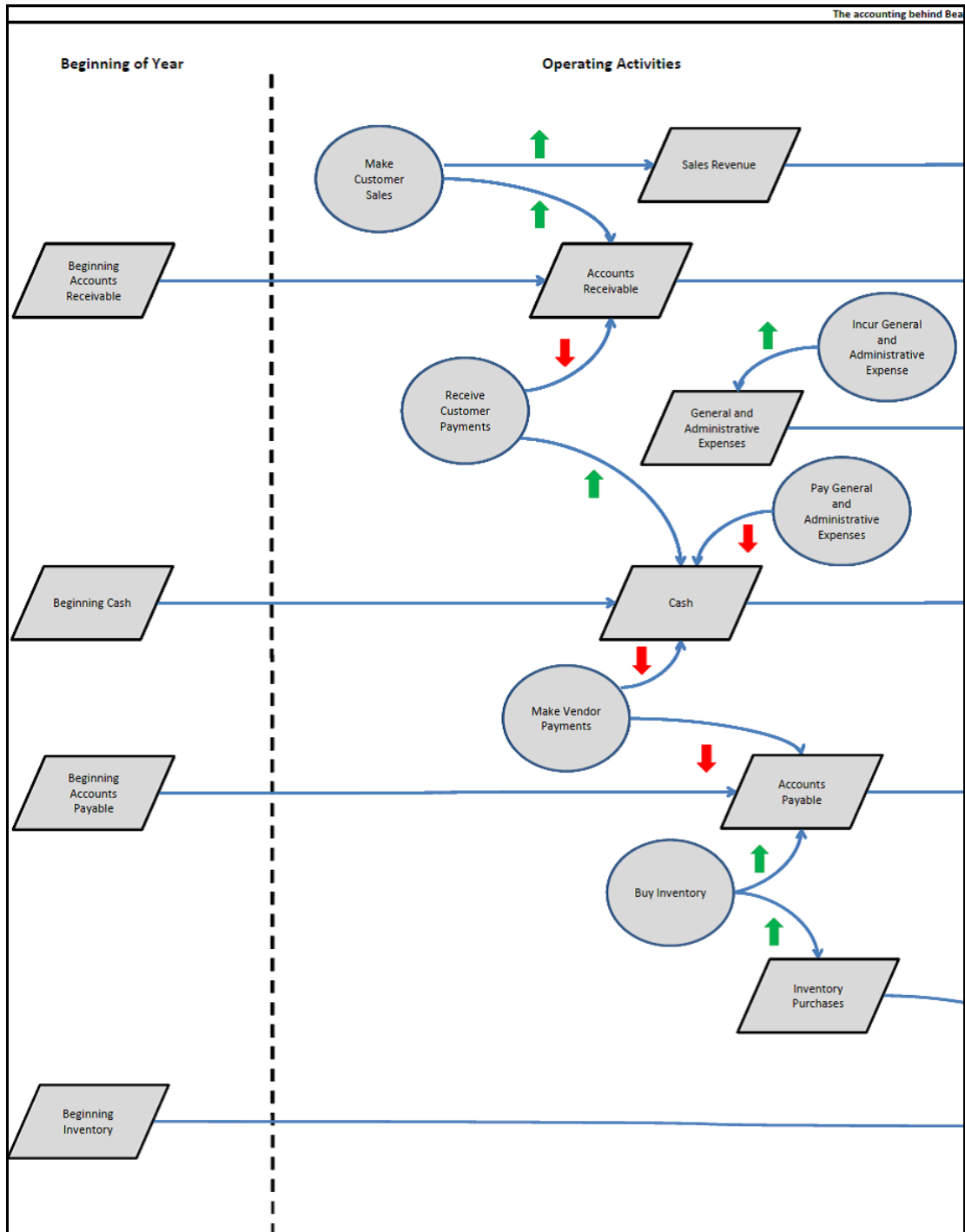


FIGURE 2a. The activity relationship diagram. (Part 1 of 2).

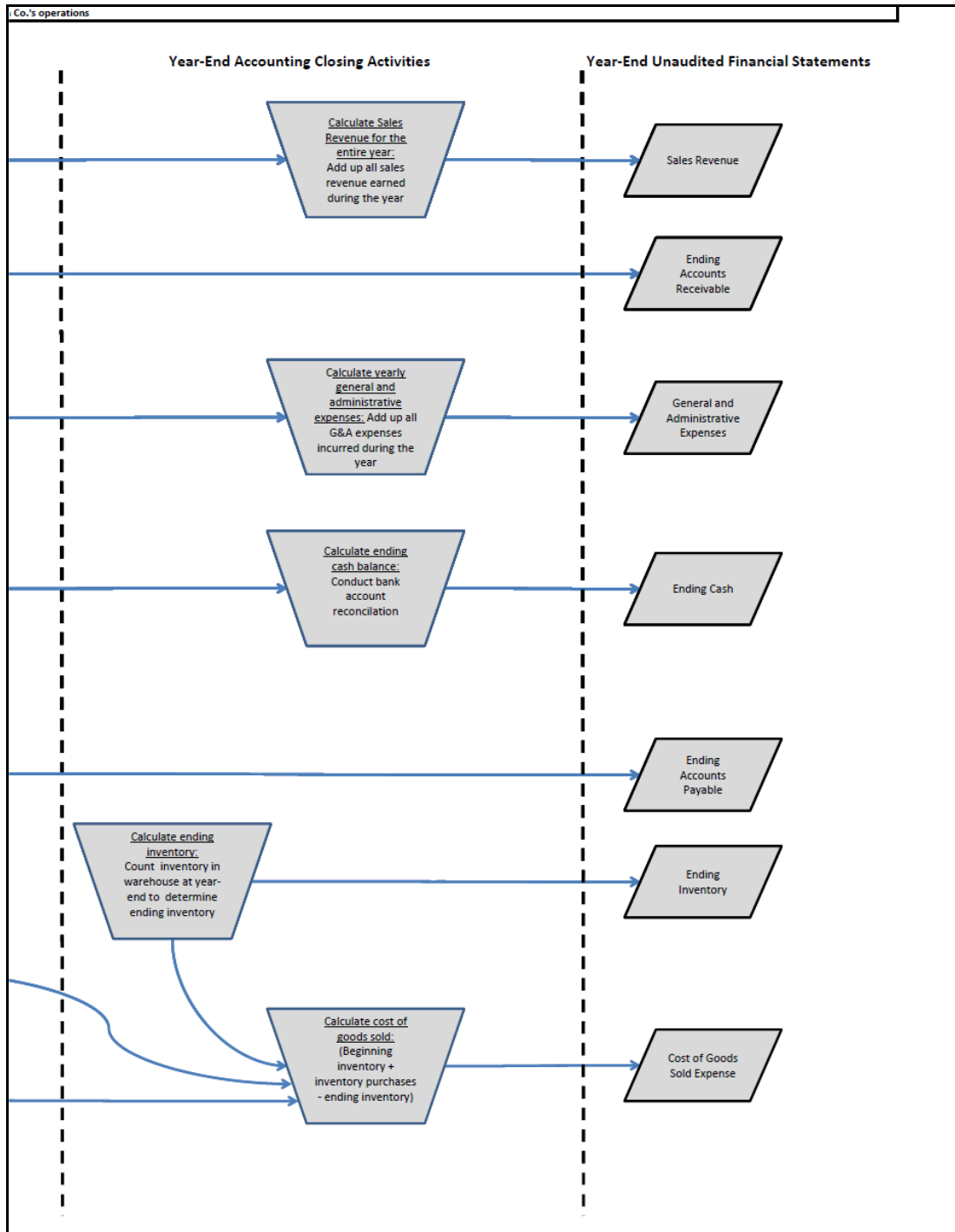


FIGURE 2b. The activity relationship diagram. (Part 2 of 2).

2.9.2.1 Hypothesis Generation Stage Predictions.

The ARD should help auditors during the hypothesis generation stage because it provides them with a graphical depiction of common accounting relationships, permitting auditors to accurately recall domain-specific knowledge. Specifically, by depicting common client operations, how these operations impact the appropriate accounts, and how these accounts are related to other accounts, the ARD should reduce the extraneous cognitive load imposed by the task during hypothesis generation since it eliminates the need for auditors to recall domain-specific information (i.e., accounting knowledge) and because it provides auditors with a place outside of working memory to maintain this information.³¹ In other words, the ARD should improve the hypothesis generation *process* by reducing extraneous cognitive load.

Since I predict the ARD will reduce extraneous cognitive load, aid users should have more working memory available for hypothesis generation. By increasing the amount of working memory available for problem-solving and by heightening the salience of account relationships, I predict that auditors within the activity relationship diagram intervention will conduct hypothesis generation more effectively by proposing more plausible hypotheses and by generating the correct hypothesis more often compared

³¹ There is a wealth of studies examining how information presentation impacts decision effectiveness. Generally, these studies look at whether information provided in tabular or graphical form impacts performance. An important distinction between this stream of literature and my study is that the decision aids utilized in this study do not present new information, and therefore, my study does not compare the effect of presenting information in one format over another on performance. Consequently, I do not focus on how presentation format affects performance, but rather whether presenting information that is already known to participants can reduce their cognitive load during analytical review, leading to better performance.

to auditors within the no-aid intervention. These predictions are stated formally in the following hypotheses:

H1a: A larger number of plausible hypotheses will be proposed by inexperienced auditors using the activity relationship diagram aid than by inexperienced auditors not using any aid.

H1b: The correct hypothesis will be proposed more often by inexperienced auditors using the activity relationship diagram aid than by inexperienced auditors not using any aid.

2.9.2.2 Hypothesis Evaluation Stage Predictions.

Since discerning the correct cause of the unexpected fluctuation is the goal of the hypothesis evaluation stage, an auditor's performance during this stage is effective in two situations. In the first situation, the auditor decides to return to the hypothesis generation stage because he or she realizes that none of the proposed hypotheses adequately explain the cause of the unexpected fluctuation. In the second situation, the auditor appropriately discerns the correct hypothesis by ruling out all incorrect hypotheses, leaving the correct hypothesis available for selection during the hypothesis selection stage. Therefore, a necessary prerequisite for effective hypothesis evaluation is that the correct hypothesis must be obtained.

To evaluate the plausibility of each hypothesis, the auditor must draw upon his or her accounting knowledge to determine whether the account relationships embedded within each hypothesis are consistent with the unexpected fluctuation. If the auditor does not properly construct and maintain the proper accounting relationships, he or she may fail to remove hypotheses that involve accounts that are not related to the unexpected fluctuation. For example, suppose the client's gross margin unexpectedly increases by \$120,000. Utilizing his or her knowledge of accounting, the auditor should start the

hypothesis evaluation stage by ruling out proposed hypotheses that involve accounts not associated with gross margin. Thus, a hypothesis proposing the gross margin increase is caused by a reduction in advertising should be removed since advertising has no impact on gross margin.³² Once an auditor has obtained the correct hypothesis, a decision aid can be said to improve the auditor's hypothesis evaluation effectiveness if using the aid helps the auditor to discern the correct hypothesis from the set of proposed hypotheses.

I predict that when an auditor evaluates a hypothesis set containing the correct hypothesis, the ARD should reduce the extraneous cognitive load experienced by the auditor during hypothesis evaluation by eliminating the need for the auditor to tie up working memory in recalling and maintaining account relationships. By lowering cognitive load, the ARD should increase the amount of working memory available for problem-solving. Further, by heightening the salience of account relationships, I predict it will be easier for auditors in the ARD intervention to rule out hypotheses that involve accounts that are not related to the unexpected fluctuation.

Consequently, by lowering cognitive load and heightening the salience of account relationships, the ARD should improve the auditor's hypothesis evaluation effectiveness. As a result, the correct hypothesis should be identified more often by auditors in the ARD

³² Scenarios similar to these can occur because the analytical review process is not static. Rather, the auditor's understanding of the unexpected fluctuation frequently changes during the information search process (which is classified as part of the hypothesis generation stage) based on the auditor's application of her accounting knowledge, the identification of cues believed to be relevant, and the formation of the cues into a meaningful pattern. Thus, auditors can (and do) propose implausible hypotheses during the analytical review.

intervention as compared to auditors in the no-aid intervention. This prediction is stated formally in the following hypothesis:

H2: When the hypothesis set contains the correct hypothesis, inexperienced auditors using the activity relationship diagram aid will identify the correct hypothesis more often when compared to inexperienced auditors not using any aid.

2.9.2.3 Hypothesis Selection Stage Predictions.

Recognizing that hypothesis selection effectiveness is a direct result of the effectiveness of both hypothesis generation and hypothesis evaluation, I predict that by heightening the salience of account relationships throughout the entire analytical review task it will be easier for auditors within the activity relationship diagram intervention to select the correct hypothesis compared to auditors within the no-aid intervention. I predict this because ARD aid users should more easily recall and reference the accounting relationships necessary to select the correct hypothesis and because they should experience lower extraneous cognitive load, leaving them with more working memory available for problem solving. Thus, the ARD should improve the effectiveness of the hypothesis selection stage; this prediction is stated formally in the following hypothesis:

H3: The correct hypothesis will be selected more often by inexperienced auditors in the activity relationship diagram intervention when compared to inexperienced auditors in the no-aid intervention

2.9.3 The Pattern-Consideration Aid Intervention.

Conducting effective analytical review requires an auditor to accurately recall, simultaneously consider, and successfully manipulate the cues relevant to the unexpected fluctuation. Without a decision aid, the auditor must perform these activities using working memory. In light of the high degree of cue interactivity existing within the

analytical review setting (i.e., that the consideration of each relevant cue must take into account the relationship existing between it and the other cues to be meaningfully processed), the cognitive load associated with these three activities is likely to be high. A decision aid that provides the auditor with a place outside of working memory to store and accurately recall the relevant cues should reduce the cognitive load placed upon the auditor during analytical review.

Consequently, I have created a pattern-consideration aid (PCA), a textual decision aid that stores and accurately recalls the auditor-identified relevant cues.³³ A key aspect of the PCA is that it does not provide users with any information that they do not self-identify. That is, the PCA does not provide users with any information that they do not type into the electronic PCA application.³⁴ The PCA should help an auditor to accurately recall and simultaneously consider the relevant information cues during the analytical review task, reducing the auditor's cognitive load by eliminating the need to expend cognitive resources recalling and maintaining the relevant cues in working memory. I predict that by reducing the cognitive load placed upon an auditor during analytical review, the PCA will free cognitive resources that the auditor can use for task problem solving, leading to improved analytical review effectiveness compared to auditors who conduct the task without a decision aid. Further, by textually presenting the auditor-

³³ I do not imply that auditors do not currently record the cues they identify as relevant. Further, studies which examine analytical review in the literature (i.e., Bedard and Biggs 1991; Bedard et al. 1998, etc) do not prohibit auditors from taking notes or recording cues. Rather, I examine whether enhancing the salience of the auditor-identified cues by automatically presenting them to the auditor at the appropriate time during analytical review can improve task effectiveness.

³⁴ Although discussed in depth in subsequent pages, the PCA will automatically recall the participant-identified relevant cues during the hypothesis generation, hypothesis evaluation, and hypothesis selection stages. Further, the PCA will also require the user to evaluate each proposed hypothesis against the participant-identified relevant cues during the hypothesis selection stage.

identified relevant cues, the PCA should also heighten the salience of the cues, helping the auditor to simultaneously consider them during hypothesis generation, hypothesis evaluation, and hypothesis selection.

2.9.3.1 Hypothesis Generation Stage Predictions.

Although research suggests auditors may reduce the number of cues they consider to decrease their cognitive load during analytical review (Bonner 2008), proposing the correct hypothesis is most likely when an auditor considers as many relevant cues as possible. Consequently, the use of a reduced-cue set strategy is not desirable during hypothesis generation. The PCA should help auditors perform hypothesis generation because it assists with cue recall and helps auditors to consider all cues simultaneously. Specifically, by storing and automatically presenting auditors with the cues they identify as relevant, the PCA should reduce the extraneous cognitive load associated with maintaining and recalling the cues in working memory. Further, by textually presenting auditors with a list that contains all of the cues, the PCA should heighten the salience of the cues, permitting auditors to more effectively consider the cues simultaneously while generating hypotheses. Auditors in the no-aid intervention must use their working memory to maintain the identified cues, correctly arrange the cues into a meaningful pattern, and propose explanations that address the relevant cues.

I predict that since auditors within the pattern-consideration aid intervention should experience lower extraneous cognitive load, they will have more working memory available for hypothesis generation. By increasing the amount of working memory available for problem-solving and by heightening the salience and processability of the auditor-identified relevant cues, I predict auditors within the pattern-consideration aid

intervention will conduct hypothesis generation more effectively by proposing more plausible hypotheses and by generating the correct hypothesis more often than auditors within the no-aid intervention. These predictions are stated formally in the following hypotheses:

H4a: A larger number of plausible hypotheses will be proposed by inexperienced auditors using the pattern-consideration aid than by inexperienced auditors not using any aid.

H4b: The correct hypothesis will be proposed more often by inexperienced auditors using the pattern-consideration aid than by inexperienced auditors not using any aid.

2.9.3.2 Hypothesis Evaluation Stage Predictions.

A decision aid can be said to improve an auditor's hypothesis evaluation effectiveness if using the aid helps the auditor to discern the correct hypothesis. Recall that obtaining the correct hypothesis is a necessary prerequisite for effective hypothesis evaluation since the auditor cannot identify the correct hypothesis if it has not been obtained.³⁵ To evaluate the completeness of a proposed hypothesis, the auditor must evaluate how well it addresses the cues identified as relevant to the unexpected fluctuation. In other words, the auditor must determine how well each proposed hypothesis matches the overall cue pattern. Ideally, the auditor will evaluate each proposed hypothesis to find one that is supported by the relevant cues. Because the correct hypothesis is most likely the one that is supported by every relevant cue,

³⁵ This does not imply that the correct hypothesis must be proposed the first time the auditor engages in hypothesis generation. The auditor may discover that none of his or her proposed hypotheses satisfactorily explain the unexpected fluctuation while conducting hypothesis evaluation. Thus, the auditor would return to the hypothesis generation stage to propose a hypothesis that appropriately explains the unexpected fluctuation.

employing a reduced-cue set strategy during hypothesis evaluation is likely to lead to ineffective hypothesis evaluation, ultimately leading to incorrect hypothesis selection.

The pattern-consideration aid intervention should reduce the extraneous cognitive load placed upon auditors during the hypothesis evaluation stage by giving auditors an external place to maintain, recall, and view the cues they identify as relevant. During the hypothesis evaluation stage, the PCA automatically presents an auditor with the cues the auditor previously identified as relevant. This should reduce the auditor's extraneous cognitive load because he or she will not tie up working memory maintaining and recalling the relevant cues during the hypothesis evaluation process.

In addition to reducing cognitive load and increasing the amount of working memory available for problem-solving, the PCA contains check-box functionality designed to help auditors determine how many cues support each proposed hypothesis. During the hypothesis evaluation stage, auditors are asked to click a box next to each cue the hypothesis under evaluation supports.³⁶ Then, a check mark appears next to the cue to indicate that the cue supports the hypothesis. After all the hypotheses have been evaluated, the PCA stores and displays the number of cues that support each hypothesis.³⁷ PCA users (i.e., auditors) can then click on each proposed hypothesis to see which specific cues support it. Logically, one would expect the correct hypothesis to be the one that is supported by all the relevant cues. Auditors who do not receive the PCA must use their working memory to determine how many cues support each of their proposed

³⁶ The cues contained within the PCA will be the ones entered into the application by each individual user. The PCA will not contain any cues or other information besides the information entered by the user.

³⁷ Pattern-consideration aid users can also “drill down” by double clicking on any hypothesis to see which specific cues support the hypothesis.

hypotheses. Further, auditors who do not receive the PCA must also use their working memory to keep track of the specific cues that support a given hypothesis.

Consequently, I predict the PCA's check-box functionality should reduce the cognitive load required to determine how many cues support each proposed hypothesis. Thus, it should be easier for auditors to discern which (if any) of their proposed hypotheses is best supported by the relevant cues, making it easier for auditors to identify the correct hypothesis. This prediction is stated formally in the following hypothesis:

H5: When the hypothesis set contains the correct hypothesis, inexperienced auditors using the pattern-consideration aid will identify the correct hypothesis more often when compared to inexperienced auditors not using any aid.

2.9.3.3 Hypothesis Selection Stage Predictions.

I suggest the most important aspect of hypothesis selection is the auditor's ability to determine which hypothesis best explains the cause of an unexpected fluctuation. By heightening the salience of the relevant cues and by reducing the cognitive load required to identify which of an auditor's proposed hypotheses is supported by the greatest number of relevant cues, I predict that the PCA's check-box functionality will lead auditors to select the correct hypothesis more often than auditors in the no-aid intervention. My prediction is formally stated in the following hypothesis:

H6: The correct hypothesis will be selected more often by inexperienced auditors using the pattern-consideration aid when compared to inexperienced auditors not using any aid.

2.9.4 Combined-Aid Intervention Predictions

Auditors receiving the combined-aid intervention conduct the analytical review task using both the activity relationship diagram and the pattern-consideration aid.

Because I expect each decision aid to reduce the cognitive load required to perform unique aspects of the analytical review task, I predict that auditors receiving both decision aids will experience the lowest levels of cognitive load compared to all other interventions. In other words, since I expect the ARD to reduce cognitive load by eliminating the auditor's need to expend cognitive resources recalling and maintaining common accounting relationships and since I expect the PCA to reduce cognitive load by eliminating the auditor's need to expend cognitive resources recalling and maintaining the relevant cues in working memory, I expect that auditors receiving both decision aids will experience less overall cognitive load than auditors receiving one (or none) of the decision aids because each aid should reduce the auditor's cognitive load in an independent way.

Although I suggest that giving auditors both decision aids should reduce their cognitive load, there is some empirical evidence to suggest that providing individuals with multiple decision aids may actually increase the level of cognitive load through a phenomenon termed the "split-attention effect" (Tarmizi and Sweller 1988). The cognitive literature demonstrates that the split-attention effect can occur in tasks that require individuals to use working memory to hold information from one source while simultaneously requiring individuals to integrate information from other sources for successful task completion. The split-attention effect increases cognitive load within problem-solving contexts. Consequently, empirical evidence demonstrates that requiring individuals to consider information from multiple sources of information can create cognitive load (Tarmizi and Sweller 1988; Ward and Sweller 1990; Sweller, Chandler, Tierney, and Cooper, 1990; Chandler and Sweller, 1992; Rose and Wolfe 2000).

Recognizing the split-attention effect research, I maintain that auditors who use both decision aids should experience the lowest levels of cognitive load compared to all other interventions because each of my decision aids is expected to reduce cognitive load through a different mechanism. The ARD should reduce cognitive load by providing an auditor with a place to store *common accounting relationships* outside of working memory, while the PCA should reduce cognitive load by providing an auditor with a place to store the *task relevant cues* outside of working memory. Further, in contrast to the tasks utilized in split-attention studies, which (by design) require an individual to maintain and integrate information from multiple sources in working memory, my decision aids provide auditors with a place to *store* information that is external to working memory. Consequently, I suggest that my decision aids do not constitute separate sources of information, but rather separate storage repositories that the auditor can use to offload information he or she would otherwise be required to maintain in working memory.

Since I expect each decision aid to incrementally reduce the cognitive load placed upon an auditor during analytical review by providing them with a place to store information outside of working memory, I predict that auditors within the combined-aid intervention will experience the lowest level of cognitive load, leading them to exhibit the most effective task performance. Consequently, I suggest that auditors in the combined-aid intervention will conduct hypothesis generation more effectively by proposing more

plausible hypotheses and by generating the correct hypothesis more often than any other intervention. These predictions are formally stated in the following hypotheses:

H7a: A larger number of plausible hypotheses will be proposed by inexperienced auditors using both aids than by inexperienced auditors using each aid individually or no aid.

H7b: The correct hypothesis will be proposed more often by inexperienced auditors using both aids than by inexperienced auditors using each aid individually or no aid.

Because I suggest both the ARD and PCA interventions will lead to more effective hypothesis evaluation as compared to the no-aid intervention, I propose that the combined-aid intervention will result in the most effective hypothesis evaluation because auditors within it will receive the predicted benefits of both decision aids during the task.

Thus, I propose:

H8: When the hypothesis set contains the correct hypothesis, inexperienced auditors using both aids will identify the correct hypothesis more often than inexperienced auditors using each aid individually or no aid.

Finally, because auditors within the combined-aid intervention have access to decision aids that both heighten the salience of the account relationships and the relevant cues, I predict they will select the correct hypothesis more often because it will be easier for them to identify which of their proposed hypotheses most plausibly and sufficiently explains the cause of the unexpected fluctuation. This prediction is formally stated:

H9: The correct hypothesis will be selected more often by inexperienced auditors using both aids than by inexperienced auditors using each aid individually or no aid.

Table 2.1 presents this study's hypotheses by decision aid and analytical review stage.

TABLE 2.1. Hypotheses by Decision Aid Intervention and Stage.

	Hypothesis	Decision Aid Intervention	Analytical Review Stage
H1a	A larger number of plausible hypotheses will be proposed by inexperienced auditors using the activity relationship diagram aid than by inexperienced auditors not using any aid.	Activity Relationship Diagram	Hypothesis Generation
H1b	The correct hypothesis will be proposed more often by inexperienced auditors using the activity relationship diagram aid than by inexperienced auditors not using any aid.	Activity Relationship Diagram	Hypothesis Generation
H2	When the hypothesis set contains the correct hypothesis, inexperienced auditors using the activity relationship diagram aid will identify the correct hypothesis more often when compared to inexperienced auditors not using any aid.	Activity Relationship Diagram	Hypothesis Evaluation
H3	The correct hypothesis will be selected more often by inexperienced auditors in the activity relationship diagram intervention when compared to inexperienced auditors in the no-aid intervention	Activity Relationship Diagram	Hypothesis Selection
H4a	A larger number of plausible hypotheses will be proposed by inexperienced auditors using the pattern-consideration aid than by inexperienced auditors not using any aid.	Pattern Consideration Aid	Hypothesis Generation
H4b	The correct hypothesis will be proposed more often by inexperienced auditors using the pattern-consideration aid than by inexperienced auditors not using any aid.	Pattern Consideration Aid	Hypothesis Generation
H5	When the hypothesis set contains the correct hypothesis, inexperienced auditors using the pattern-consideration aid will identify the correct hypothesis more often when compared to inexperienced auditors not using any aid.	Pattern Consideration Aid	Hypothesis Evaluation
H6	The correct hypothesis will be selected more often by inexperienced auditors using the pattern-consideration aid when compared to inexperienced auditors not using any aid.	Pattern Consideration Aid	Hypothesis Selection
H7a	A larger number of plausible hypotheses will be proposed by inexperienced auditors using both aids than by inexperienced auditors using each aid individually or no aid.	Combined-Aid Intervention	Hypothesis Generation
H7b	The correct hypothesis will be proposed more often by inexperienced auditors using both aids than by inexperienced auditors using each aid individually or no aid.	Combined-Aid Intervention	Hypothesis Generation
H8	When the hypothesis set contains the correct hypothesis, inexperienced auditors using both aids will identify the correct hypothesis more often than inexperienced auditors using each aid individually or no aid.	Combined-Aid Intervention	Hypothesis Evaluation
H9	The correct hypothesis will be selected more often by inexperienced auditors using both aids than by inexperienced auditors using each aid individually or no aid.	Combined-Aid Intervention	Hypothesis Selection

2.9.5 Planned Mediation Analysis.

A mediator is a variable that helps account for the relationship existing between an independent variable and a dependent variable (Baron and Kenny 1986). Thus, a mediator is a third variable that helps to explain how or why the relationship between the independent variable and dependent variable exists. The basic causal chain involved in mediation is illustrated in Figure 3, Panel A. The basic model is comprised of a three-variable system where two causal paths impact the dependent variable. The first path is the direct impact of the independent variable (Path A) upon the dependent variable. In the second path, the independent variable directly impacts the mediating variable (Path B), which in turn causes the mediating variable to directly impact the dependent variable (Path C). Thus, the second path involves an indirect relationship between the independent variable and the dependent variable.

Baron and Kenny (1986) note that a variable functions as a mediator when it meets the following three conditions: (1) the level of the independent variable significantly impacts the presumed mediator, (2) variations in the mediator significantly impact the dependent variable, and (3) when Paths *B* and *C* are controlled, a previously significant relationship between the independent and dependent variable is either decreased in significance or is no longer significant.³⁸

To test for mediation, Baron and Kenny suggest a series of regression models be estimated. Within this study, I apply Cognitive Load Theory to suggest that the level of cognitive load mediates the relationship between decision aid use and analytical review effectiveness. Accordingly, to test my assertion I conduct a planned mediation analysis.

³⁸ If the significance level is decreased, the mediation is said to be partial; if the direct relationship between the independent and dependent variable is no longer significant, mediation is said to be full.

Using the mediation analysis steps recommended by Baron and Kenny (1986), I use a mediation model to test whether the auditor’s use of a decision aid results in better analytical review effectiveness through a reduction in cognitive load, a mediating factor. Figure 3, Panel B illustrates the mediating relationship that I predict exists between the use of decision aids, the level of cognitive load, and analytical review effectiveness.

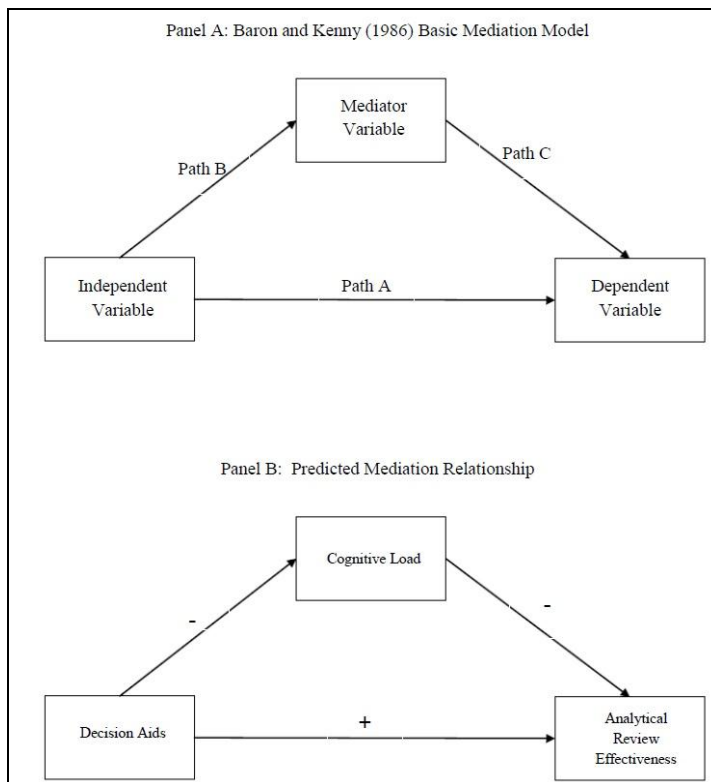


FIGURE 3. Mediation Models.

Specifically, I examine the extent to which decision aid use reduces cognitive load and the extent to which reductions in cognitive load are associated with improved analytical review effectiveness. A detailed discussion of the mediation analysis and results is presented in the Results section. Next, I discuss the experimental design, participants, and variables of interest in the Method section.

3.0 METHOD

3.1 Experimental Design

To test the hypotheses, I employ a 1 x 4 between-subjects design. I use this design because although I examine two distinct treatment interventions, both are manifestations of the same factor (i.e., decision aids). Thus, the two interventions are not independent factors, precluding a 2 x 2 factorial design.

The decision aid factor is manipulated at four levels: a no-aid intervention, an activity relationship diagram intervention, a pattern-consideration aid intervention, and a combined-aid intervention. The research design is illustrated in Figure 4.

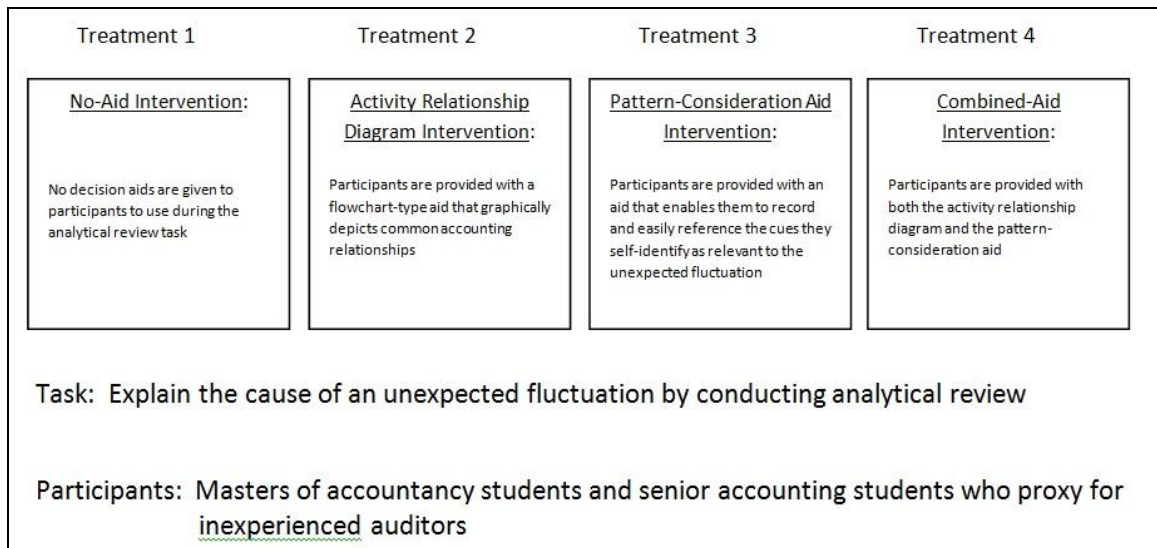


FIGURE 4. Research Design.

3.2 Case Design Considerations.

The case materials used in this study are developed to achieve three goals: to appropriately reflect the diagnostic nature of analytical review, to achieve reasonable external validity, and to permit experimental tractability.

To appropriately reflect the diagnostic nature of analytical review, the case requires participants to consider multiple cues to identify the correct cause of the unexpected fluctuation. To achieve reasonable external validity the case materials have been validated by practicing Big Four audit managers and seniors. Lastly, for experimental tractability the case is designed to have only one correct answer. However, I recognize this is abstraction from practice where an unexpected fluctuation may be the result of multiple underlying causes.³⁹ While the case materials used in this study are original, they have been modeled after prior analytical review research (Libby 1985; Bedard and Biggs 1991; Asare and Wright 1997; Asare and Wright 2001; Green and Trotman 2003). I abstract from practice for experimental tractability, particularly with regards to data coding. Employing only one correct answer simplifies the data coding analysis, which permits me to more accurately interpret the results. While designing a task to have more than one correct answer may improve external validity, doing so introduces more variation into the results, reducing strength of the inferences that can be made for each aid. For example, as the number of correct answers increases, it becomes increasingly difficult to discern the direct impact that each aid has in facilitating

³⁹ Though by no means the definitive guide to the proportion of times an unexpected fluctuation is comprised of more than one cause, a field study examining 19 unexpected fluctuations revealed that approximately 47 percent of the time an unexpected fluctuation was the result of more than one cause (Coglitore and Berryman 1988).

participants to a specific correct answer.⁴⁰ Further, designing a task with only one correct answer biases against finding results since participants must hypothesize, evaluate, and select *the* correct answer as opposed to *a* correct answer.⁴¹

3.3 Task

The task requires participants to explain the cause of an unexpected decline in an audit client's cost of sales ratio. The case was developed by establishing the correct account and ratio balances and then seeding the appropriate discrepancy cues. My seeded cue pattern was adopted from Kinney (1987) and there are six relevant cues that participants need to identify and consider to deduce the cause of the unexpected fluctuation: (1) the inventory turnover ratio is lower than expected, (2) the accruals ratio is higher than expected, (3) the gross margin ratio is higher than expected, (4) purchases are lower than expected, (5) accounts payable is less than expected, and (6) inventory costs did not decrease (i.e., the inventory purchase price, direct labor costs, and shipping costs did not change from the prior period). When the six relevant cues are considered as one pattern, the resulting conclusion is that the client failed to record an inventory purchase in the same period that it was subsequently recorded as a sale. In accordance with Kinney (1987), there are four alternative plausible explanations for this cue pattern. To rule out the four plausible alternatives participants are told the client, a fictitious company named Bean Co., buys and sells only one product - commoditized Jamaican coffee beans. Participants are also told that ending inventory is not miscounted or

⁴⁰ Using a case that contains more than one correct answer raises a host of coding issues, which ultimately affect the interpretation of results. For example, when there is more than one correct answer the researcher needs to consider whether both answers equally correct.

⁴¹ The more correct answers there are, the more likely it is that a participant will choose a correct answer through chance as opposed to successful task performance.

overpriced and that there is no evidence of inventory theft. Participants are also presented with information that explicitly tells them that inventory costs (i.e., the price per pound of coffee purchased, direct labor, indirect labor, shipping costs, overhead, and “all other costs required to make coffee beans available for resale”) have not changed from the prior year. Consequently, the seeded error, relevant cue pattern, and correct answer are consistent with those originally developed and used in Kinney (1987).

To increase task complexity, two irrelevant information cues are seeded into the case materials: (1) a significant reduction in general and administrative expenses, (2) and a significant increase in net income. While both of these cues significantly deviate from expectations, both cues are inconsistent with the cause of the unexpected fluctuation because the each cue is not related to the cost of sales ratio. The seeding of these two irrelevant cues is modeled after Bedard and Biggs (1991), who use irrelevant cues to increase task complexity.

3.4 Case Materials

Case materials inform participants, who assume the role of staff auditors on a financial statement audit, that their engagement senior has identified an unexpected decline in the cost of sales ratio during fieldwork. The case materials are fashioned after audit work papers and contain client background information, account balances, and ratios. The case materials are seeded with the appropriate cues necessary to explain the cause of the decline in the cost of sales ratio. Account balances and ratios are presented as “Audited,” “Expected,” and “Unaudited.” Audited information is described as the result of last year’s audit and, therefore, can be considered completely accurate. Expected information is described as the account expectations developed by the engagement

partner based upon his past client experience and current industry trends. To prevent participants from questioning the reliability of the partner's expectations, participants are told that the partner's expectations can be considered completely reliable. Unaudited information is described as current year information that has been provided by the client without any verification to its accuracy. As a result, participant attention should not be focused on the accuracy or reliability of the information provided by their audit firm, but rather on the information provided by the client. The information given to participants contains a column labeled "Threshold" which provides the upper and lower bounds of the partner's expectations. During training, participants learn that account balances or ratios that fall outside of threshold bounds indicate a material departure from audit expectations. The case materials specify that the Partner set the threshold to three percent of the expectation. Appendix A presents the case materials.

3.5 Participants

One hundred and twenty nine accounting students from a large southeastern public university served as this study's participants. To best reflect inexperienced audit staff, only MAcc and senior accounting students are included in the study's analysis.⁴² Participants were drawn from a rigorous accounting program whose audit curriculum emphasizes the audit standards.⁴³ The participants assumed the role of audit staff asked to find the cause of an unexpected fluctuation. Since MAcc and senior accounting students are almost identical demographically to staff auditors, these participants are an

⁴² Of the 129 participants, five were junior accounting students. All of the junior accounting students were excluded from this study's analysis because they were not exposed to Audit I. Additionally, all of the junior accounting students failed the accounting knowledge test (discussed in further detail under the heading titled *Pre-Test Knowledge Assessment* in the Results section).

⁴³ Excluding the five junior accounting students, all of this study's participants had exposure to Audit I.

appropriate proxy for staff auditors who now conduct analytical review 48 percent of the time (Trompeter and Wright 2010). Additionally, following the recommendation of Peecher and Solomon (2001), participants in all conditions were given a detailed analytical review training session before they started the task to impart the requisite task-specific domain knowledge and to equalize participants' knowledge of analytical review. To best reflect the content and presentation style of analytical review training received by audit practitioners, this study's training session was crafted using analytical review training materials obtained from a Big Four public accounting firm. Consequently, the training session provided to participants reflected the appropriate analytical review standards (i.e., AU 329, *Analytical Procedures*) and was similar to analytical review training received by staff auditor in public accounting.

3.6 Analytical Review Software Application

Participants perform the analytical review task using a customized Excel software application. The application displays Bean Co.'s financial information, permitting participants to manipulate the information as they would a normal Excel workbook. The application contains an administrator menu button, six action buttons, and three stage control buttons. Figure 5 presents a screenshot of the top part of the application's main screen and all ten buttons.

The administrator menu button is used by the researcher to manage the application. The six action buttons are used by participants to conduct the analytical review task. By clicking the first action button, a participant can record a cue they believe to be relevant to the cause of the unexpected fluctuation. Clicking the second action button brings up a screen that permits a participant to review, edit, or delete the cues the

participant has previously identified as relevant. Clicking the third action button brings up a screen where a participant can enter a proposed hypotheses. Clicking the fourth

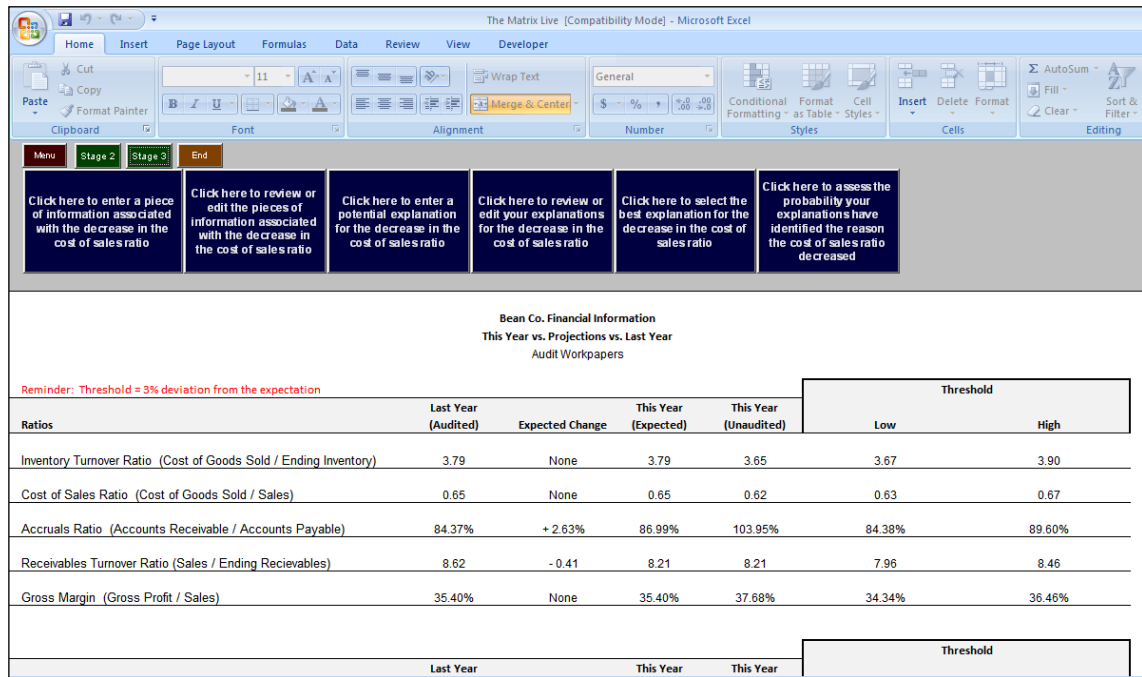


FIGURE 5. Screenshot of the Analytical Review Application's Main Screen.

action button brings up a screen that permits a participant to review, edit, or delete their proposed hypotheses. The fifth action button permits a participant to engage in hypothesis evaluation and selection. Finally, the sixth action button solicits a participant's input regarding the likelihood that each of their proposed hypotheses is the actual cause of the unexpected fluctuation.

The three stage control buttons are designed to structure the analytical review task in a manner that is consistent with the literature's classification of the steps in analytical review. The three stage control buttons help structure the task by restricting the action buttons a participant can see during each stage of the analytical review task. When the application is first launched a participant can only see the first two action buttons; these two buttons permit the participant to identify and edit the cues they believe to be relevant

to the unexpected fluctuation. Clicking the password-protected Stage 2 button permits a participant to see the next two action buttons, enabling the participant to engage in hypothesis generation. Pressing the password-protected Stage 3 button displays the final two action buttons which permit a participant to engage in hypothesis evaluation and selection. Finally, clicking the End button prompts the application to verify that a participant has both selected a hypothesis to explain the cause of the unexpected fluctuation and that the participant has rated the probability that each of their remaining proposed hypotheses were the “correct” explanation. If a participant failed to perform these two actions, the application displays a pop-up box instructing the participant to engage in the appropriate action. If a participant successfully completed these two actions, the application automatically saves the participant’s work, closes the application, returns the participant to the desktop, and prompts the participant to raise his or her hand for the last part of the experiment, the post-experiment questionnaire.

3.7 Experimental Procedure

The experiment consists of six parts: an online pre-test, analytical review training, case familiarization, hypothesis generation, hypothesis evaluation and selection, and an online post-experiment questionnaire. Figure 6 presents an illustration of the experimental procedure.

All experimental sessions took place within a university lab setting. Upon arrival, participants were asked to sit in front of an open laptop. The laptop contained the analytical review application which automatically recorded each participant’s work. At the appointed start time, the researcher closed the lab doors close and no additional

participants were admitted. After completing requisite IRB informed consent forms, participants completed an online pre-test designed to gauge their understanding of common account relationships. The pre-test also contained an activity designed to measure the size of each participant’s working memory. Appendix B contains the online pre-test administered to participants. Next, participants were given a 20 minute training session. The purpose of training was to provide all participants with a basic

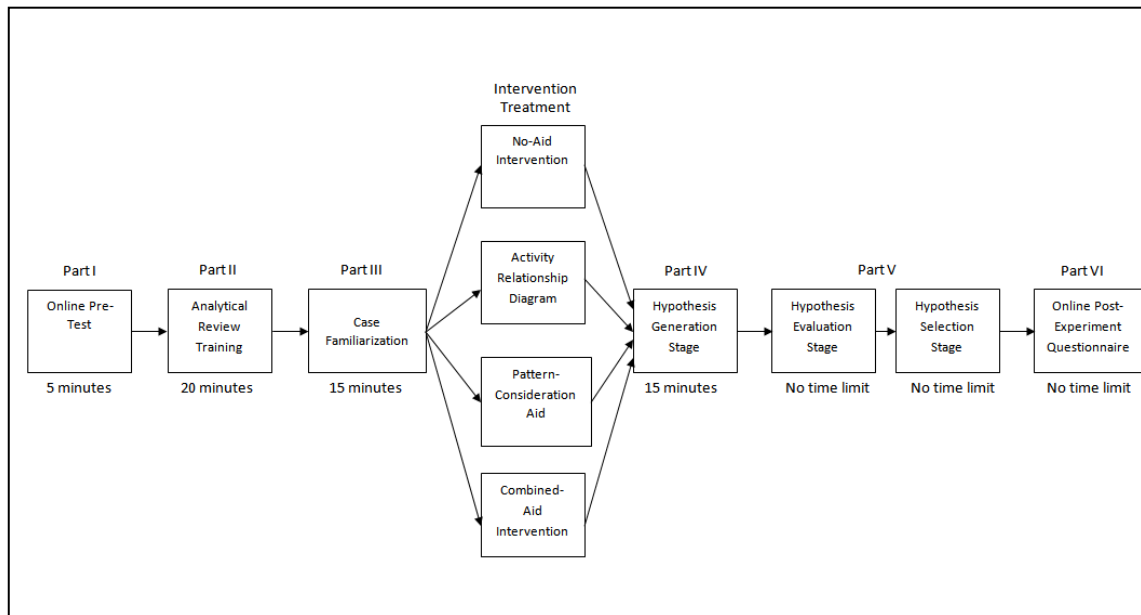


FIGURE 6. Experimental Procedure

understanding of analytical review. The training session explained why audit firms conduct analytical review and provided insight regarding the way analytical review is commonly conducted. Although the training session was much briefer than training provided in practice, the session was based on Big Four analytical review training materials. Appendix C contains the training materials used.

Although multiple experimental sessions were conducted, every participant within a given session experienced the same intervention treatment. The intervention

administered for any given session was determined by randomly selecting one of four numbered slips of paper from a hat. The purpose of this technique was to ensure that the researcher was blind to the experimental treatment during the training phase, to avoid introducing bias into the training procedure.

3.7.1 The Analytical Review Task and Decision Aid Manipulations.

The analytical review task consisted of three stages. After the training session, participants were instructed to open an envelope containing the case materials. In Stage 1, participants were advised that their engagement senior had discovered the audit client's cost of sales ratio was lower than expected during audit fieldwork. Participants then spent 15 minutes becoming familiar with the case materials.

Participants in all treatment conditions were explicitly asked to identify and record the cues they felt were relevant to the unexpected fluctuation. Participants in the ARD intervention, the PCA intervention, and the combined-aid intervention were asked to electronically record each cue by clicking the appropriate button on their analytical review application (i.e., button one) and by typing each cue into a prompt. Participants in the no-aid intervention were explicitly told that they could take notes in the task application's spreadsheet, in a separate electronic word processing document (i.e., Microsoft Word), or on any of the paper materials given to them. Additionally, to facilitate note taking across all interventions, participants in every intervention were given a new pen they could use to take notes.

Although the cues a participant typed into the software application were later used by pattern-consideration aid users in Stage 2, no explicit decision aid interventions occurred in Stage 1. After 15 minutes, the researcher instructed all participants to move

to Stage 2, hypothesis generation. The analytical review application prevented participants from moving to Stage 2 until instructed.

In Stage 2, the researcher orally asked participants to spend at least 15 minutes generating as many hypotheses as possible to explain the cause of the unexpected fluctuation. Participants within all conditions typed their proposed hypotheses into the analytical review application by clicking the appropriate button (i.e., button three). After the researcher introduced Stage 2, but before participants started proposing hypotheses, the decision aid manipulations occurred.

Participants who received the activity relationship diagram were then handed a one-page flowchart that depicted common accounting relationships. After every participant had a copy of the ARD, the researcher orally announced that the ARD might help participants formulate their proposed hypotheses because it might make it easier to see the links between the unexpected fluctuation and its' related accounts.

Participants in the pattern-consideration aid treatment were electronically presented with the cues they previously identified as relevant each time they clicked the button to propose a new hypothesis. Thus, PCA participants could easily refer to their cues each time they proposed a new hypothesis. To make the presentation of cues more salient, the application told participants that they might find it helpful to refer to the cues in formulating their proposed hypotheses. Figure 7 presents an example that illustrates how the PCA electronically presented the cues to participants during hypothesis generation. After 15 minutes, participants were permitted to move to Stage 3, hypothesis evaluation and selection. A participant was not forced to move to Stage 3 until they were ready. The application prevented participants from moving to Stage 3 until instructed.

When a participant was ready to move to Stage 3, the participant raised his or her hand. The researcher then handed the participant a sheet of paper asking the participant to evaluate each of their proposed hypotheses and to select the hypothesis the participant believed best explained the reason the cost of sales ratio declined. The handout also explained how to use the electronic application to complete Stage 3. For participants in the ARD intervention, the handout also informed participants that they might find it

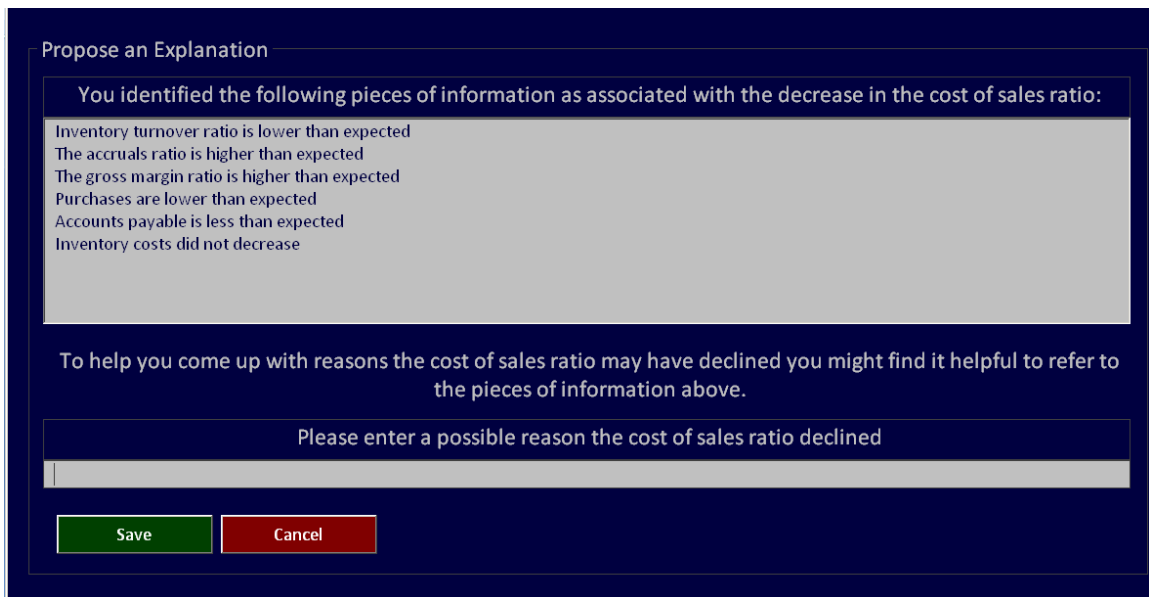


FIGURE 7. Screenshot Example of Participant-Identified Cues Displayed by the PCA During Hypothesis Generation.

helpful to refer to the ARD during the stage because the accounting relationships displayed in the ARD might make it easier to evaluate how well each proposed hypotheses addressed the cost of sales ratio decline.

To evaluate their proposed hypotheses, a participant clicked the appropriate button in the analytical review application (i.e., button five). The application then displayed all the participant's proposed hypotheses. To select a hypothesis as the correct reason the cost of sales ratio declined, participants in both the no-aid intervention and ARD intervention double clicked the desired hypothesis from a screen that displayed all

of the participant's proposed hypotheses. After a participant double-clicked the desired hypothesis, a pop-up box asked the participant to confirm his or her selection. Figure 8 presents an example of a screenshot that illustrates how participants in both the no-aid intervention and ARD intervention viewed their proposed hypotheses during hypothesis evaluation and selection.

The application required participants who received the PCA to evaluate each of their proposed hypotheses against the cues they previously identified as relevant before the application would permit them to select a "correct" hypothesis. To evaluate their



FIGURE 8. Screenshot Example of the Hypothesis Evaluation and Selection Screen. This is an example of the screen viewed by participants in the no-aid and activity relationship diagram interventions during hypothesis evaluation and hypothesis selection.

proposed hypotheses, a PCA user clicked on a proposed hypothesis, opening a new screen that displayed the participant-identified relevant cues. A check-box appeared to the left of each cue and the PCA user was instructed to check the box of each cue that supported the hypothesis. Figure 9 provides a screenshot example illustrating the PCA's

check-box functionality. After a PCA user evaluated all of their proposed hypotheses individually, the application permitted the user to select a hypothesis they felt best explained the cause of the unexpected fluctuation.

Reconcile Supporting Information with Explanations

You suggested the decrease in the cost of sales ratio may have been caused by:

Bean Co. may have failed to record an inventory purchase

Here are the pieces of information you identified as associated with the decrease in the cost of sales ratio. Please click the check box next to every piece of information that supports your proposed explanation:

- Inventory turnover ratio is lower than expected
- The accruals ratio is higher than expected
- The gross margin ratio is higher than expected
- Purchases are lower than expected
- Accounts payable is less than expected
- Inventory costs did not decrease

Save Cancel Select as Best Explanation

FIGURE 9. Screenshot Example of the Hypothesis Evaluation Checklist. This is an example of the screen viewed by participants in the pattern-consideration aid and combined-aid interventions during hypothesis evaluation.

Figure 10 provides a screenshot example that illustrates how PCA users selected the correct hypothesis.

In all treatment conditions participants were free to propose new cues and potential hypotheses during Stage 3. A participant might choose to do so if he or she felt none of the proposed hypotheses correctly explained the cause of the decline in the cost of sales ratio. After selecting the “correct” hypothesis, the application asked participants to assess the probability that each of their proposed hypotheses (including the hypothesis they selected as the correct one) was the actual cause of the cost of sales ratio decline. Participants entered their probability assessment for each proposed hypothesis into a

prompt provided by the application. Due to the iterative nature of analytical review, there was no specific time limit enforced for the completion of Stage 3.⁴⁴

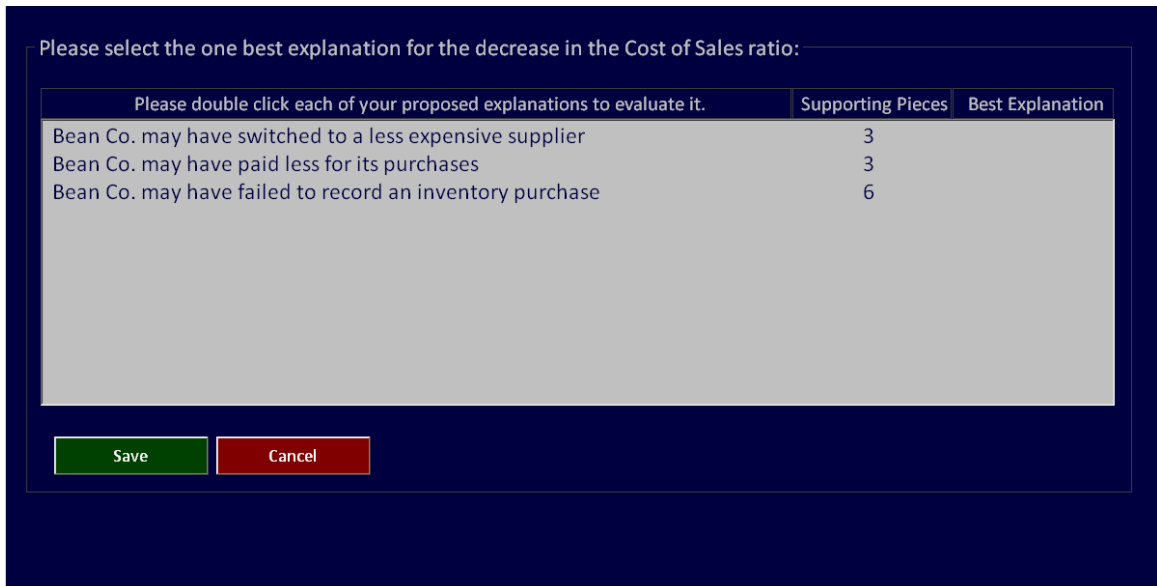


FIGURE 10. Screenshot Example of the Hypothesis Selection Screen. This is an example of the screen viewed by participants in the pattern-consideration aid and combined-aid interventions during hypothesis selection.

After completing Stage 3, participants completed an online post-experiment questionnaire. Once a participant completed the post-experiment questionnaire they were dismissed from the lab. Appendix D presents the questions asked in the post-experiment questionnaire.

To more clearly illustrate the oral and written instructions given to participants during the main task Appendix E contains the oral script for Stage 1 and Stage 2.

Appendix F contains the written instructions given to participants after they finished Stage 2.

⁴⁴ Although it may seem desirable to impose a time limit on participants for control purposes, such a limit actually constitutes both a significant departure from practice and could lead to false results. In practice, aside from budgetary considerations, there is no time limit imposed upon auditors to evaluate and select the cause of an unexpected fluctuation. Within this experimental setting, imposing a time limit may force some participants to choose an explanation they would not otherwise select due to time considerations.

3.8 Dependent Measures

I use four dependent measures to test my hypotheses. The first dependent variable is the mean number of plausible hypotheses generated by a participant within each intervention. A hypothesis is deemed plausible if it is consistent with the case information. That is, a hypothesis is plausible if it is both supported by information contained within the case and not contradicted by information provided in the case. The second dependent variable is the percentage of times the correct hypothesis is proposed by participants within each intervention. The third dependent variable is the percentage of times the correct hypothesis is identified by participants within each intervention. I operationalize this dependent variable by first discarding the data from those instances where participants did not propose the correct hypothesis. I do not use this data to analyze participants' hypothesis evaluation effectiveness because a participant cannot possibly identify the correct hypothesis if the participant did not propose it. Then, in those instances where the correct hypothesis was proposed, I calculate the percentage of times the correct hypothesis was actually selected by intervention condition. This proportion serves as my third dependent variable. I assert that this is an appropriate way to operationalize participants' hypothesis evaluation effectiveness because a participant will ultimately select the proposed hypothesis that he or she has evaluated to best explain the cause of the unexpected fluctuation. If my decision aids help participants to evaluate their proposed hypotheses, I would expect that in cases where the correct hypothesis is proposed, the correct hypothesis will be selected more often by aid users than by non-aid users. The fourth dependent variable is the percentage of times the correct hypothesis was selected by participants within each intervention condition.

The following example may help to illustrate the differences between dependent variables two, three, and four. Suppose the ARD intervention was administered to ten participants. Suppose further that of the ten participants, six proposed the correct hypothesis. The value of dependent variable two would be 60 percent for the ARD intervention. Now, suppose that of the six participants who proposed the correct hypothesis, three ultimately select it as the correct cause of the unexpected fluctuation. Thus, the value of dependent variable three would be 50 percent for the ARD intervention. Last, because three out of the ten participants in the ARD condition ultimately select the correct hypothesis, the value of dependent variable four would be 30 percent.

3.9 Data Coding

The researcher and one Ph.D candidate in the final stages of a doctoral program served as data coders. To prevent coding bias, neither coder was provided with information that could be used to identify the intervention treatment (i.e., no-aid, ARD, PCA, or combined-aid interventions) of the data. 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. Further, both coders independently classified the data in separate locations. 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.

3.9.1 Hypotheses Coding.

Coders were first asked to identify instances where a participant proposed the same hypothesis more than once. After removing duplicate hypotheses, coders were then instructed to use their judgment in determining whether each of the participant's proposed hypotheses were consistent with the case information. The coders independently classified each hypothesis as either "plausible" or "implausible" and then reconciled any discrepancies in their classification. Prior to resolving discrepancies, the Cohen's Kappa coefficient, a measure of inter-rater reliability, was 0.858 for 508 items coded, suggesting a high degree of coding agreement.⁴⁵ These coding results formed the data used to calculate the average number of plausible hypotheses proposed by participants in each intervention condition.

Next, coders were asked to review the set of plausible hypotheses to identify instances where the correct hypothesis was proposed. The coders independently classified each plausible hypothesis as either "correct" or "incorrect" and then reconciled any discrepancies. Prior to resolving discrepancies, the Cohen's Kappa coefficient was 0.927 for 79 items coded, suggesting a high degree of coding agreement. These coding results formed the data used to calculate the percentage of participants who proposed the correct hypothesis. Additionally, these coding results were also used to identify the instances where a participant's hypotheses set contained the correct hypothesis.

Lastly, coders were asked to examine each participant's selected hypothesis to determine if the participant chose the correct hypothesis. The coders independently

⁴⁵ Zwick (1988) notes that Cohen's Kappa is the most popular index for assessing inter-rater agreement of nominal categories. The degree of inter-rater agreement is determined by the value of the Cohen's Kappa coefficient. The strength of the inter-rater agreement (based on the Cohen's Kappa coefficient) is generally considered to be as follows: 0.00 to 0.20 poor, 0.21 to .40 fair, 0.41 to 0.60 moderate, 0.61 to 0.80 good, 0.81 to 1.00 very good (Fleiss et al. 2003).

classified each participant's hypothesis selection as "right" or "wrong" and then reconciled any discrepancies. Prior to resolving discrepancies, the Cohen's Kappa coefficient was 0.815 for 78 items coded, suggesting a high degree of coding agreement. These coding results formed the data used to calculate the percentage of times participants selected the correct hypothesis by intervention condition.

4.0 RESULTS

The Results section begins with a discussion of this study's participants. First, I present the criteria used to identify participants who failed the experiment's manipulation check, then I present the criteria used to assess participants' knowledge of the general accounting relationships depicted within the decision aid, discuss the criteria used to exclude these participants from my analysis, and present descriptive statistics on the final set of participants. Then, I test each of the hypotheses and report the results. I then present the results of the planned mediation analysis. Finally, I conclude the section by discussing the results of supplemental analysis.

One hundred and twenty nine individuals participated in my experiment. Participants were recruited from masters and upper-level accounting undergraduate courses at a large university in the Southeastern United States. Because my study examines the analytical review effectiveness of inexperienced auditors and because accounting seniors and MAcc students are hired by public accounting firms as new audit staff, I suggest that upper-level accounting students are an appropriate proxy for inexperienced auditors.

4.1 Analysis of Participant Demographic Information and Responses

Although one hundred and twenty nine individuals participated in my experiment, not all participant responses were useable. One participant was excluded from the analysis because their Excel application became corrupted during the experiment, making it

impossible to extract the data. Additionally, two questions in the post experiment questionnaire served as manipulation checks. The purpose of these questions was to identify whether participants were attentive to the task. The first manipulation check question asked participants to “Please select the unexpected fluctuation you investigated today.” The correct response to this question was “The unexpected fluctuation involved Bean Co.’s cost of sales ratio.” The second manipulation check question asked participants to “Please indicate what materials were available to you while conducting analytical review today (Check all that apply).” The available responses were: (1) Background information about Bean Co., (2) a flowchart depicting common accounting relationships, (3) an Excel application / workbook, (4) an electronic check-box screen that helped you evaluate your proposed explanations. Eleven participants were eliminated from the analysis because they failed the manipulation check questions, suggesting these participants did not conduct the experiment conscientiously. Of the 11 participants who failed the manipulation check questions, nine were accounting seniors and two were MAcc students.⁴⁶ There was no significant difference between the proportion of MAcc students and senior accounting students failing the manipulation check questions ($\chi^2 = 0.252$, $df=1$, $p=0.615$).

4.1.1 Pre-Test Knowledge Assessment.

A maintained assumption of this study is that participants already know the general accounting relationships depicted within the ARD. To verify this assumption, all participants took a pre-test before the main task to assess their understanding of the general accounting relationships depicted within the ARD. Excluding the participants

⁴⁶ Further, all 11 participants who failed the manipulation check questions also failed the accounting knowledge test (described in the next paragraph). This provides further evidence to support the claim that these 11 participants were not attentive to the task.

that failed the manipulation check questions, thirty-five participants failed the accounting relationship knowledge test by indicating that a sale of goods causes general and administrative expenses to increase. Participants were asked the question, “Which of the following would occur after a sale is made and the product is delivered?” Four clickable choices were presented and participants were asked to select all the choices that applied. The four choices presented were (1) sales revenue increases, (2) sales revenue increases but only when payment is received, (3) general and administrative expenses increase, and (4) general and administrative expenses increase but only when actually paid. The only correct answer is that sales revenue increases. However, 35 participants responded that general and administrative expenses would increase. A participant may have answered incorrectly due to a variety of factors: a participant may have not known the correct answer (i.e., they incorrectly believe 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 taken the question seriously. Since the unexpected fluctuation employed within this study is related to cost of sales, these participants were excluded from the analysis because it was unclear as to whether or not they possessed the accounting knowledge required to successfully complete the task. As expected, participants who failed the accounting relationship question were significantly less effective at analytical review than those who did not miss the question ($\chi^2=4.063$, $df=1$, $p=0.044$). Further, because the purpose of the ARD is to help a participant recall accounting relationship knowledge the participant already knows, rather than to provide the participant with new accounting relationship knowledge it was important to remove those who failed the accounting relationship test because the ARD might have presented

them with new accounting knowledge. Excluding the participants who failed the accounting relationship test biased against finding results because the relationships depicted in the ARD made users more effective in conducting the task. The analytical review effectiveness results for both the ARD and the combined-aid intervention were stronger when the participants who failed the accounting relationship knowledge test were not removed.

Of the 35 participants who failed the accounting relationship test, five were accounting juniors, 23 were accounting seniors, and seven were MAcc students. While none of the accounting juniors who participated in this experiment failed the manipulation check questions, all of the accounting juniors failed the accounting relationship knowledge test. There was no significant difference between the proportion of MAcc students and accounting senior students failing the accounting relationship test ($\chi^2 = 0.061$, $df=1$, $p=0.805$). Of those who failed the accounting relationship test, seven were in the no-aid intervention, nine in the ARD intervention, nine in the PCA intervention, and 10 in the combined-aid intervention. There was no significant difference in the proportion of those who failed the accounting relationship test by intervention condition ($\chi^2 = 2.717$, $df=3$, $p=0.437$).

4.1.2 Participants Reporting No Decision Aid Reliance.

To properly attribute a participant's task effectiveness to the decision aid(s) they received, my statistical analysis only includes participants who reported that they placed reliance on the decision aid(s) provided to them. Four participants that received a decision aid were excluded from my analysis because they reported placing no reliance on it while conducting the task. Of the four, one participant proposed the correct cause of

the unexpected fluctuation. The same participant ultimately chose the correct cause.

Table 4.1 provides a summary of all the participants excluded from this study.

4.1.3 Participant Descriptive Statistics.

Data from 78 participants were used to test this study's hypotheses.⁴⁷ Table 4.2 presents participant descriptive statistics. Aside from the amount of time spent on the task and participant reported grade point averages between the ARD intervention and the combined-aid intervention (which are both discussed in detail below), the demographics of participants across the four interventions were not found to be significantly different ($p < 0.10$), providing evidence that participants were randomly assigned to the four treatment conditions.

Twenty-four percent of the participants were MAcc students and 76 percent were accounting seniors.⁴⁸ On average, participants completed 5.5 accounting courses and 0.69 audit courses, with approximately 21 percent of participants reporting public accounting internship experience. The average size of each participant's working memory was 4.4 pieces of information, proxied as the number of historical phrases a participant was able to recall. On average, participants spent 39.7 minutes conducting the main task. There was a significant difference in the amount of time spent on the task between the no-aid intervention and the PCA intervention ($p=0.02$) and the No-Aid intervention and the combined-aid interventions ($p=0.00$). This difference is not unexpected since participants within PCA and combined-aid interventions were required

⁴⁷ Although data gathered from 78 participants provides the foundation from which this study's results are drawn, it should be noted that not every statistical test contains 78 observations due to the fact that useable participant responses were not always obtained. Thus, the actual number of useable observations may vary slightly between tests.

⁴⁸ There were no significant main effects or interactions for analytical review effectiveness based on class standing (MAcc or accounting senior) at $p < 0.10$.

to use the PCA's checkbox functionality to evaluate each of their proposed hypotheses before they could select the hypothesis they believe best explained why the cost of sales ratio declined. Finding that participants within these interventions took longer to complete the task provides evidence they attended to the decision aids and took the task

TABLE 4.1. Descriptive Statistics on Participant Exclusion.

Panel A: Participant Exclusion Descriptive Information					
	No-Aid	ARD	PCA	Combined-Aid	Total
Total Participants					129
Corrupted Data File					(1)
Accessible Participant Responses	31	31	33	33	128
Failed Manipulation Check Questions	(1)	(2)	(5)	(3)	(11)
Failed Accounting Knowledge Test	(7)	(9)	(9)	(10)	(35)
Reported No Decision Aid Reliance	N/A	(4)	0	0	(4)
Participants Used in Analysis	23	16	19	20	78

No significant difference in manipulation check failure rate or accounting knowledge failure rate between treatment conditions (p=0.437, Pearson chi-squared)

Panel B: Descriptive Information on Participants Who Failed the Manipulation Check Questions				
	Junior Accounting Students	Senior Accounting Students	Masters of Accountancy Students	Total
Accessible Participant Responses	5	93	30	128
Failed Manipulation Check	0	(9)	(2)	(11)
Proportion Who Failed	0.0%	9.7%	6.7%	8.6%

No significant difference in the manipulation check failure rate between masters of accountancy students and senior accounting students between treatment conditions (p=0.615, Pearson chi-squared)

Panel C: Descriptive Information on Participants Who Failed the Accounting Knowledge Test				
	Junior Accounting Students	Senior Accounting Students	Masters of Accountancy Students	Total
Participants Who Passed the Manipulation Check	5	84	28	117
Failed Accounting Knowledge Test	(5)	(23)	(7)	(35)
Proportion Who Failed the Knowledge Test	100.0%	27.3%	25.0%	29.9%

No significant difference in the accounting knowledge failure rate between masters of accountancy students and senior accounting students between treatment conditions (p=0.805, Pearson chi-squared)

seriously. To examine whether the amount of time taken on the task was related to analytical review effectiveness, task time was included as a covariate in each of the

logistic regression models used to test hypothesis generation, evaluation, and selection. However, task time was not found to be significant at $p < 0.10$.

Using a 6-point interval scale to capture participants' grade point averages (where One = grade point averages of 1.0 to 1.4; Two = 1.5 to 1.9; Three = 2.0 to 2.4; Four = 2.5 to 2.9; Five = 3.0 to 3.4; Six = 3.5 to 4.0), grade point average was found to be marginally different between conditions ($F = 2.348$, $df = 3$, $p = 0.079$). Results of a one-way ANOVA test revealed this difference was caused by a difference between ARD ($\bar{x}_{\text{Response_Scale}} = 5.13$) and combined-aid ($\bar{x}_{\text{Response_Scale}} = 4.50$) interventions ($F = 5.433$, $df = 1$, $p = 0.025$). There were no statistically significant GPA differences between the other intervention conditions. Because this study's hypotheses examine the analytical review effectiveness of participants within decision aid interventions to participants within the no-aid intervention, the GPA difference that exists between the participants in the ARD intervention and participants in the combined-aid intervention should have no impact on this study's results. To provide empirical evidence that GPA does not affect this study's results, I conducted sensitivity analysis by including GPA as a covariate in the statistical models used to test this study's hypotheses. I found that including GPA as a covariate had no effect on this study's results.

To measure participant responses to questions asked in the post-experiment questionnaire, I used a variety of 7-point Likert scales. Using a 7-point Likert scale (1 = Very Easy, 7 = Very Difficult), participants generally found the case material easy to understand ($\bar{x} = 2.1$) and found it somewhat easy to identify the pieces of information related to the unexpected fluctuation ($\bar{x} = 2.8$). Using a 7-point Likert scale (1 = Not at All Rushed, 7 = Completely Rushed), participants indicated they did not feel the task was

hurried ($\bar{x} = 2.1$). On a 7-point Likert scale (1 = Very Low, 7 = Very High), participants reported exerting a relatively high amount of mental effort ($\bar{x} = 4.9$). Finally, participant responses (1 = Not at All Important, 7 = Completely Important) indicate that it was of high importance for participants to find the correct answer to the task ($\bar{x} = 5.7$).⁴⁹ These demographics suggest that participants did not find the case confusing or rushed and also suggest that participants were diligent in conducting the task.

Additionally, the researcher closely and conspicuously observed participants during all experiment sessions to encourage them to conscientiously apply themselves during the task. Thus, the researcher's monitoring activity provides additional assurance that participants were diligent in conducting the task.

4.1.4 Participant Note-Taking Descriptive Statistics.

Participants in decision aid interventions were explicitly asked to identify and record the pieces of information they felt were relevant to the decline in the cost of sales ratio by typing them into the electronic software application used during the task. Every participant in the decision aid interventions complied with these instructions.⁵⁰

Like participants in decision aid interventions, participants in the no-aid intervention were also explicitly asked to identify the pieces of information they felt were relevant to the decline in the cost of sales ratio. To facilitate note-taking in the no-aid condition, participants were verbally told that they were free to take notes in the task application spreadsheet, in a separate electronic word processing document, or on any of the paper

⁴⁹ All participant response means were significantly different from the 7-point Likert scale neutral point of four at $p=0.000$.

⁵⁰ Although all decision aid participants complied with the note-taking instructions, not all of them necessarily successfully identified relevant pieces of information.

TABLE 4.2. Participant Descriptive Statistics.

Panel A: Participant Demographics by Treatment Condition					
	No-Aid	ARD	PCA	Combined-Aid	Overall
Total Participants	23	16	19	20	78
Masters of Accountancy Students	17.3% (n = 4)	43.8% (n = 7)	21.1% (n = 4)	25.0% (n = 4)	24.4% (n = 19)
Mean Number of Accounting Courses Taken (Standard Deviation)	5.5 (1.5)	5.8 (1.5)	5.4 (1.6)	5.4 (1.6)	5.5 (1.5)
Mean Number of Audit Courses Taken (Standard Deviation)	0.6 (0.7)	0.9 (0.7)	0.6 (0.6)	0.8 (0.7)	0.69 (0.7)
Public Accounting Internship Experience	21.7%	20.0%	15.7%	25.0%	21.0%
Mean Size of Working Memory (Standard Deviation)	4.2 (1.2)	4.5 (1.6)	4.4 (2.1)	4.6 (2.0)	4.4 (1.7)
Mean Time Spent on the Main Task in Minutes (Standard Deviation)	37.7 (3.4)	38.4 (3.8)	40.9 (5.1)	42.0 (4.9)	39.7 (4.7)*
Grade Point Average (Standard Deviation) Where one = 1.0 to 1.4, two = 1.5 to 1.9, three = 2.0 to 2.4, four = 2.5 to 2.9, five = 3.0 to 3.4, six = 3.5 to 4.0	4.65 (1.0)	5.13 (0.9)	5.00 (0.7)	4.50 (0.9)	4.8 (0.9)**

*Significant difference between the No-Aid intervention, PCA intervention, and Combined-Aid intervention (p=0.02 and p=0.00, respectively)

** Significant difference between the ARD intervention and the Combined-Aid intervention (p = 0.025)

Panel B: Participant Task Perceptions by Treatment Condition: Means and (Standard Deviation)					
	No-Aid	ARD	PCA	Combined-Aid	Overall
Ease or difficulty understanding case material (1 = Very Easy, 7 = Very Difficult)	2.1 (1.1)	2.0 (1.2)	2.1 (1.0)	2.0 (1.1)	2.1 (1.1)
Ease or difficulty in identifying the pieces of information relevant to the unexpected fluctuation (1 = Very Easy, 7 = Very Difficult)	2.8 (1.3)	3.3 (1.5)	2.6 (1.2)	2.5 (1.1)	2.8 (1.3)
How hurried was the pace of the task? (1 = Not at all Hurried, 7 = Completely Hurried)	2.0 (1.1)	2.3 (1.2)	1.9 (1.1)	2.3 (1.6)	2.1 (1.2)
Mental Effort Exerted (1 = Very Low, 7 = Very High)	4.7 (1.2)	5.1 (1.4)	4.9 (1.2)	4.7 (1.3)	4.9 (1.2)
How important was it to you to find the correct answer? (1 =Not at All Important, 7 = Completely Important)	5.8 (0.9)	6.0 (0.7)	5.6 (0.9)	5.6 (0.9)	5.7 (0.9)

materials given to them. To facilitate non-electronic note taking across all interventions, participants in every intervention were given a new pen they could use to take notes.

An ex-post examination of note taking among no-aid intervention participants was conducted for two reasons: (1) to determine the extent to which no-aid participants complied with the experimental instructions, and (2) to empirically test whether the decision to take notes is significantly related to this study's results. Consequently, two coders examined no-aid participants' paper case materials and electronic files to determine the extent of note-taking. Note-taking was coded at three levels: no notes, light notes, and heavy notes. "No notes" means that no participant markings were found, "light notes" means the participant made markings of some kind, and "heavy notes" means the participant made many markings and expressed developed thoughts. Eighty-seven percent of no-aid participants ($n = 20$) were found to take heavy notes, four percent took light notes ($n = 1$), and nine percent took no notes ($n = 2$). Two coders independently classified the level of note-taking and then reconciled any discrepancies in their classification. Prior to resolving discrepancies, the Cohen's Kappa coefficient, a measure of inter-rater reliability, was 0.805 for 23 items coded. The analytical review effectiveness between those who took notes and those who did not was not significant at $p < 0.10$. Additionally, a sensitivity analysis of this study's results was conducted by excluding participants who did not take notes from the statistical analysis. Excluding participants who did not take notes had no effect on the results.⁵¹ Thus, there is empirical evidence that note-taking did not drive this study's results. Table 4.3 presents descriptive statistics regarding no-aid intervention note-taking.

⁵¹ I also conducted another sensitivity analysis that excluded the participants who took no notes or light notes ($n=3$). Excluding these participants had no significant effect on this study's results.

4.1.5 Decision-Aid Reliance and Usefulness Descriptive Statistics.

Descriptive statistics regarding decision aid reliance indicate that participants placed reliance upon the decision aids in conducting the task. Table 4.4 presents decision aid reliance descriptive statistics. Using a 7-point Likert scale to measure decision aid

TABLE 4.3. Ex-Post Note-Taking Analysis Among No-Aid Intervention Participants.

Participant	Intervention	Level of Note-Taking
Rocky3	No-Aid	Heavy Notes
Rocky4	No-Aid	Light Notes
Rocky5	No-Aid	Heavy Notes
Rocky6	No-Aid	Heavy Notes
Rocky7	No-Aid	Heavy Notes
Rocky8	No-Aid	Heavy Notes
Rocky9	No-Aid	Heavy Notes
Rocky10	No-Aid	No Notes
Rocky11	No-Aid	Heavy Notes
Rocky12	No-Aid	Heavy Notes
Rocky13	No-Aid	Heavy Notes
Rocky33	No-Aid	Heavy Notes
Rocky34	No-Aid	Heavy Notes
Rocky51	No-Aid	Heavy Notes
Rocky52	No-Aid	Heavy Notes
Rocky55	No-Aid	Heavy Notes
Rocky56	No-Aid	No Notes
Rocky57	No-Aid	Heavy Notes
Rocky61	No-Aid	Heavy Notes
Rocky122	No-Aid	Heavy Notes
Rocky127	No-Aid	Heavy Notes
Rocky129	No-Aid	Heavy Notes
Rocky130	No-Aid	Heavy Notes

Level of note-taking was not significantly related to analytical review effectiveness at $p < 0.10$. Excluding No-Aid intervention participants who did not take notes has no effect on this study's results. Further, excluding No-Aid intervention participants who took either no notes or light notes has no effect on this study's results.

reliance (1 = Not at all, 7 = A Great Deal) during the task, both ARD and PCA users

reported placing reliance on the aids during the task ($\bar{x}_{ARD \text{ reliance}} = 3.9$, $\bar{x}_{PCA \text{ reliance}} = 5.3$).⁵²

⁵² Four participants reported placing “no reliance” on the decision aids. These four participants were in the ARD intervention and were removed from my statistical analysis because their task performance cannot be appropriately attributed to the ARD. Of the four, one participant proposed the correct hypothesis and subsequently evaluated and selected it as the cause of the unexpected fluctuation.

Users of both the ARD and PCA reported finding the aids as helpful during hypothesis generation, hypothesis evaluation, and hypothesis selection ($\bar{x}_{ARD} = 3.7$, $\bar{x}_{PCA} = 4.9$).⁵³ To test whether the degree to which a participant relied upon a decision aid was significantly related to task effectiveness, two LOGIT models were estimated (one model for ARD users and one model for PCA users) regressing decision aid reliance upon whether a participant selected the correct cause of the unexpected fluctuation. Decision aid reliance

TABLE 4.4. Decision Aid Reliance and Usefulness Descriptive Statistics.

Decision Aid Reliance and Usefulness: Means and (Standard Deviation)		
Reliance Question (1 = Not at All, 4 = Somewhat, 7 = A Great Deal)	ARD Decision Aid ⁺	PCA Decision Aid [^]
How much did you rely on the decision aid to conduct the task?	3.9 (1.7) n=34 ⁺⁺	5.3 (1.5) n=37 ^{^^}
How much did the decision aid help you to generate potential reasons to explain why the cost of sales ratio declined?	4.1 (1.9) n=34	5.3 (1.4) n=37
How much did the decision aid help you to evaluate each of your proposed reasons?	3.5 (1.7) n=34	4.6 (1.8) n=37
How much did the decision aid help you to select the best reason the cost of sales ratio declined?	3.6 (1.9) n=34	4.4 (1.9) n=37
Overall, how helpful was the decision aid?	4.2 (2.0) n=34	5.1 (1.5) n=37

⁺Reliability analysis of the five-question “ARD reliance and usefulness” 7-point Likert scale indicated high reliability (Cronbach’s alpha = 0.963)

⁺⁺The sample size reflects participant responses from both ARD interventions and combined-aid interventions since the ARD was made available to participants in both intervention conditions

[^]Reliability analysis of the six-question “PCA reliance and usefulness” 7-point Likert scale indicated high reliability (Cronbach’s alpha = 0.913)

^{^^}The sample size reflects participant responses from both PCA interventions and combined-aid interventions since the PCA was made available to participants in both intervention conditions

⁵³ These means reflect participants’ responses to a multiple question (five questions for ARD users, six questions for PCA users), 7-point Likert scale reliance and usefulness questionnaire administered at the end of the experiment. Chronbach’s alpha for these questions and responses were 0.963 for ARD users and 0.913 for PCA users.

was not found to be significantly related to analytical review effectiveness for ARD users ($\chi^2 = 0.023$, $df=1$, $p=0.628$) or for PCA users ($\chi^2 = 0.083$, $df=1$, $p=0.773$). In sum, there is evidence that participants within decision aid interventions placed reliance upon the aids and found them to be helpful in conducting analytical review.

4.1.6 Participant Proposed Hypotheses.

The purpose of conducting analytical review is to identify the cause of an unexpected fluctuation. Consequently, to properly perform analytical review each participant must generate at least one potential reason to explain why the unexpected fluctuation occurred. Within this study, each participant generated at least one potential reason to explain the cause of the unexpected fluctuation.⁵⁴

4.2 Assumption Testing

4.2.1 Univariate Analysis of Variance (ANOVA).

Before hypotheses testing, I first investigate whether any of the statistical procedures I plan to use violate their related assumptions. ANOVA makes several assumptions regarding the sampling units employed and the overall sample distribution. ANOVA assumes that units are randomly sampled from the population of interest, that observations are statistically independent of each other, that dependent variables are normally distributed, and that within-intervention variances are homogenous.

Each participant self-selected the experimental session they participated in. Thus, I did not have control over the distribution of participants to any specific experimental session. Further, to promote random assignment and guard against introducing bias into a specific experimental session, the intervention treatment administered for a given

⁵⁴ The mean number (standard deviation) of hypotheses generated by intervention condition was as follows: $\bar{x}_{No-Aid} = 4.26$ (1.54), $\bar{x}_{ARD} = 3.31$ (2.02), $\bar{x}_{PCA} = 4.00$ (1.86), $\bar{x}_{Combined-aids} = 3.70$ (1.72). There was no significant difference in the number of hypotheses proposed by intervention condition at $p < 0.10$.

session was determined by randomly selecting one of four numbered slips of paper from a hat. An empirical investigation of participant demographics between treatment conditions (previously discussed above in detail in the section sub-titled “Participant Descriptive Statistics”) reveals that all interventions were not significantly different from each other in terms of the proportion of MAcc students participating, the number of accounting courses completed by participants, the number of audit courses completed by participants, the working memory size of participants, and participants’ public accounting experience.

The assumption of statistical independence of observations is satisfied by using a between-subjects experiment design. Normality of the distribution of the dependent variable “Number of plausible hypotheses proposed” was examined by creating a scatterplot that depicted the number of plausible hypotheses proposed by participant. There appeared to be no extreme deviations from normality for this dependent variable. To test whether within-intervention variances were homogenous, a Levene’s test was conducted and test results indicated that the homogeneity assumption was not violated ($F = 0.598$, $df = 3$, $p = 0.619$).

4.2.2 Logistic Regression (LOGIT).

LOGIT is a statistical test used to estimate the odds of a discrete outcome. LOGIT assumes that the dependent variable is discrete and that observations are independent. The dependent variables I propose examining with LOGIT are (1) whether the correct hypothesis is proposed, (2) whether the correct hypothesis is selected, and (3) in cases where the correct hypothesis is proposed, whether the correct hypothesis is selected. These three dependent variables are dichotomous in nature, thus they are

discrete. The assumption of observation statistical independence is satisfied by using a between-subjects experiment design.

4.3 Hypotheses Testing

Hypothesis 1a (H1a) predicts that participants in the ARD intervention will generate more plausible hypotheses compared to those in the no-aid intervention. Contrary to my expectations, there is no statistical difference in the mean number of plausible hypotheses proposed by participants in the ARD intervention ($\bar{x}_{ARD} = 1.13$) compared to participants in the no-aid intervention ($\bar{x}_{No-Aid} = 0.91$) as reported in Panel A in Table 4.5 ($F = 0.521$, $df = 3$, $p=0.669$). Further, a significant difference was not found using either Dunnett’s post hoc analysis or by planned contrast testing. Thus, H1a is not supported.

TABLE 4.5. Number of Plausible Hypotheses Proposed by Intervention.

Panel A: Treatment Descriptive Statistics				
Intervention	Range (# of Plausible Hypotheses Proposed)	Mean	Standard Deviation	N
No-Aid	0 – 5	0.91	1.345	23
ARD	0 – 4	1.13	1.31	16
PCA	0 – 5	1.21	1.398	19
Combined-Aid	0 – 3	0.75	0.967	20

Dependent Variable: The average number of plausible hypotheses proposed by each participant

Panel B: ANOVA Model and Significance					
Analysis of variance source of variation	Sum of Squares	Degrees of Freedom	Mean Square	F-stat	F-Significance
Model	2.503	3	0.834	0.521	0.669
Intercept	76.643	1	76.643	47.868	0.000
Intervention	2.503	3	0.834	0.521	0.669
Error	118.484	74	1.610		
Total	197.000	78			

Hypothesis 1b (H1b) proposes that participants in the ARD intervention will propose the correct hypothesis more often compared to participants in the no-aid

intervention. Panel A in Table 4.6 presents descriptive statistics regarding the number of times the correct hypothesis was proposed. Seven out of 16 participants (43.8 percent) in the ARD intervention proposed the correct hypothesis, while seven out of 23 participants (30.4 percent) proposed the correct hypothesis in the no-aid intervention. To put these percentages in perspective with prior research, analytical review studies find between nine and twenty-nine percent of experienced auditors propose the correct hypothesis (Bedard, Biggs, and Maroney 1998; Bedard and Biggs 1991, respectively). Panel B in Table 4.6 presents the results of a LOGIT model that regresses whether a participant received the ARD against whether a participant proposed the correct hypothesis, controlling for cognitive load, the number of non-audit accounting courses completed, and the number of audit courses completed. As the results in Panel C of Table 4.6 indicate, the odds a participant proposes the correct hypothesis increase by a multiplicative factor of 3.834 when a participant uses the ARD as compared to conditions where a participant is not given any decision aid ($\chi^2 = 2.383$, $df = 1$, $p=0.062$, one-tailed). Thus, there is evidence that the ARD helps to improve participant hypothesis generation effectiveness, in support of hypothesis H1b.

Results indicate that participants in the ARD intervention are more effective at hypotheses evaluation than those who do not receive any decision aid. Participants were able to identify the correct hypothesis six out of the seven times (85.7 percent) it was proposed in the ARD intervention, as compared to three out of seven times (42.8 percent) in the no-aid intervention. Panel A in Table 4.7 presents descriptive statistics regarding the number of times participants were able to identify the correct hypothesis. Using instances where a participant proposed the correct hypothesis, a LOGIT model was

estimated regressing whether a participant received the ARD against whether a participant selected the correct hypothesis, controlling for cognitive load. As the results

TABLE 4.6. Hypothesis 1b: Hypothesis Generation Effectiveness (ARD vs. No-Aid).

Panel A: Hypothesis Generation Effectiveness Descriptive Statistics by Intervention			
Intervention	N	Number of Times Correct Hypothesis Proposed	Percentage of Times Correct Hypothesis Proposed
No-Aid	23	7	30.4%
ARD	16	7	43.8%

Panel B: LOGIT Model and Significance			
LOGIT	Overall Model χ^2	Degrees of Freedom	Significance
Model	12.358	4	0.015

Panel C: LOGIT Model Detailed Statistics						
	B	Exp(B)	Standard Error	Wald χ^2	df	Significance
Decision Aid (0 = No-Aid, 1 = ARD)	1.344	3.834	0.871	2.383	1	0.062*
Covariate: Cognitive Load (1 = Very Low, 7 = Very High)	-0.735	0.480	0.281	6.812	1	0.009**
Covariate: Number of Non-Audit Accounting Courses Completed	2.644	14.071	1.634	2.617	1	0.106
Covariate: Number of Audit Courses Completed	-1.525 [†]	0.218	0.968	2.483	1	0.115
Constant	-2.881	0.056	3.570	0.651	1	0.165

Dependent Variable: The proportion of times the correct hypothesis was proposed

*Significant at $p < 0.10$, one-tailed

**Significant at $p < 0.05$

[†]The non-intuitive coefficient direction is due to overconfidence and is discussed in the post-hoc analysis

presented in Panel B of Table 4.7 reveal, the odds that a participant discerns the actual cause of the unexpected fluctuation increase when the participant uses the ARD as compared to participants in the no-aid intervention ($\chi^2=2.250$, $df = 1$, $p=0.067$, one-tailed). Results indicate that the odds the participant will discern the correct hypothesis increase by a multiplicative factor of 160.99 if the ARD is used, providing support for H2.

Hypothesis 3 (H3) proposes that the correct hypothesis will be selected more often by participants in the ARD intervention when compared to participants in the no-

aid intervention. Six out of 16 participants (37.5 percent) selected the correct hypothesis in the ARD intervention, as compared to three out of 23 participants (13.0 percent) in the no-aid intervention. Panel A in Table 4.8 presents descriptive statistics regarding the

TABLE 4.7. Hypothesis 2: Hypotheses Evaluation and Effectiveness (ARD vs. No-Aid).

Panel A: Hypothesis Evaluation Effectiveness Descriptive Statistics by Intervention			
Intervention	Number of Times Correct Hypothesis Proposed	Number of Times Correct Hypothesis Identified	Percentage of Times Correct Hypothesis Identified When Proposed
No-Aid	7	3	42.8%
ARD	7	6	85.7%

Panel B: LOGIT Model and Significance			
LOGIT	Overall Model χ^2	Degrees of Freedom	Significance
Model	11.420	2	0.003

Panel C: LOGIT Model Detailed Statistics						
	B	Exp(B)	Standard Error	Wald χ^2	df	Significance
Decision Aid (0 = No-Aid, 1 = ARD)	5.081	160.999	3.388	2.250	1	0.067*
Covariate: Cognitive Load (1 = Very Low, 7 = Very High)	-1.964	0.140	1.033	3.612	1	0.057**
Constant	4.388	80.447	2.682	2.677	1	0.102

Dependent Variable: The proportion of times the correct hypothesis was selected when it was proposed

*Significant at $p < 0.10$, one-tailed

**Significant at $p < 0.10$

number of participants that selected the correct hypothesis. To test H3, a LOGIT model was estimated regressing whether a participant received the ARD against whether a participant selected the correct hypothesis, controlling for cognitive load. Consistent with H3, the results presented in Panels B and C of Table 4.8 indicate that the odds a participant selects the correct hypothesis increase by a multiplicative factor of 5.464 when the ARD is used ($\chi^2=3.167$, $df = 1$, $p=0.038$, one-tailed). Thus, there is evidence that the ARD improves overall analytical review effectiveness, supporting H3.

TABLE 4.8. Hypothesis 3: Hypothesis Selection Effectiveness (ARD vs. No-ARD).

Panel A: Hypothesis Selection Effectiveness Descriptive Statistics by Intervention			
Intervention	N	Number of Times Correct Hypothesis Selected	Percentage of Times Correct Hypothesis Selected
No-Aid	23	3	13.0%
ARD	16	6	37.5%

Panel B: LOGIT Model and Significance			
LOGIT	Overall Model χ^2	Degrees of Freedom	Significance
Model	12.568	2	0.002

Panel C: LOGIT Model Detailed Statistics						
	B	Exp(B)	Standard Error	Wald χ^2	df	Significance
Decision Aid (0 = No-Aid, 1 = ARD)	1.698	5.464	0.954	3.167	1	0.038*
Covariate: Cognitive Load (1 = Very Low, 7 = Very High)	-0.906	0.404	0.376	5.792	1	0.016**
Constant	0.652	1.920	1.087	0.360	1	0.548

Dependent Variable: The proportion of times participants selected the correct hypothesis

*Significant at $p < 0.10$, one-tailed

**Significant at $p < 0.10$

Hypothesis 4a (H4a) proposes that participants in the PCA intervention will generate more plausible hypotheses compared to participants in the no-aid intervention. Contrary to my expectations, there is no statistical difference in the mean number of plausible hypotheses proposed by participants in the PCA intervention ($\bar{x}_{PCA} = 1.21$) compared to those generated by participants in the no-aid intervention ($\bar{x}_{No-Aid} = 0.91$) as reported in Panel A in Table 4.5 ($F = 0.521$, $df = 1$, $p = 0.669$). Further, a significant difference was not found using either Dunnett's post hoc analysis or by planned contrast testing. Thus, H4a is not supported.

Hypothesis 4b (H4b) proposes that participants in the PCA intervention will propose the correct hypothesis more often compared to participants in the no-aid intervention. Seven out of 19 (36.8 percent) participants in the PCA intervention proposed the correct hypothesis, while seven out of 23 participants (30.4 percent)

proposed the correct hypothesis in the no-aid intervention. These results are consistent with prior research which finds that between nine and twenty-nine percent of experienced auditors propose the correct hypothesis (Bedard, Biggs, and Maroney 1998; Bedard and Biggs 1991, respectively). Although the proportion of participants that proposed the correct hypothesis is higher in the PCA intervention, the results are not statistically significant. Panel B in Table 4.9 presents the results of a LOGIT model which was estimated by regressing whether a participant received the PCA against whether a participant proposed the correct hypothesis, controlling for cognitive load. As the results in Panel B of Table 4.9 indicate, neither the overall regression model ($\chi^2=3.635$, $df = 2$, $p=0.162$) or the PCA intervention were significant ($\chi^2=0.582$, $df = 1$, $p=0.223$, one-tailed). Thus, the PCA is not found to significantly increase the odds that a participant will propose the correct hypothesis as compared to participants in the no-aid intervention. Consequently, there is no evidence the PCA helps improve participants' hypothesis generation effectiveness. Therefore, H4b is not supported.

Results reported in Table 4.10 indicate that participants in the PCA intervention are more effective at hypotheses evaluation than those in the no-aid intervention. Participants were able to identify the correct hypothesis six out of the seven times (85.7 percent) it was proposed in the PCA intervention, as compared to three out of seven times (42.8 percent) in the no-aid intervention. Panel A in Table 4.10 presents descriptive statistics regarding the number of times participants were able to identify the correct hypothesis. Using cases where a participant proposed the correct hypothesis, a LOGIT

TABLE 4.9. Hypothesis 4b: Hypothesis Generation Effectiveness (PCA vs. No-Aid).

Panel A: Hypothesis Generation Effectiveness Descriptive Statistics by Intervention			
Intervention	N	Number of Times Correct Hypothesis Proposed	Percentage of Times Correct Hypothesis Proposed
No-Aid	23	7	30.4%
PCA	19	7	36.8%

Panel B: LOGIT Model and Significance			
LOGIT Model	Overall Model χ^2	Degrees of Freedom	Significance
Model	3.635	2	0.162

Panel C: LOGIT Model Detailed Statistics							
	B	Exp(B)	Standard Error	Wald χ^2	df	Significance	
Decision Aid (0 = No-Aid, 1 = PCA)	0.536	1.709	0.702	0.582	1	0.223	
Covariate: Cognitive Load (1 = Very Low, 7 = Very High)	-0.380	0.684	0.218	3.052	1	0.081*	
Constant	0.562	1.754	0.892	0.397	1	0.529	

Dependent Variable: The proportion of times the correct hypothesis was proposed

*Significant at $p < 0.10$

model was estimated regressing whether a participant received the PCA against whether a participant discerned the correct hypothesis, controlling for cognitive load. Consistent with expectations, the results presented in Panel B of Table 4.10 indicate the odds that a participant is able to discern the correct hypothesis increase when the participant uses the PCA as compared to participants in the no-aid intervention ($\chi^2=2.850$, $df = 1$, $p=0.046$, one-tailed). The odds that a participant discerns the correct hypothesis increase by a multiplicative factor of 100.09 if the PCA is used, providing support for H5.

Hypothesis 6 (H6) proposes that the correct hypothesis will be selected more frequently by participants in the PCA intervention when compared to participants in the no-aid intervention. Six out of 19 participants (31.6 percent) selected the correct hypothesis in the PCA intervention, as compared to three out of 23 participants (13.0 percent) in the no-aid intervention.

TABLE 4.10. Hypothesis 5: Hypothesis Evaluation Effectiveness (PCA vs. No-Aid).

Panel A: Hypothesis Evaluation Effectiveness Descriptive Statistics by Intervention			
Intervention	Number of Times Correct Hypothesis Proposed	Number of Times Correct Hypothesis Identified	Percentage of Times Correct Hypothesis Identified When Proposed
No-Aid	7	3	42.8%
PCA	7	6	85.7%

Panel B: LOGIT Model and Significance			
LOGIT	Overall Model χ^2	Degrees of Freedom	Significance
Model	6.783	2	0.034

Panel C: LOGIT Model Detailed Statistics						
	B	Exp(B)	Standard Error	Wald χ^2	df	Significance
Decision Aid (0 = No-Aid, 1 = PCA)	4.606	100.095	2.728	2.850	1	0.046*
Covariate: Cognitive Load (1 = Very Low, 7 = Very High)	-1.171	0.310	0.792	2.187	1	0.139
Constant	2.539	12.667	2.045	1.541	1	0.214

Dependent Variable: The proportion of times the correct hypothesis was selected when it was proposed

*Significant at $p < 0.10$, one-tailed

regarding the number of participants that selected the correct hypothesis. To test H6, a LOGIT model was estimated regressing whether a participant received the PCA against whether a participant selected the correct hypothesis, controlling for cognitive load, the number of audit courses completed, and the number of non-audit accounting courses completed. Consistent with H6, results in Table 4.11 indicate that the odds a participant selects the correct hypothesis significantly increase by a multiplicative factor of 10.06 when the PCA is used ($\chi^2=4.004$, $p=0.023$, one-tailed). Thus, there is evidence that the PCA improves overall analytical review effectiveness, supporting H6.

Hypothesis 7a (H7a) proposes that participants in the combined-aid intervention condition will generate more plausible hypotheses compared with participants in all other conditions. However, ANOVA results indicate that there is no statistical difference in the

TABLE 4.11. Hypothesis 6: Hypothesis Selection Effectiveness (PCA vs. No-Aid).

Panel A: Hypothesis Selection Effectiveness Descriptive Statistics by Intervention			
Intervention	N	Number of Times Correct Hypothesis Selected	Percentage of Times Correct Hypothesis Selected
No-Aid	23	3	13.0%
PCA	19	6	31.6%

Panel B: LOGIT Model and Significance			
LOGIT Model	Overall Model χ^2	Degrees of Freedom	Significance
Model	17.116	4	0.002

Panel C: LOGIT Model Detailed Statistics						
	B	Exp(B)	Standard Error	Wald χ^2	df	Significance
Decision Aid (0 = No-Aid, 1 = PCA)	2.309	10.066	1.154	4.004	1	0.023*
Covariate: Cognitive Load (1 = Very Low, 7 = Very High)	-0.721	0.486	0.314	5.264	1	0.022**
Covariate: Number of Audit Courses Completed	-4.372	0.013	1.797	5.919	1	0.015***
Covariate: Number of Non- Audit Accounting Courses Completed	1.780	5.931	0.918	3.757	1	0.053**
Constant	-6.531	0.001	3.935	2.755	1	0.097**

Dependent Variable: The proportion of times the correct hypothesis was selected when it was proposed

*Significant at $p < 0.05$, one-tailed

**Significant at $p < 0.10$

***Significant at $p < 0.10$. The non-intuitive coefficient direction is due to participant overconfidence and is discussed in the post-hoc analysis

mean number of plausible hypotheses proposed by participants in the combined-aid intervention ($\bar{x}_{\text{Combined-Aid}} = 0.75$) compared to those generated by participants in the no-aid intervention ($\bar{x}_{\text{No-Aid}} = 0.91$) as reported in Panel A in Table 4.5 ($F = 0.521$, $df = 3$, $p = 0.669$). Further, because a visual examination of the descriptive statistics regarding the number of plausible hypotheses generated by combined-aid users reveals that participants in this condition actually proposed the least average number of plausible hypotheses, no further statistical analysis was conducted ($\bar{x}_{\text{Combined-Aid}} = 0.75$, $\bar{x}_{\text{ARD}} = 1.13$, $\bar{x}_{\text{PCA}} = 1.21$). Consequently there is no support for H7a.

Hypothesis 7b (H7b) proposes that participants in the combined-aid intervention condition will generate the correct hypothesis more often compared to participants in all

other conditions. A simple visual examination of the descriptive statistics presented in Panel A in Table 4.12 reveals that participants in the combined-aid intervention did not propose the correct hypothesis more often than any other intervention. Five out of 20 participants (25.0 percent) in the combined-aid intervention proposed the correct hypothesis as compared to seven out of 23 participants (30.4 percent) in the no-aid intervention, seven out of 16 participants (43.8 percent) in the ARD intervention, and seven out of 19 participants (36.8 percent) in the PCA intervention. Thus, H7b is not supported.

Hypothesis 8 (H8) proposes that when the correct hypothesis is proposed, it will be identified more often by participants in the combined-aid intervention when compared to participants in all other interventions. A visual examination of the descriptive statistics presented in Panel A in Table 4.13 reveals that participants in the combined-aid intervention did not identify the correct hypothesis more often than all other interventions, as participants in both the ARD intervention and the PCA intervention identified the correct hypothesis more often, providing no support for H8 (Percent Identified_{Combined-Aid} = 80.0 percent, Percent Identified_{ARD} = 85.7 percent, Percent Identified_{PCA} = 85.7 percent). However, results reported in Table 4.13 indicate that combined-aid users are more effective at hypotheses evaluation than those in the no-aid intervention. When a participant proposes the correct hypothesis during the task, the odds that the participant is able to discern it as the actual cause of the unexpected fluctuation increase by a multiplicative factor of 34.46 when the participant uses both decision aids as compared to when the participant is provided with no decision aids

TABLE 4.12. Descriptive Statistics: Hypothesis Generation, Evaluation, and Selection.

Panel A: Hypothesis Generation Descriptive Statistics			
Intervention	N	Number of Times Correct Hypothesis Proposed	Percentage of Times Correct Hypothesis Proposed
No-Aid	23	7	30.4%
ARD	16	7	43.8%
PCA	19	7	36.8%
Combined-Aid	20	5	25.0%

Panel B: Hypothesis Evaluation Descriptive Statistics			
Intervention	Number of Times Correct Hypothesis Proposed	Number of Times Correct Hypothesis Identified	Percentage of Times Correct Hypothesis Identified When Proposed
No-Aid	7	3	42.8%
ARD	7	6	85.7%
PCA	7	6	85.7%
Combined-Aid	5	4	80.0%

Panel C: Hypothesis Selection Descriptive Statistics			
Intervention	N	Number of Times Correct Hypothesis Selected	Percentage of Times Correct Hypothesis Selected
No-Aid	23	3	13.0%
ARD	16	6	37.5%
PCA	19	6	31.6%
Combined-Aid	20	4	20.0%

($\chi^2=2.590$, $df=1$, $p=0.054$, one-tailed). When the correct hypothesis was proposed, participants in the combined-aid intervention were able to identify it four out of the five times (80.0 percent) as compared to three out of seven times (42.8 percent) in the no-aid intervention. Thus, participants using both decision aids performed hypothesis evaluation more effectively than participants who did not receive any decision aids.

Hypothesis 9 (H9) proposes that the correct hypothesis will be selected more often by participants in the combined-aid intervention when compared to participants

TABLE 4.13. Hypothesis Evaluation Effectiveness (Combined-Aid vs. No-Aid).

Panel A: Hypothesis Evaluation Effectiveness Descriptive Statistics by Intervention			
Intervention	Number of Times Correct Hypothesis Proposed	Number of Times Correct Hypothesis Identified	Percentage of Times Correct Hypothesis Identified When Proposed
No-Aid	7	3	42.8%
Combined-Aid	5	4	80.0%
ARD	7	6	85.7%
PCA	7	6	85.7%

Panel B: LOGIT Model and Significance: Combined-Aid vs. No-Aid			
LOGIT Model	χ^2	Degrees of Freedom	Significance
	7.737	2	0.021

Panel C: LOGIT Model Detailed Statistics: Combined-Aid vs. No-Aid						
	B	Exp(B)	Standard Error	Wald χ^2	df	Significance
Decision Aid (0 = No-Aid, 1 = Combined-Aid)	3.540	34.459	2.199	2.590	1	0.054*
Covariate: Cognitive Load (1 = Very Low, 7 = Very High)	-2.247	0.106	1.341	2.805	1	0.094**
Constant	5.063	158.000	3.428	2.181	1	0.140

Dependent Variable = The proportion of times the correct hypothesis was selected when it was proposed

*Significant at $p < 0.10$, one-tailed

**Significant at $p < 0.10$

within all other interventions. A visual examination of the descriptive statistics presented in Panel A in Table 4.14 reveals that participants in the combined-aid intervention did not select the correct hypothesis more often than all other interventions, as participants in both the ARD intervention and the PCA intervention identified the correct hypothesis more often (Proportion Selected_{Combined-Aid} = 20.0 percent, Proportion Selected_{ARD} = 37.5 percent, Proportion Selected_{PCA} = 31.6 percent). Thus, H9 is not supported. Further, the results presented in Panel B of Table 4.14 indicate that there was no significant hypothesis selection effectiveness difference between participants in the combined-aid intervention and participants in the no-aid intervention ($\chi^2=1.535$, $df = 1$, $p=0.108$).

TABLE 4.14. Hypothesis 9: Hypothesis Selection Effectiveness (Combined-Aid vs. No-Aid).

Panel A: Hypothesis Selection Effectiveness Descriptive Statistics by Intervention			
Intervention	N	Number of Times Correct Hypothesis Selected	Percentage of Times Correct Hypothesis Selected
No-Aid	23	3	13.0%
Combined-Aid	20	4	20.0%
ARD	16	6	37.5%
PCA	19	6	31.6%

Panel B: LOGIT Model and Significance: Combined-Aid vs. No-Aid			
LOGIT	Overall Model χ^2	Degrees of Freedom	Significance
Model	12.658	4	0.013

Panel C: LOGIT Model Detailed Statistics: Combined-Aid vs. No-Aid						
	B	Exp(B)	Standard Error	Wald χ^2	df	Significance
Decision Aid (0 = No-Aid, 1 = Combined-Aid)	1.412	4.104	1.140	1.535	1	0.108
Covariate: Cognitive Load (1 = Very Low, 7 = Very High)	-1.178	0.308	0.544	4.683	1	0.031*
Covariate: Number of Audit Courses Completed	-2.878	0.056	1.616	4.683	1	0.075**
Covariate: Number of Non-Audit Accounting Courses Completed	1.367	3.924	0.771	3.142	1	0.076*
Constant	-4.779	0.008	3.657	1.708	1	0.191

Dependent Variable: The proportion of times the correct hypothesis was selected

*Significant at $p < 0.10$

**Significant at $p < 0.10$. The non-intuitive coefficient direction is due to participant overconfidence and is discussed in the post-hoc analysis

In summary, the study's results indicate that use of the decision aids investigated in this study significantly improve participants' analytical review performance during hypotheses evaluation and hypothesis selection, leading to better overall analytical review effectiveness as compared to participants who are not provided with any decision aids. Further, although participants within the ARD intervention proposed the correct hypothesis significantly more often than the no-aid intervention, no significant difference was found among PCA or combined-aid interventions. Results indicate that decision aid users do not generate more plausible hypotheses than non-aid users. Finally, results indicate that although the analytical review effectiveness of participants within the combined-aid intervention is sometimes significantly better than that of the no-aid

intervention, participants within the combined-aid intervention do not outperform participants within the ARD intervention or the PCA intervention. Table 4.15 presents the study's hypothesis and identifies whether each was supported.

TABLE 4.15. Summary of Hypotheses and Results.

	Hypothesis	Supported?
H1a	A larger number of plausible hypotheses will be proposed by inexperienced auditors using the activity relationship diagram aid than by inexperienced auditors not using any aid.	No.
H1b	The correct hypothesis will be proposed more often by inexperienced auditors using the activity relationship diagram aid than by inexperienced auditors not using any aid.	Yes.
H2	When the hypothesis set contains the correct hypothesis, inexperienced auditors using the activity relationship diagram aid will identify the correct hypothesis more often when compared to inexperienced auditors not using any aid.	Yes.
H3	The correct hypothesis will be selected more often by inexperienced auditors in the activity relationship diagram intervention when compared to inexperienced auditors in the no-aid intervention.	Yes.
H4a	A larger number of plausible hypotheses will be proposed by inexperienced auditors using the pattern-consideration aid than by inexperienced auditors not using any aid.	No.
H4b	The correct hypothesis will be proposed more often by inexperienced auditors using the pattern-consideration aid than by inexperienced auditors not using any aid.	No.
H5	When the hypothesis set contains the correct hypothesis, inexperienced auditors using the pattern-consideration aid will identify the correct hypothesis more often when compared to inexperienced auditors not using any aid.	Yes.
H6	The correct hypothesis will be selected more often by inexperienced auditors using the pattern-consideration aid when compared to inexperienced auditors not using any aid.	Yes.
H7a	A larger number of plausible hypotheses will be proposed by inexperienced auditors using both aids than by inexperienced auditors using each aid individually or no aid.	No.
H7b	The correct hypothesis will be proposed more often by inexperienced auditors using both aids than by inexperienced auditors using each aid individually or no aid.	No.
H8	When the hypothesis set contains the correct hypothesis, inexperienced auditors using both aids will identify the correct hypothesis more often than inexperienced auditors using each aid individually or no aid.	No.
H9	The correct hypothesis will be selected more often by inexperienced auditors using both aids than by inexperienced auditors using each aid individually or no aid.	No.

4.4 Planned Mediation Analysis

I conducted mediation analysis to empirically test my prediction that the level of cognitive load mediates the relationship between decision aid use and analytical review effectiveness. Applying the Baron and Kenny (1986) mediation model I examine whether the three conditions necessary for mediation exist within this study: (1) the independent variable should significantly impact the mediator variable, (2) changes in the mediator variable should significantly impact the dependent variable, and (3) a previously significant relationship between the independent and dependent variable should decrease in significance when the mediator is introduced. I present these three conditions in equation form below:

$$\text{Equation 1: } Y = i_1 + cX + e_1$$

$$\text{Equation 2: } Y = i_2 + c'X + bM + e_2$$

$$\text{Equation 3: } M = i_3 + aX + e_3$$

In the equations above Y serves as the dependent variable, X is the independent variable, M is the mediator variable. i_1 and i_2 and i_3 are the regression intercepts, c is the coefficient relating the independent variable and the dependent variable, c' is the coefficient relating the independent variable to the dependent variable, which is adjusted for the mediator variable, b is the coefficient relating the mediator to the dependent variable, which is adjusted for the independent variable, a is the coefficient relating the independent variable to the mediator, and e_1 and e_2 and e_3 are the residuals (MacKinnon et al. 2007). Applying the equations to my study, Y is analytical review effectiveness (measured by each of the three dependent variables used to test my hypothesis), X is the intervention (i.e., the ARD, PCA, or combined-Aid interventions), and M is cognitive load as reported by study participants using a 7-point Likert scale administered during the

post-experimental questionnaire. The coefficients represent the values obtained by estimating the regression.

4.4.1 Measuring Cognitive Load.

The question of how to measure cognitive load has proven difficult for researchers to solve (Paas et al. 2003). Researchers generally use two methods to measure cognitive load: analytical methods and empirical methods (Paas et al. 2003). Analytical methods seek to estimate cognitive load through the use of mathematical models and task analysis. Analytical methods are not widely used to measure cognitive load, as Paas et al. (2003) report that only one study has applied this method (p. 66). Thus, I decided not to measure cognitive load using an analytical method.

Empirical methods attempt to measure cognitive load using psychophysiological techniques and ratings scales (Paas et al. 2003). Psychophysiological techniques are based on the presumption that changes in cognitive functioning are reflected in individuals' psychological responses. Some common psychophysiological variables of interest are heart activity, brain activity, and eye activity (Paas et al. 2003). Because I did not have access to instruments capable of measuring physiological variables, I decided not to measure cognitive load in this way.

This study uses a 7-point Likert rating scale to measure participants' perceived level of cognitive load. Measuring cognitive load through the use of a rating scale is based on the assumption that individuals are able to accurately report the amount of cognitive load they experience. While the use of self-rating scales may appear questionable, research demonstrates that individuals are adept at giving an accurate indication of their perceived cognitive load (Paas 1992; Paas and van Merriënboer 1994;

Paas et al. 1994). Consequently, the 7-point Likert scale used to measure cognitive load in this study is adapted from Paas (1992).⁵⁵

4.4.1.1 Where I Measured Cognitive Load During the Task.

Within my study, determining the optimal point to measure cognitive load presented a formidable challenge. Although asking participants to assess their cognitive load at the end of each analytical review stage (i.e., after hypothesis generation, hypothesis evaluation, and hypothesis selection) seemed to be the optimal point to measure cognitive load, having participants assess and report their cognitive load at the end of each stage also seemed to constitute a distracter task. Consequently, I feared having participants self-report their cognitive load at the end of each stage would clear out at least a portion of their working memory, leaving participants with fewer relevant cues to call upon in performing the task, artificially reducing participants' analytical review effectiveness. The risk was especially relevant to my task because asking participants to self-report their cognitive load in the middle of the task might have artificially increased the effectiveness of the decision aid interventions since participants that do not receive any decision aids must rely more heavily upon their working memory as compared to aid-users in conducting the task.

Consequently, I chose to measure cognitive load for each of the three stages at the end of the experiment in the post-experiment questionnaire even though I knew *ex ante*

⁵⁵ Underscoring the difficulty involved with measuring cognitive load, I originally proposed measuring cognitive load using the NASA TLX Task Load Index. However, consistent with the findings of Rubio et al. 2004, the pilot study results suggested the NASA TLX instrument was not sensitive enough to pick up changes in cognitive load levels across intervention conditions. Thus, after consultation with my dissertation chairman, the decision was made not to use the NASA TLX instrument to measure cognitive load within this study.

that a considerable amount of time and activity would occur between the point that cognitive load was experienced and the point that cognitive load was measured.

4.4.2 Determining Whether a Mediating Relationship Exists.

To conclude that cognitive load mediates the relationship between analytical review effectiveness and decision aid use, I must first demonstrate that each decision aid intervention is significantly related to analytical review effectiveness without controlling for cognitive load. To test this, I estimate a regression that excludes cognitive load from the model. I expect this regression to estimate the coefficient c to be positive and significantly related to analytical review effectiveness at conventional significance levels.⁵⁶

Next, I need to demonstrate that cognitive load is significantly related to analytical review effectiveness when both decision aid and cognitive load are modeled as independent variables in equation 2. In other words, both cognitive load *and* the decision aid must be statistically significant when estimating analytical review effectiveness. Further, the coefficient of the decision aid in equation 1 should be inversely related to the coefficient of the decision aid in equation 2.

Finally, I must demonstrate that there is a significant relationship between the decision aid and cognitive load. To test this, I estimate a regression where cognitive load is the dependent variable and the decision aid is the independent variable. I expect the regression estimate of equation 3 to return a negative a coefficient since I predict that decision aids reduce cognitive load. Further, I expect the regression estimate will indicate the decision aid is significantly associated with cognitive load at conventional significance levels.

⁵⁶ Within this study conventional significance levels are defined as those below $p=0.10$.

A variable may be considered a mediator to the extent that it accounts for the relationship between the independent variable and the dependent variable (Baron and Kenny 1986). Full mediation occurs when the relationship, previously found to be significant, between the independent variable and the dependent variable is reduced to zero when the mediator is introduced. Partial mediation occurs when the relationship between the independent variable and the dependent variable is reduced in the presence of the mediator. Table 4.16 illustrates the four steps performed (and the required outcome) to test whether cognitive load mediates the relationship between decision aids and analytical review effectiveness.

TABLE 4.16. The Four Steps and Required Outcomes Necessary to Provide Evidence that Cognitive Load Mediates Decision Aid Use and Analytical Review Effectiveness.

Step	Requirement	Test Applied	Required Outcome
1	Demonstrate the decision aid is significantly related to analytical review effectiveness.	Estimate a regression where analytical review effectiveness is the dependent variable and the decision aid and other covariates are independent variables. Cognitive load may not serve as a covariate.	The decision aid must be significantly related to analytical review effectiveness.
2	Demonstrate that both the decision aid and cognitive load are significantly related to analytical review effectiveness.	Estimate a regression where analytical review effectiveness is the dependent variable and decisions aid and cognitive load are independent variables.	Cognitive load <i>and</i> the decision aid must be significant when estimating analytical review effectiveness.
3	Demonstrate that cognitive load explains a significant portion of analytical review effectiveness.	Compare the decision aid coefficient estimates from regression models in Step 1 and Step 2.	The decision aid coefficient from step 2 should be inversely related to the decision aid coefficient from step 1.
4	Demonstrate there is a significant relationship between the decision aid and cognitive load.	Estimate a regression where cognitive load is the dependent variable and the decision aid is the independent variable.	Decision aid users should experience significantly less cognitive load than no-aid users.

4.4.3 Activity Relationship Diagram Mediation Results.

Because hypotheses testing reveals the ARD improves the analytical review effectiveness of participants during all three stages of analytical review (i.e., hypothesis

generation, hypothesis evaluation, and hypothesis selection), I examine whether cognitive load mediates the relationship between the ARD and analytical review effectiveness in these stages.

Examining whether cognitive load mediates the relationship between the ARD and analytical review effectiveness in hypothesis generation, Step 1 results indicate that the ARD does not significantly improve analytical review effectiveness when cognitive load is removed as a covariate from the model ($\chi^2 = 4.753$, $p = 0.11$, one-tailed). Thus, cognitive load does not mediate the relationship between the ARD and analytical review effectiveness during hypothesis generation. However, results do indicate that cognitive load is significantly related to ARD hypothesis generation effectiveness ($\chi^2 = 7.763$, $p = 0.01$).

Results indicate the ARD is significantly related to hypothesis evaluation effectiveness (Step 1: $\chi^2 = 3.506$, $p = 0.045$, one-tailed) and adding cognitive load as a predictor variable in the model is significant (Step 2: $\chi^2 = 12.258$, $p = 0.055$). Further, the coefficient of the ARD is inversely related to the level of cognitive load (Step 3: $\beta = 2.234$ and 5.123 , respectively). However, there is no significant relationship between cognitive load and the ARD (Step 4: $F = 0.057$, $p = 0.815$). Thus, there is no evidence that cognitive load mediates the relationship between the ARD and hypothesis evaluation effectiveness.

The ARD is significantly related to hypothesis selection effectiveness (Step 1: $\chi^2 = 7.276$, $p = 0.017$, one-tailed) and adding cognitive load as a predictor variable in the model is significant (Step 2: $\chi^2 = 19.514$, $p = 0.006$). Further, the coefficient of the ARD is inversely related to the level of cognitive load (Step 3: $\beta = 2.037$ and 2.576 , respectively). However, there is no significant relationship between cognitive load and the ARD (Step

4: $F=0.144$, $p=0.707$). Thus, there is no evidence that cognitive load mediates the relationship between the ARD and hypothesis selection effectiveness.

4.4.4 Pattern-Consideration Aid Mediation Results.

Hypothesis testing reveals the PCA improves the analytical review effectiveness of participants during hypothesis evaluation and hypothesis selection, so I examine whether cognitive load mediates the relationship between the PCA and analytical review effectiveness in stages.

Examining whether cognitive load mediates the relationship between the PCA and analytical review effectiveness in hypothesis evaluation, Step 1 results indicate that the PCA significantly improves analytical review effectiveness when cognitive load is removed as a covariate from the model (Step 1: $\chi^2= 2.947$, $p=0.058$, one-tailed) although adding cognitive load as a predictor variable in the model is not significant (Step 2: $\chi^2= 6.783$, $p=0.139$). Consequently, there is no evidence that cognitive load mediates the relationship between the PCA and hypothesis evaluation effectiveness.

The PCA is significantly related to hypothesis selection effectiveness (Step 1: $\chi^2= 9.259$, $p=0.048$, one-tailed) and adding cognitive load as a predictor variable in the model is significant (Step 2: $\chi^2= 17.116$, $p=0.022$). Further, the coefficient of the PCA is inversely related to the level of cognitive load (Step 3: $\beta=1.495$ and 2.309 , respectively). However, there is no significant relationship between cognitive load and the PCA (Step 4: $F=0.470$, $p=0.497$). Thus, there is no evidence that cognitive load mediates the relationship between the PCA and hypothesis selection effectiveness.

4.4.5 Combined-Aid Mediation Results.

Because hypothesis testing reveals the combined-aid intervention improves the analytical review effectiveness of participants during hypothesis evaluation, I examine whether cognitive load mediates the relationship between decision aids and analytical review effectiveness in this stage.

Examining whether cognitive load mediates the relationship between the combined-aid intervention and analytical review effectiveness in hypothesis evaluation, Step 1 results indicate that the combined-aid intervention does not significantly improve analytical review effectiveness when cognitive load is removed as a covariate from the model (Step 1: $\chi^2 = 1.736$, $p = 0.108$, one-tailed). Thus, there is no evidence that cognitive load mediates the relationship between the combined-aid intervention and hypothesis selection effectiveness.

4.4.6 Summary of Mediation Analysis Results and Discussion of the Implications.

In summary, I find no empirical evidence to suggest that cognitive load mediates the relationship between the decision aids examined within this study and the positive effect the aids are found to have on analytical review effectiveness.

Although I predicted that decision aids would improve participants' analytical review effectiveness by reducing the amount of cognitive load they experienced during the task, I did not find a significant relationship between the decision aids examined in this study and the level of cognitive load reported by participants. A potential explanation for the reason I did not find cognitive load to be significantly lower among decision aid users could be the timing of my cognitive load measurement. I measured cognitive load during the post-experimental questionnaire, where participants were asked

to indicate the level of cognitive load they experienced during each of the three analytical review stages. A considerable amount of time and activity occurred between the point where participants experienced the cognitive load of interest and the point that they self-reported their cognitive load. Thus, it may be possible that participants were unable to accurately recall the amount of cognitive load they experienced in each stage.

I was aware, *ex ante*, that measuring cognitive load for each of the three analytical review stages at the end of the experiment might diminish participants' ability to accurately recall how much cognitive load they experienced during each stage. However, I decided against asking participants to gauge their cognitive load at the end of each analytical review stage because I feared having participants respond in this manner would essentially constitute a distracter task. It is likely that a distracter task would clear out at least a portion of no-aid participants' working memory, leaving them with fewer relevant cues in working memory to call upon in performing the task, ultimately reducing their analytical review effectiveness. Thus, a cognitive load questionnaire placed in the middle of the task could artificially increase the effectiveness of the decision aid interventions since it could clear the working memory of participants that do not receive a decision aid(s) and who need to rely more heavily upon their working memory to effectively conduct the task.

4.5 Post Hoc Analysis

In this section, I present a discussion of the post hoc testing that I conducted. The purpose of conducting post hoc analysis was to empirically investigate several phenomena of interest that were identified during hypotheses testing.

4.5.1 The ARD and the PCA as Memory Aids.

Another potential reason that I may not have found cognitive load to be lower among decision aid users could be that the two decision aids examined in this study may be more akin to memory aids than tools that reduce the cognitive load placed upon participants during analytical review. Memory aids are decision aids whose purpose is to assist individuals in recovering knowledge from memory that is relevant to a given judgment and decision making task (Bonner 2008).

Research demonstrates that individuals are more accurate when using their memory to recognize information as opposed to using their memory to recall information (MacDougall 1904; Ratcliff 1978; Mandler 1980; Hintzman 1990). Within the context of analytical review, auditors need to accurately recall accounting relationships because the cause of the unexpected fluctuation is deduced by analyzing unanticipated fluctuations in one or more specific accounts. Further, auditors need to accurately recall the cues they believe to be relevant to the cause of the unexpected fluctuation in order to generate, evaluate, and select the correct cause of the unexpected fluctuation.

Because people often fail to retrieve all the relevant information while performing a task, memory aids can help users with the cognitive process of recovering knowledge from memory (Bonner 2008). Within audit practice, the use of standard audit programs are an example of a memory aid used to help the auditor recall relevant information in conducting an audit engagement (Bonner 2008). Thus, the use of memory aids can help individuals to improve the quality of the information they consider in conducting a judgment and decision making task (Bonner 2008).

4.5.1.1 The ARD as a Memory Aid.

Even though auditors learn the common accounting 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. Consequently, I suggest the ARD might serve as a memory aid by helping auditors to recognize common accounting relationships while performing an analytical review task. If the ARD functions as a memory aid, I would expect participants using the ARD to have a better understanding of the accounting relationships depicted in it. To find evidence of this, I examined the accuracy of participant responses to the question, "When a company uses a periodic inventory system, how is cost of goods sold (COGS) calculated at the end of the year?" Using a Fishers Exact Test to examine the proportion of ARD participants who answered the question correctly as compared to participants who did not receive the ARD, I find marginal evidence to support the assertion that participants who relied on the ARD correctly answered the question more often than participants who did not receive the ARD ($\chi^2 = 2.318$, $df = 1$, $p = 0.129$). This finding provides some evidence to suggest that the ARD may function as a memory aid.

4.5.1.2 The PCA as a Memory Aid.

Even though auditors identify the cues they believe to be relevant to the cause of the unexpected fluctuation during analytical review, an auditor's ability to accurately recall these cues may be negatively affected by the same cognitive factors that prevent them from accurately recalling common accounting relationships. Consequently, I suggest that the PCA may serve as a memory aid because it automatically presents an

electronic list of the user-identified relevant cues, helping PCA users to accurately recall and consider the cues during analytical review. If the PCA serves as a memory aid, I would expect PCA users to more accurately estimate the likelihood that each of their proposed hypotheses is the correct cause of the unexpected fluctuation because PCA users should be more aware of the number of cues that support each hypothesis as compared to those who do not use the aid. For example, to properly evaluate the merits of each proposed hypothesis, non-PCA users must accurately recall the cues they previously identified as relevant. Because auditors may not be able to accurately recall all of the cues they identified due to cognitive constraints (Bonner 2008), I suggest that non-PCA users may conduct hypothesis evaluation using a reduced cue set. If so, when a non-PCA user evaluates the likelihood that each of his or her proposed hypotheses is the correct cause of the unexpected fluctuation, the hypothesis that best matches the individual's reduced cue set should be rated as more probable. However, a non-PCA user is likely to under-estimate the likelihood that each of the remaining proposed hypotheses are the correct cause because he or she will not evaluate them against the full set of relevant cues. Therefore, if the PCA serves as a memory aid, I would expect the probability assessments of non-PCA users to exhibit greater variability between the hypothesis rated most likely to be correct and the hypothesis rated the *next* most likely to be correct.

Constructing an ANOVA model where I compare the difference between the hypothesis rated most likely to be correct and the hypothesis rated the next most likely to be correct, I find the average percentage difference for PCA users is 30.23 percent and

the average percentage difference for no-aid users is 49.83 percent ($F = 3.840$, $df = 1$, $p = 0.058$), providing some empirical evidence that the PCA may serve as a memory aid.⁵⁷

4.5.2 Overconfidence.

During hypothesis testing the number of audit courses completed by a participant was unexpectedly found to be negatively related to hypothesis generation effectiveness and hypothesis selection effectiveness during analytical review. This finding was unexpected because it seems intuitive to assume that the number of audit courses completed should improve task effectiveness. Consequently, I conducted post hoc analysis to examine why the number of audit courses completed was negatively correlated with task effectiveness.

A potential explanation for this counter-intuitive finding is a bias known within the information-processing literature as “overconfidence” (Koriat et al. 1980; Lichtenstein et al. 1982). Overconfidence is frequently defined as increases in decision confidence without associated improvements in decision quality (Rose 2002). Overconfidence is displayed when an individual’s confidence in the accuracy of their response is greater than their actual accuracy (Bonner 2008). Empirical evidence demonstrates that overconfidence results from an individual’s belief in their own knowledge or expertise (Whitecotton 1996; Rose 2002). Thus, the more knowledgeable individuals perceive they are in regards to a subject matter, the more confident they may become in their ability to effectively perform a task related to the subject matter. Because confidence comes from an individual’s perception of their knowledge, rather

⁵⁷ There was also a significant difference between PCA users (\bar{x} PCA percentage difference = 30.23 percent) and ARD users (\bar{x} ARD percentage difference = 55.95 percent), suggesting that the reduction in variance is uniquely associated with the PCA and is not solely associated with providing a decision aid to participants.

than from an individual's actual knowledge, overconfidence often leads to suboptimal task performance.

A considerable body of literature supports the notion that overconfidence is directly related to task difficulty; in a review of the overconfidence literature, Lichtenstein, Fischhoff, and Phillips (1982) conclude that "overconfidence is most extreme with tasks of great difficulty" (p. 315).⁵⁸ Due to the inherent complexity of analytical review, it is a relatively difficult task. Thus, finding evidence that an individual's exposure to auditing concepts leads to overconfidence is consistent with prior research.⁵⁹ Although participants who completed more than one audit course felt it was significantly easier to propose the correct hypothesis than those who had not completed any audit courses (Panel A, Table 4.17: $F=2.723$, $df = 2$, $p=0.072$), the odds that the correct hypothesis was proposed decrease by a multiplicative factor of 0.282 for every audit course completed ($\chi^2 = 4.273$, $df = 1$, $p=0.039$) after controlling for cognitive load and the number of non-audit accounting courses completed. Results also indicate that participants who completed more than one audit course felt it was significantly easier to evaluate their proposed hypotheses (Panel B, Table 4.17: $F=2.922$, $df = 2$, $p=0.060$), were significantly more confident that the hypothesis they selected was the "correct" one (Panel C, Table 4.17: $F=2.437$, $df = 2$, $p=0.094$), and felt it was significantly easier to conduct hypothesis selection (Panel D, Table 4.17: $F=4.279$, $df = 2$, $p=0.018$) as compared to participants who completed more than one audit course. However, the odds that the correct hypothesis was actually selected decreased by a multiplicative factor of

⁵⁸ In the research reviewed by Lichtenstein, Fischhoff, and Phillips (1982), difficulty was defined on the basis of participants' performance.

⁵⁹ Within this study the most audit courses completed by any participant was two.

0.163 for every audit course completed ($\chi^2 = 5.820$, $df = 1$, $p=0.016$) after controlling for cognitive load, the number of non-audit accounting courses completed, and whether a participant used a decision aid. Thus, while participants who completed more audit courses felt it was easier to conduct analytical review and were more confident that they successfully selected the correct cause of the unexpected fluctuation, they actually performed analytical review less effectively than participants who completed fewer audit courses. These results suggest that participants who completed more audit courses suffered from overconfidence, which ultimately reduced their analytical review effectiveness.

TABLE 4.17. Post Hoc Overconfidence Descriptive Statistics.

Panel A: Participant Reported Ease or Difficulty in Proposing the Correct Answer			
Number of Audit Courses Completed	Mean Difficulty (1 = Very Easy, 7 = Very Difficult)	Standard Deviation	N
0	4.03*	1.616	32
1	4.12	1.452	34
2	2.78*	1.856	9

*Significant difference at $p<0.10$

Panel B: Participant Reported Ease or Difficulty in Hypothesis Evaluation			
Number of Audit Courses Completed	Mean Difficulty (1 = Very Easy, 7 = Very Difficult)	Standard Deviation	N
0	3.41*	1.563	32
1	3.39	1.001	34
2	2.22*	1.481	9

*Significant difference at $p<0.10$

Panel C: Participant Reported Confidence That the Correct Hypothesis Was Selected			
Number of Audit Courses Completed	Mean Confidence (0% - 100%)	Standard Deviation	N
0	70.09%*	21.511	32
1	70.37%	20.398	34
2	85.89%*	8.905	9

*Significant difference at $p<0.10$

4.5.3 Effective Hypothesis Generation.

Although participants in the ARD intervention proposed the correct hypothesis more often than participants who did not receive any decision aid, neither the PCA nor the combined-aid interventions were found to help participants propose the correct cause of the unexpected fluctuation. Thus, I conducted post hoc analysis in an attempt to identify other non-decision aid factors that may be significantly related to hypothesis generation effectiveness.

The covariates included in the LOGIT model estimated to test Hypothesis 1b (H1b) suggest that cognitive load, the number of non-audit courses completed, and the number of audit courses completed are at least marginally significant predictors of whether a participant generates the correct hypothesis. The beta of these coefficients in H1b's LOGIT model have intuitive appeal: As a participant's cognitive load increases, the beta shows that it is less likely the participant will propose the correct hypothesis. This suggests cognitive load is negatively associated with analytical review effectiveness. Consistent with intuition that accounting knowledge should be positively associated with accounting task effectiveness, the odds that a participant proposes the correct hypothesis increase with each non-audit accounting course a participant completes. This suggests that a participant's analytical review effectiveness improves as the participant's knowledge of accounting concepts increases. Finally, due to overconfidence (which was discussed earlier), it is not surprising to find the odds that a participant proposes the correct hypothesis decrease as the number of audit courses the participant completed increases. In sum, it seems reasonable to predict that these three covariates might be factors associated with hypothesis generation effectiveness.

To test the suggestion that cognitive load, accounting knowledge, and overconfidence are related to hypothesis generation effectiveness, I estimate a Logistic regression model that regresses these three variables against whether a participant proposes the correct hypothesis. Table 4.18 presents the logistic regression model and descriptive statistics. The logistic regression model is populated with data from all interventions. Results indicate that the model is significant ($\chi^2 = 11.347$, $df = 3$, $p=0.010$) and each of the three independent variables are significant ($p_{\text{cognitive load}} = 0.018$, $p_{\text{number of non-audit accounting courses completed}} = 0.024$, $p_{\text{overconfidence}} = 0.043$). The odds that the correct hypothesis is proposed by a participant decrease by a multiplicative factor of 0.663 for every one unit cognitive load increase that a participant experiences. For every non-audit accounting course a participant completes, the odds that the participant generates the correct hypothesis increase by a multiplicative factor of 2.462. Presumably due to overconfidence, the odds that the correct hypothesis is proposed by a participant decrease by a multiplicative factor of 0.299 for every audit course the participant completes.

TABLE 4.18. Post Hoc Analysis: Hypothesis Generation Effectiveness.

Panel A: LOGIT Model and Significance						
LOGIT	Overall Model χ^2	Degrees of Freedom	Significance			
Model	11.347	3	0.010			

Panel B: LOGIT Model Descriptive Statistics						
	B	Exp(B)	Standard Error	Wald χ^2	df	Significance
Cognitive Load (1 = Very Low, 7 = Very High)	-0.411	0.663	0.173	5.632	1	0.018*
Accounting Knowledge (Number of Non-Audit Courses Completed)	0.901	2.462	0.597	5.076	1	0.024*
Overconfidence (Number of Audit Accounting Courses Completed)	-1.208	0.299	0.597	4.093	1	0.043*
Constant	-2.641	0.071	1.739	2.307	1	0.129

Dependent Variable: Whether the correct hypothesis was proposed

*Significant at $p < 0.05$

**Significant at $p < 0.10$

4.5.4 Summary of Post Hoc Findings.

To summarize my post hoc analysis, there is empirical evidence to suggest that a participant's exposure to auditing concepts may engender a sense of overconfidence, ultimately leading to reduced analytical review effectiveness. Further, I find some empirical evidence that cognitive load, accounting knowledge, and overconfidence are all associated with a participant's hypothesis generation effectiveness. Finally, I found some empirical evidence to suggest that the ARD and the PCA may function as memory aids.

5.0 SUMMARY AND CONCLUSIONS

5.1 Summary of Key Findings

This study uses a realistic analytical procedures task in which inexperienced auditors conducted analytical review across three stages: hypothesis generation, hypothesis evaluation, and hypothesis selection. Prior research finds it is difficult for experienced auditors to propose, evaluate, and select the correct cause of an unexpected fluctuation (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). As a result, experienced auditors often have difficulty conducting effective analytical review.

The results of this study indicate that like experienced auditors, inexperienced auditors have difficulty conducting analytical review effectively. As predicted, both of the decision aids examined in this study were found to increase auditors' analytical review effectiveness when auditors placed at least some reliance upon them. However, contrary to my predictions, auditors who used both decision aids did not perform analytical review more effectively than auditors who used only one decision aid.

I found empirical evidence to suggest that the level of cognitive load is negatively related to analytical review effectiveness. As an auditor's self-reported cognitive load increased, the auditor's analytical review effectiveness declined. Cognitive load was

found to be negatively associated with hypothesis generation effectiveness, hypothesis evaluation effectiveness, and hypothesis selection effectiveness.

My post hoc analysis results suggest that even inexperienced auditors can fall prey to the overconfidence bias. As the number of audit courses completed by an inexperienced auditor increased, the auditor became more confident that he or she selected the correct cause of the unexpected fluctuation. Further, the auditor also felt it easier to propose, evaluate, and select the correct cause. However, my results show that the odds that an inexperienced auditor conducted analytical review effectively actually decreased with every additional audit course completed. Although the judgment and decision making literature has extensively established the link between knowledge and overconfidence, finding that even inexperienced auditors can fall victim to overconfidence was unexpected. Extending this finding to experienced auditors, who have more accounting knowledge than inexperienced auditors, a potential implication is that experienced auditors may exhibit overconfidence to a greater degree than inexperienced auditors within the context of analytical review. Therefore, the analytical review effectiveness of experienced auditors is likely to be more negatively affected by the overconfidence bias as compared to the analytical review effectiveness of inexperienced auditors. Thus, overconfidence may help to explain the existence of the analytical review paradox.

Within my post hoc analysis, I find evidence to suggest that accounting knowledge is a significant factor that increases the odds that an auditor will propose the correct hypothesis. I also find some evidence to support the notion that the decision aids examined within this study may function as memory aids.

Finally, given that all of the junior-level accounting students who participated in my experiment were unable to answer questions regarding common accounting relationships, this suggests that junior-level accounting students may not have a developed grasp of basic accounting principles. This finding suggests that junior-level accounting students may not be an inappropriate proxy for auditors. Consequently, researchers should exercise caution when generalizing the results of studies that proxy junior-level accounting students as auditors.

5.2 Contributions

The major contribution of this study is that I find evidence that decision aids can significantly improve auditors' analytical review effectiveness without unduly increasing the time needed to conduct the task.⁶⁰ The ARD was found to improve auditors' analytical review effectiveness during hypothesis generation, hypothesis evaluation, and hypothesis selection without increasing the time needed to conduct the task.⁶¹ The PCA was found to improve auditors' analytical review effectiveness during hypothesis evaluation and hypothesis selection. Although it took approximately 3 minutes longer for auditors using the PCA to conduct analytical review than auditors that did not use the aid, this relatively small time increase does not seem likely to dissuade auditors from using the PCA since it improves auditors' hypothesis evaluation effectiveness and hypothesis selection effectiveness.

⁶⁰ Time is always an important practitioner consideration because auditors have to accomplish their work under tight time budgets.

⁶¹ While participants that relied on the ARD were found to generate the correct hypothesis more often than those who did not use any decision aid (Percent Generated_{ARD} = 43.75 vs. Percent Generated_{No-Aid} = 30.43), care should be taken when making inferences about the practical significance of this difference due to the fact that participants within both interventions generated the correct hypothesis seven times.

These findings have two practical implications: First, because auditors using either decision aid significantly outperformed auditors not provided with a decision aid, it seems logical to suggest that auditors should use some type of decision aid during analytical review. Thus, a major contribution of this study is in providing audit practitioners with two relatively easy-to-deploy decision aid tools that should improve auditors' analytical review effectiveness. The second implication is that practitioners should exercise care with regard to the number of decision aids they provide to auditors. Although the use of one decision aid was found to improve auditors' analytical review effectiveness, the use of more than one decision aid was not found to incrementally improve auditors' analytical review effectiveness.

This study contributes to the decision aid reliance literature by finding that individuals can improve their task effectiveness by placing even a light degree of reliance upon a decision aid. Since both of the decision aids examined within this study were found to improve aid users' analytical review effectiveness regardless of the self-reported degree of reliance, this also suggests that it may not be necessary or desirable to force individuals to place complete reliance on a decision aid. This finding also suggests that in some situations individuals may be able to effectively determine the appropriate degree of reliance to place on an aid. Of course, more research is needed before definitive conclusions can be drawn.

Finding cognitive load to be negatively related to task effectiveness contributes to the Cognitive Load Theory stream of literature by providing empirical evidence that cognitive load can have task performance implications. Consequently, in finding cognitive load to be negatively related to analytical review effectiveness, I extend the

tenets of cognitive load theory outside its traditional learning context by providing evidence that cognitive load can also inhibit task effectiveness. One practical implication of this finding is to suggest that audit practitioners seek to minimize the amount of cognitive load placed upon them while conducting analytical review. Reducing the cognitive load placed upon an auditor during analytical review should lead to improved analytical review effectiveness.

This study contributes to the audit overconfidence literature by finding evidence that even inexperienced auditors can succumb to the overconfidence bias. A practical implication of this finding is that both accounting practitioners and accounting educators need to take greater efforts to educate auditors on the dangers of overconfidence.

Finding that accounting knowledge increases the odds that an auditor proposes the correct hypothesis has two primary contributions: First, it contributes to the academic literature by providing evidence of a positive relationship between accounting knowledge and analytical review effectiveness. Second, this finding provides evidence that domain-specific knowledge is positively related to task effectiveness. One practical implication of this finding is that audit practitioners need to be aware that a task is more likely to be performed effectively when the individual(s) conducting the task possess greater degrees of domain-specific knowledge.

5.3 Limitations

This study's findings and conclusions should be considered in light of its limitations. First, because I employ students as surrogates for inexperienced auditors, the results of this study may not generalize to settings where experienced auditors conduct analytical review. However, prior research finds that even experienced auditors have

great difficulty conducting analytical review (Bedard and Biggs 1991; Bedard, Biggs, and Maroney 1998). 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.

Second, the unexpected fluctuation used in this study had only one cause. Thus, this study's results may be less generalizable as the number of causes associated with an unexpected fluctuation rises. Lastly, this study only examines an unexpected fluctuation arising due to 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, whether the effectiveness of the two decision aids investigated in this study will hold outside of client error remains an empirical question.

5.4 Future Research Opportunities Arising from this Study's Results

The results of this study may provide some future research opportunities. Contrary to my predictions, I found that auditors who used both decision aids did not perform analytical review more effectively than auditors using only one decision aid. A potential explanation for this result may be that the use of both decision aids increased the level of cognitive load placed upon the auditor. This explanation is consistent with the "split attention effect" (Rose and Wolfe 2000), where an individual who receives multiple decision aids actually experiences a cognitive load increase, instead of a cognitive load decrease, because the individual's working memory must be split between the multiple aids. Finding that auditors in the combined-aid intervention proposed the fewest mean number of plausible hypotheses provides some evidence that combined-aid users may have experienced a split attention effect.

Future research could investigate whether providing auditors with both decision aids in a sequential order might improve their analytical review effectiveness compared to auditors that only receive one decision aid. For example, since the ARD alone was found to improve hypothesis generation effectiveness, perhaps the ARD should be provided by itself to auditors during the hypothesis generation stage. After hypothesis generation ends, the researcher could then remove the ARD, leaving auditors to complete hypothesis evaluation and hypothesis selection using only the PCA, which was found to be as effective as the ARD in both these stages. Providing auditors with both decision aids in a sequential fashion might be a way to eliminate the split attention effect, leading to better analytical review effectiveness.

Future research could also examine the impact of decision aid training on task effectiveness. Within this study, the fact that auditors received both decision aids without receiving any training on how to use each aid may have overwhelmed auditors. This suggests training can play a vital role in decision aid effectiveness.

Researchers may find it fruitful to examine whether decision aids can be a suitable substitute for domain-specific knowledge. Additionally, researchers could examine whether decision aids can help individuals recall domain-specific knowledge that may have receded from memory. Thus, future research could examine whether decision aids can permit individuals to more effectively perform tasks for which their domain-specific knowledge has atrophied.

Future research could also examine the process-level strategies of inexperienced auditors during analytical review by comparing and contrasting the process-level steps of successful and unsuccessful inexperienced auditors during an analytical review task.

This extension would complement the work of Bedard and Biggs (1991), who examine the process-level strategies of experienced auditors during analytical review.

5.5 Future Research Opportunities Extending this Study's Decision Aids

Extending the two decision aids investigated within this study 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 inexperienced 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 decision aids used within this study 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 correct hypothesis (Bedard and Biggs 1991; Bedard, Biggs, and Maroney 1998), the PCA may be especially useful to experienced auditors during hypothesis generation because it automatically displays the auditor-identified relevant cues, making the cues more available to the auditor while he or she attempts to propose the correct hypothesis. Additionally, given that experience is often negatively correlated with decision aid reliance, this study's findings may not hold among experienced auditors because experienced auditors may opt not to rely upon the aids during analytical review.

Second, future research could extend the two decision aids investigated within this study to examine whether these aids help auditors to more effectively evaluate the sufficiency of client-provided explanations. The two decision aids used within this study may help debias auditor judgment and decision making when evaluating client-provided explanations. This extension appears particularly promising considering evidence that once an unexpected fluctuation is uncovered, auditors usually make client inquiry their first step (Trompeter and Wright 2010) and given that research finds auditors inappropriately fixate on client-provided explanations (Bedard and Biggs 1991; Anderson and Koonce 1995; Green 2004).

Third, future research could extend the use of inexperienced 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). The decision aids used within this study could serve as a manipulated variable. There are a few reasons to expect that investigating the analytical review performance of inexperienced auditors may yield different results than those previously found in the experienced auditor group literature. For example, an inexperienced auditor likely has less overall accounting knowledge than an experienced auditor. Thus, putting two inexperienced auditors together may increase the group's pool of available accounting knowledge more than a group comprised of two experienced auditors, whose pool of available knowledge may not increase as much because each experienced auditor possesses a larger accounting knowledge base. Further, due to their lack of audit experience, inexperienced 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 thus be less willing to consider alternative points of view. Consequently, it may be possible that inexperienced auditor groups may conduct analytical review more effectively than experienced auditor groups. Finally, I suggest that inexperienced auditors are more likely to acknowledge their limitations when conducting analytical review as compared to experienced auditors. Thus, inexperienced auditors may be more open to working with others while conducting analytical review as compared to experienced auditors who may view working with another auditor as unnecessary and a nuisance.

Future research is needed to establish a theoretical basis to explain why the decision aids used within this study improved auditors' analytical review effectiveness. Although this study provides considerable evidence that both the ARD and the PCA improve the analytical review effectiveness of inexperienced auditors, the mediation analysis I performed to test my application of Cognitive Load Theory to this setting does not allow me to draw definitive conclusions regarding *why* the decision aids improved auditor effectiveness. Future research using alternative methods, such as a more refined measure of cognitive load or obtaining a measure of cognitive load before the end of the study, may provide evidence that the decision aids examined within this study decrease the cognitive load placed upon auditors during analytical review. However, researchers should be aware that asking auditors to gauge their cognitive load during the task is likely to invalidate any subsequent analytical review task results due to the distracter confound discussed in the Method section.

Additionally, because I find some evidence in my post-hoc analysis to suggest that the decision aids examined within this study may function as memory aids, future

research could further investigate this link in an attempt to better explain the mechanics behind why the decision aids investigated in this study improved auditors' analytical review effectiveness.

Finally, because the two decision aids that I developed and investigated within this study were found to improve auditors' analytical review effectiveness, future research might find it fruitful to expand upon the design of these two aids or to extend their application to similar, but new, problem-solving contexts.

REFERENCES

- Abdolmohammadi, M. J. 1999. A comprehensive taxonomy of audit task structure, professional rank and decision aids for behavioral research. *Behavioral Research in Accounting*. 11: 51-92.
- Abdolmohammadi, M., and C. Usoff. 2001. A longitudinal study of applicable decision aids for detailed tasks in a financial audit. *International Journal of Intelligent Systems in Accounting, Finance, and Management*. 10: 139-154.
- AICPA. 2007. American Institute of Certified Public Accountants. *Risk Assessment Standards Overview of the Standards for AICPA PCPS Members*. Online AICPA tool kit:
http://www.aicpa.org/InterestAreas/PrivateCompaniesPracticeSection/Resources/KeepingUp/DownloadableDocuments/Overview_SASs_104_111_082007.doc
- AICPA. 2010a. American Institute of Certified Public Accountants. AU 329: Analytical Procedures. *AICPA Professional Standards*. AICPA: New York, NY: 339-344.
- AICPA. 2010b. American Institute of Certified Public Accountants. AU 314: Understanding the Entity and Its Environment. *AICPA Professional Standards*. 2010. AICPA: New York, NY: 117-159.
- AICPA. 2010c. American Institute of Certified Public Accountants. AU 316: Consideration of Fraud in a Financial Statement Audit. *AICPA Professional Standards*. 2010. AICPA: New York, NY: 169-220.
- 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. 1990. Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback, and justification. *Journal of Accounting Research*. 28: 148-180.
- 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.
- Baron, M. B., and D. A. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*. 51(6): 1173-1182.
- 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.
- Benbasat, I., and B. R. Nault. 1990. An evaluation of empirical research in managerial support systems. *Decision Support Systems*. 6: 203-226.

- 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.
- Blocher, E., and R. S. Esposito, and J. J. Willingham. 1983. Auditors' analytical review judgments for payroll expense. *Auditing: A Journal of Practice & Theory*. 3(1): 75-91.
- 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.
- Brewster, B. E. 2010. Enhancing auditor expertise: Examining how a reinforcing feedback loop between knowledge and ability affects task performance. Working paper.
- Butler, S. A. 1985. Application of a decision aid in the judgmental evaluation of substantive tests of details samples. *Journal of Accounting Research*. 23: 513-526.
- Chandler, P., and J. Sweller. 1991. Cognitive load theory and the format of instruction. *Cognition and Instruction*. 8: 293-332.
- Chandler, P., and J. Sweller 1992. The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*. 62(2):233-246 as cited on p. 124 by Jacob Rose in 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, Florida. 111-133
- 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.

- Eining, M. M., and D. R. Jones, and J. K. Loebbecke. 1997. Reliance on decision aids: An examination of auditors' assessment of management fraud. *Auditing: A Journal of Practice & Theory*. 16(2): 1-19.
- Engle, R. W., and M. J. Kane. 2003. Executive attention, working memory, capacity, and a two-factor theory of cognitive control. *The Psychology of Learning and Motivation*. 44:145-199.
- Fleiss, J.L., and B. Levin, and M. C. Paik. 2003. *Statistical methods for rates and proportions*. 2003. Third Edition. John Wiley & Sons: Hoboken, New Jersey.
- Glover, S., and D. Prawitt, and B. Spilker. 1997. The influence of decision aids on user behavior: Implications for knowledge acquisition and inappropriate reliance. *Organizational Behavior and Human Decision Processes*. 72: 232-255.
- 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.
- Haist, F., and A. P. Shimanura, and L. R. Squire. 1992. On the relationship between recall and recognition in memory. *Journal of Experimental Psychology*. 18(4): 691-702.
- 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.
- Hennes, K., and A. Leone, and B. Miller. 2008. The importance of distinguishing errors from irregularities in restatement research: The case of restatements and CEO/CFO turnover. *The Accounting Review*. 83: 1487-1519.
- Hintzman, D. 1990. Human learning and memory: Connections and dissociations. *Annual Review of Psychology*. 41: 109-139.
- Hirst, D. E. 1995. Auditors' sensibility to source reliability. *Journal of Accounting Research*. 32(1): 113-126.

- Hirst, D. E., and L. Koonce. 1996. Audit analytical procedures: A field investigation. *Contemporary Accounting Research*. (13)2: 457-486.
- Horcher, A., and G. P. Tejay. 2009. Building a better password: The role of cognitive load in information security training. In the *Proceedings of the 2009 IEEE International Conference on Intelligence and Security Informatics*. 113-118.
- Hylas, R. E., and R. H. Ashton. 1982. Audit detection of financial statement errors. *The Accounting Review*. 57(4): 751-765.
- IAASB. 2010a. 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.
- IAASB. 2010b. International Auditing and Assurance Standards Board. International Standard on Auditing 315: Identifying and assessing the risks of material misstatement through understanding the entity and its environment. *Handbook of International Quality Control, Auditing, Review, Other Assurance, and Related Services Pronouncements*. 2010. IFAC: New York, NY: 262-312.
- Jamal, K. P., and P. Johnson, and R. Berryman. 1995. Detecting framing effects in financial statements. *Contemporary Accounting Research*. 12(1): 85-105.
- Johnson, V. E., and S. E. Kaplan. 1996. Auditors' decision-aided probability assessments: An analysis of the effects of list length and response format. *Journal of Information Systems*. 10(2): 78-101.
- Kinney, W. E. 1987. Attention-directing analytical review using accounting ratios: A case study. *Auditing: A Journal of Practice & Theory*. 6(2): 59-73.
- Kirschner, P. A. 2002. Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*. 12(1): 1-10.
- Kleinmuntz, B. 1990. Why we still use our heads instead of formulas: Towards an integrated approach. *Psychological Bulletin*. 107(3): 296-310.
- Koonce, L. 1993. A cognitive characterization of audit analytical review. *Auditing: A Journal of Practice & Theory*. 12(Supplement): 57-76.

- Koriat, A., and S. Lichtenstein, and B. Fischhoff. 1980. Reasons for confidence. *Journal of Experimental Psychology: Human Learning*. 6: 107-118
- Kowalczyk, T. K., and C. J. Wolfe. 1998. Anchoring effects associated with recommendations from expert decision aids: An experimental analysis. *Behavioral Research in Accounting*. 10(Supplement): 147-169.
- Libby, R. 1985. Availability and the generation of hypotheses in analytical review. *Journal of Accounting Research*. 23(2): 648-667.
- Lichtenstein, S., and B. Fischhoff, and L. Phillips. 1982. Calibration of Probabilities: The state of the art to 1980. *In Judgment under Uncertainty: Heuristics and Biases*, edited by D. Kahneman, P. Slovic, and A. Tversky. New York: Cambridge University Press, 1982.
- 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.
- Lynch, A. L., and U. Murthy, and T. J. Engle. 2009. Fraud brainstorming using computer-mediated communication: The effects of brainstorming technique and facilitation. *The Accounting Review*. 84(4): 1209-1232.
- McDaniel, L. S., and W. R. Kinney. 1995. Expectation-formation guidance in the auditor's review of interim financial information. *Journal of Accounting Research*. 33(1): 59-76.
- MacDougall, R. 1904. Recognition and Recall. *The Journal of Philosophy, Psychology and Scientific Methods*, 1(9): 229-233.
- Mackay, J. M., and S. H. Barr, and M. G. Kletke. 1992. An empirical investigation of the effects of decision aids on problem-solving processes. *Decision Sciences*. 23: 648-672.
- MacKinnon, D. P., and A. J. Fairchild, and M. S. Fritz. 2007. Mediation analysis. *Annual Review of Psychology*. 58: 593-614.
- Mandler, G. 1980. Recognizing: The judgment of previous occurrence. *Psychological Review*, 87: 252-271.

- Miller, G. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*. 63: 81-97.
- Mousavi, S. Y., and R. Low, and J. Sweller. 1995. Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*. 87: 319-334.
- Mueller, J. M., and J. C. Anderson. 2002. Decision aids for generating analytical review alternatives: The impact of goal framing and audit-risk level. *Behavioral Research in Accounting*. 14: 157-177.
- O'Donnell, E., and J. D. Perkins. 2011. Pattern recognition during analytical procedures: The influence of systems-thinking support on auditor risk assessments. Forthcoming in the *International Journal of Accounting Information Systems*.
- 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 Merriënboer. 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 Merriënboer, and J. J. Adam. 1994. Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*. 79: 419-430.
- Paas, F., and J. E. Tuovinen, and H. Tabbers, and P. van Gerven. 2003. Cognitive load measurements as a means to advance cognitive load theory. *Educational Psychologist*. 38(1): 63-71.
- Palmrose, Z. V., and V. Richardson, and S. Scholtz. 2004. Determinants of market reactions to restatement announcements. *Journal of Accounting and Economics*. 37: 59-89.
- PCAOB. 2010. Public Company Accounting Oversight Board. *Auditing Standards Related to the Auditor's Assessment of and Response to Risk and Related Amendments to PCAOB Standards*. Available at the Public Company Accounting Oversight Board's website: http://pcaobus.org/Rules/Rulemaking/Docket%202026/Release_2010-004_Risk_Assessment.pdf

- Peecher, M. E., and I. Solomon. 2001. Theory and experimentation in studies of audit judgments and decisions: Avoiding common research traps. *International Journal of Auditing*. 5: 193-203.
- Penney, C. G. 1989. Modality effects and the structure of short-term verbal memory. *Memory and Cognition*. 17: 398-422.
- Peterson, D. K., and G. F. Pitz. 1986. Effect of input from a mechanical model on clinical judgment. *Journal of Applied Psychology*. 71(1): 163-167.
- Pincus, K. V. 1989. The efficacy of a red flags questionnaire for assessing the possibility of fraud. *Accounting, Organizations, and Society*. 14(1): 156-163.
- Plumlee, M., and T. L. Yohn. 2010. An analysis of the underlying causes attributed to restatements. *Accounting Horizons*. 24(1): 41-64.
- POB. 2000. *The Panel on Audit Effectiveness Report and Recommendations*. Public Oversight Board. August 31, 2000.
- Ratcliff, R. 1978. A theory of memory retrieval. *Psychological Review*, 85: 59-108.
- Romney, P., and P.J. Steinbart. 2009. *Accounting Information Systems*. Eleventh Edition, 2009. Pearson Education: Upper Saddle River, New Jersey.
- Rose, J. M., and C. J. Wolfe. 2000. The effects of system design alternatives on the acquisition of tax knowledge from a computerized tax decision aid. *Accounting, Organizations, and Society*. 25: 285-306.
- 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, Florida. 111-133
- Rose, J. M. 2004. Performance evaluations based on financial information: How do managers use situational information? *Managerial Finance*: 45-65.
- Rubio, S., and E. Diaz, and J. Martin, and J. Puente. 2004. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology: An International Review*. 53(1): 61-86.

- Scaife, M., and Y. Rogers. 1996. External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*. 45:185-213.
- SEC. 2010. Securities and Exchange Commission. *Release No. 34-63606*. Available online at: <http://www.sec.gov/rules/pcaob/2010/34-63606.pdf>
- Simon, H. A. 1956. Rational choice and the structure of the environment. *Psychological Review*. (63)2: 129-138.
- Simon, H. A. 1974. How big is a chunk? *Science*. 183: 482-488.
- Speier, C. 2006. The influence of information presentation formats on complex task decision-making performance. *International Journal of Human-Computer Studies*. 64: 1115-1131.
- Sterman, J. D. 1989. Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Process*. 43(1): 301-335.
- Sweeney, L. B., and J. D. Sterman. 2000. Bathtub dynamics: Initial results of a systems thinking inventory system. *Systems Dynamics Review*. 16: 249-286.
- Sweller, J. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science*. 12: 257-285.
- Sweller, J., and P. Chandler, and P. Tierney, and M. Cooper. 1990. Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology*. 119(2): 176-192.
- Sweller, J. 1994. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*. 4: 295-312.
- Sweller, J., and J. Van Merriënboer, and F. Paas. 1998. Cognitive architecture and instructional design. *Educational Psychology Review*. 10(3): 251-296.
- Tarmizi, R., and J. Sweller. 1988. Guidance during mathematical problem solving. *Journal of Educational Psychology*. 80(4): 424-436.
- Trompeter, G., and A. Wright. 2010. The world has changed – Have analytical procedure practices? *Contemporary Accounting Research*. 27(2): 669-700.

- Ward, M., and J. Sweller. 1990. Structuring effective worked examples. *Cognition and Instruction*. 7(1): 1-39.
- Wells, J. T. 2001. Irrational Ratios. *Journal of Accountancy*. August: 80-83.
- Wells, J. T. 2007. *Corporate Fraud Handbook: Prevention and Detection*. 2007. John Wiley & Sons: Hoboken, New Jersey.
- Whitecotton, S. 1996. The effects of experience and confidence on decision aid reliance: A casual model. *Behavioral Research in Accounting*. 8: 194-216.
- Zwick, R. 1988. Another look at interrater agreement. *Psychological Bulletin*. 103(3): 374-378.

APPENDIX A:
CASE MATERIALS

BACKGROUND INFORMATION

Bean Co.'s Engagement Background

Bean Co. is a small subsidiary of AMEREX (which is a large, publicly traded corporation). Although Bean Co.'s financials are not material to AMEREX, the parent company requires Bean Co. to be audited every year. Your firm has audited Bean Co. for the past five years and has always given Bean Co. an unqualified opinion.

Because Bean Co. is a small company, the accounting department consists of only a few individuals. Bean Co. maintains all of its accounting records the old fashioned way - by pencil and paper. Thus, all the accounting forms and ledgers are in paper format. Even though Bean Co.'s accounting procedures are not computerized, the engagement partner has assessed Bean Co.'s inherent risk as low because its management is very ethical. Further, your firm has examined the design of Bean Co.'s internal controls and found no significant control design deficiencies.

Bean Co.'s Business Model

The operations of Bean Co. are very straightforward: First, the company buys coffee beans directly from small Jamaican coffee farms. Next, Bean Co. imports the coffee beans to the United States and stores them in a Tampa warehouse. Finally, Bean Co. ships Jamaican coffee beans to its customers when it receives orders from them. Bean Co.'s customers are American coffee houses who prefer to avoid the hassle of importing beans directly from Jamaica.

How Bean Co. Records Inventory Purchases

Bean Co. purchases inventory (coffee beans) from its network of Jamaican farms many times per year. The average coffee bean purchase is 5,000 pounds and all coffee bean purchases are made on credit. Further, all of Bean Co.'s purchases are made FOB destination. Thus, Bean Co. does not record an inventory purchase until the accounting department receives a paper "receiving ticket" from the warehouse clerk. The receiving ticket tells the accounting department how many pounds of coffee beans were received and from which Jamaican coffee farm they came.

When the accounting department gets a receiving ticket, it matches the receiving ticket to the original purchase order to make sure the purchase was authorized. Next, the accounting department debits the "Inventory Purchases" account by the amount of the purchase and credits "Accounts Payable" by the same amount. Bean Co. has 30 days to make payment once coffee beans are received.

How Bean Co. Records Coffee Bean Sales

Bean Co.'s customers are mid-to-large sized coffee houses located throughout the United States. Because all of Bean Co.'s customers buy on credit, Bean Co. has no cash sales. Further, all of Bean Co.'s sales are made FOB shipping point. This means that Bean Co.'s accounting department records a sale into the "Revenue" journal when it receives a paper "shipping ticket" from the shipping department. The shipping ticket tells the accounting department how many pounds of coffee beans were shipped and to which customer they were sent.

When the accounting department gets a shipping ticket, it matches the shipping ticket to the customer sales order. Next, the accounting department debits the "Accounts Receivable" journal by the amount of the sale and credits the "Revenue" journal by the same amount. Finally, Bean Co. sends the customer a bill. Customers have 30 days to make payment once the bill is mailed.

Bean Co.'s Inventory Method

Bean Co. uses, and has always used, a first-in-first-out (FIFO) periodic inventory system. At the end of the year, cost of goods sold is calculated by taking last year's ending inventory balance, adding the inventory purchases recorded during the year, and then subtracting the value of the coffee beans remaining in inventory. Bean Co. hires Revis, an independent third-party, to conduct a physical count of the coffee beans in inventory on the last day of each year to get the current year's ending inventory balance. Revis is an extremely reliable professional service firm who has never made a mistake determining Bean Co.'s year-end inventory balance.

Although your firm has not yet audited Bean Co.'s ending inventory balance for this year, Revis has finished counting the year-end inventory and determined there is \$895,765 of inventory in stock (this number is reflected in the financial information you have received). Consistent with prior years, Revis found no evidence of inventory theft, shrinkage, or spoilage.

The price of Jamaican coffee beans has been remarkably steady over the past thirty years due to their limited supply and very stable demand. In fact, the price Bean Co. pays for each pound of coffee purchased from Jamaican farmers has not changed in the last four years. Similarly, Bean Co. has not changed the price it charges for each pound of coffee sold to its customers in the last four years, either.

YOUR ASSIGNMENT

Congratulations! You have been hired by a Big Four public accounting firm. Your first assignment is to help with the annual audit of Bean Co., a small coffee bean distributor located in Tampa, Florida.

While comparing the year-end account balances reported by Bean Co. to the account balances the audit partner expected to see (which is an analytical review procedure), your senior has identified an unexpected fluctuation in Bean Co.'s cost of sales ratio. Recall that the cost of sales ratio is the cost of goods sold divided by total sales. In Bean Co.'s business, the cost of sales ratio identifies what percentage of every dollar of sales is consumed by the cost of goods sold.

The engagement partner expected the cost of sales ratio to remain at 64.60%, the same level as the prior year. However, Bean Co.'s unaudited cost of sales ratio unexpectedly declined to 62.32%. Although the difference may not seem significant, the engagement partner finds it surprising because the company's cost of sales ratio has barely changed over the past five years.

Your senior has provided you with financial information attached behind these instructions. Additionally, your senior has provided you with the following guidance:

- Because your accounting firm audited Bean Co.'s financials last year, the information labeled "Last Year (Audited)" can be considered accurate, reliable, and free of error.
- The information labeled as "This Year (Expected)" has been developed by your audit partner based on his extensive industry and company experience. Consequently, the partner's expectations can be considered very reliable.
- The information labeled "This Year (Unaudited)" is client-provided financial information for the current year. The unaudited information is prepared by Bean Co. and has not yet been audited; therefore, no assurance is provided regarding its accuracy or reliability.
- Looking at the financial information provided, you will see the partner has set the threshold to 3% of the expectation. Consequently, the columns labeled "Threshold" present you with information you can use to determine if the upper or lower bounds of an account or ratio fall outside the threshold boundaries.

STEPS IN YOUR ASSIGNMENT

Your senior has asked you to conduct analytical review to investigate and explain an unexpected fluctuation: Why did Bean Co.'s cost of sales ratio unexpectedly decline this year? The Partner has already determined the account balance expectations and has set the threshold to 3% of the expectation.

Your senior has asked you to conduct the remaining 3 steps (in order):

Step 1: Examine the background information and financial information to gain an understanding of Bean Co. and of the unexpected fluctuation (15 minutes)

1. The first step is to identify and record the information cues that you feel are related to the unexpected fluctuation.
2. A good way to start this step is to consider accounts related to the cost of sales ratio to see if any of them exhibited unexpected behavior.
3. If any accounts related to the unexpected fluctuation fall outside the threshold, you might want to identify them because they may help to explain why Bean Co.'s cost of sales ratio unexpectedly declined.
4. The researcher will make an announcement when 15 minutes have passed. After the announcement, you may move to the next stage if you are ready. There is no rush, so spend as much time as you need in this stage.

Step 2: Generate possible explanations for the unexpected fluctuation (15 minutes)

5. After you have identified the information cues you believe are related to the cost of sales ratio decline, the next step is to generate reasons to explain why the unexpected fluctuation occurred.
6. Please spend 15 minutes coming up with as many reasons as possible to explain why the cost of sales ratio may have declined. The researcher will make an announcement when 15 minutes have passed.
7. After the announcement, the researcher will hand out a sheet of paper that tells you how to move to Step 3. There is no rush, so only move to Step 3 when you are ready.

Step 3: Evaluate each of your proposed reasons and select the one you believe best explains why the unexpected fluctuation occurred (No time limit)

8. In other words, review each of your proposed explanations and choose the one you feel best explains why the cost of sales ratio declined.

Bean Co. Financial Information							
This Year vs. Projections vs. Last Year							
Audit Workpapers							
Reminder: Threshold = 3% deviation from the expectation						Threshold	
Ratios	Last Year (Audited)	Expected Change	This Year (Expected)	This Year (Unaudited)	Low	High	
Inventory Turnover Ratio (Cost of Goods Sold / Ending Inventory)	3.79	None	3.79	3.65	3.67	3.90	
Cost of Sales Ratio (Cost of Goods Sold / Sales)	64.60%	None	64.60%	62.32%	62.66%	66.54%	
Accruals Ratio (Accounts Receivable / Accounts Payable)	84.37%	+ 2.63%	86.99%	103.95%	84.38%	89.60%	
Receivables Turnover Ratio (Sales / Ending Recievables)	8.62	- 0.41	8.21	8.21	7.96	8.46	
Gross Margin (Gross Profit / Sales)	35.40%	None	35.40%	37.68%	34.34%	36.46%	
Account Balances	Last Year (Audited)	Expected Change	This Year (Expected)	This Year (Unaudited)	Low	High	
Sales	\$5,002,861	+ 5.00%	\$5,253,004	\$5,253,004	\$5,095,414	\$5,410,594	
Cost of Goods Sold	\$3,231,848	+ 5.00%	\$3,393,440	\$3,273,440	\$3,291,637	\$3,495,244	
Coffee Bean Inventory	\$853,110	+ 5.00%	\$895,765	\$895,765	\$868,892	\$922,638	
Accounts Payable	\$722,140	+ 1.84%	\$735,421	\$615,421	\$713,358	\$757,483	
Coffee Bean Purchases	\$3,272,472	+ 5.00%	\$3,436,096	\$3,316,096	\$3,333,013	\$3,539,178	
General and Administrative Expenses	\$620,000	+ 3.00%	\$638,600	\$497,000	\$619,442	\$657,758	
Net Income	\$426,083	+ 8.73%	\$463,282	\$724,881	\$449,384	\$477,181	
Other Information	Last Year (Audited)	Expected Change	This Year (Expected)	This Year (Unaudited)	Low	High	
Price charged per pound of coffee sold to Bean Co's customers	\$37.10	None	\$37.10	\$37.10	\$35.99	\$38.21	
Cost per pound of coffee (Includes coffee bean purchase price, freight/shipping costs, direct/indirect labor, overhead, and all other costs required to make the coffee beans available for resale)	\$23.97	None	\$23.97	\$23.97	\$23.25	\$24.69	
Gross Profit on each pound of coffee sold (the difference between the price Bean Co's customers are charged for each pound of coffee minus Bean Co's cost per pound of coffee)	\$13.13	None	\$13.13	\$13.13	\$12.74	\$13.53	

FIGURE A1. Case Materials: Financial Information Provided to All Participants.

APPENDIX B:
ONLINE PRE-TEST

Thank you for taking time out of your schedule to participate in this accounting study. Your participation is greatly appreciated.

To begin, please enter your User Name:

Thank you for entering your user name. The next screen contains a little exercise to warm up your mind.

Please do not hit the "Next" button until instructed.

Before you start the main experiment, let's get your mind warmed up with a little reading.

The nine phrases below are considered to be some of the most famous in history. Please take 60 seconds to read the phrases below:

1. To be or not to be, that is the question.
2. All is fair in love and war.
3. Four score and seven years ago...
4. I have not yet begun to fight!
5. All that we learn from history is that we do not learn from history.
6. The only thing we have to fear, is fear itself.
7. A penny saved is a penny earned.
8. Life is the only thing worth dying for.
9. I have a dream...

(PLEASE DO NOT CLICK THE NEXT BUTTON UNTIL INSTRUCTED)

How many of the phrases from the previous page can you recall?

In the spaces below, please type in as many phrases as you can recall from the previous page:

Enter a phrase here	<input type="text"/>
Enter a phrase here	<input type="text"/>
Enter a phrase here	<input type="text"/>
Enter a phrase here	<input type="text"/>
Enter a phrase here	<input type="text"/>
Enter a phrase here	<input type="text"/>
Enter a phrase here	<input type="text"/>
Enter a phrase here	<input type="text"/>
Enter a phrase here	<input type="text"/>

Now I would like to ask you five (5) accounting questions. Please answer the following questions below:

Which of the following would occur when a sale is made on account to a customer? (In other words, when a non-cash sale occurs).

Please select all that apply:

- Accounts payable increases
- Accounts payable decreases
- Sales revenue increases
- Sales revenue increases, but only after payment is received
- Accounts receivable increases
- Accounts receivable decreases

Which of the following would occur when a sale is made and the product is delivered?

Please select all that apply:

- Sales revenue increases
- Sales revenue increases, but only when payment is received
- General and administrative expenses increase
- General and administrative expenses increase, but only when actually paid

Which of the following would occur when a company receives a payment from a customer who owed money on account?

Please select all that apply:

- Cash increases
- Sales revenue increases
- Accounts receivable increases
- Accounts receivable decreases
- Accounts payable increases
- Accounts payable decreases

Which of the following would occur when a company pays its employees?

Please select all that apply:

- Cash increases
- Cash decreases
- Payables increases
- Payables decreases

Which of the following would occur when a company purchases inventory from a vendor with cash?

Please select all that apply:

- Accounts payable increases
- Accounts payable decreases
- Cash increases
- Cash decreases
- Cost of goods sold increases

Thank you for your responses. Please click the next arrow and wait until everybody is finished to start the next phase.

APPENDIX C:
TRAINING MATERIALS

These training materials were presented orally to participants

Now that your minds are warmed up, I would like to give you a little bit of background on analytical review and analytical review procedures: what the two are, the steps involved, and how new auditors often conduct them. After I briefly discuss the distinction between analytical review and analytical review procedures, I will also go over a short case before the main task begins.

If you have any questions during this little session, please feel free to ask them at any time.

Definition of Analytical Review and Analytical Review Procedures

Although the terms *analytical review procedures* and *analytical review* may seem identical, there is an important difference between the two.

- Analytical review procedures are used by auditors to reveal anomalies and departures from auditor expectations
 - Some examples of analytical review procedures are ratio analysis and trend analysis.
 - Analytical review procedures identify unexpected fluctuations
 - Using the example from above, the auditor may have used trend analysis, an *analytical review procedure*, to discover that sales increased more than expected
- Analytical review is the diagnostic process of investigating and resolving unexpected fluctuations.
 - For example, if sales increased more than the auditor expected, he or she might decide to conduct analytical review to find out what caused the increase.

So what purpose does analytical review serve?

- Analytical review is one of many fieldwork tests used to determine the validity of an account's balance.
- Analytical review is a way to test client-provided financial information by comparing it against expectations developed by the auditor.
- In other words, the auditor forms a general idea of what he or she believes client account balances should be and then compares the expectation to the client-reported balances.
- Analytical review is a useful fieldwork test because it can identify areas of the client's financial statements that do not match what the auditor expects to see.

The Analytical Review Process

To conduct analytical review account balance expectations first need to be set:

- The auditor develops an expectation of an account's balance based on his or her knowledge of the client, general business and economic conditions, and industry experience.
 - The auditor's expectation will generally be in the form of an interval estimate.
 - The more precise the auditor's expectation is, the smaller the interval will be. The smaller the interval, the more useful the analytical review test is because a smaller interval makes it more likely that material differences from the expectation will be uncovered.
 - When the client reported balance falls outside of the estimated interval, this represents a material deviation worthy of further investigation.
 - If the auditor cannot develop a precise enough estimate or a small enough interval, the auditor will typically not decide to use analytical review.
 - In public accounting firms, the partner usually develops expectations because the partner has the most client and industry experience.

- Since the partner typically sets expectations, when you conduct analytical review today, you will be given the partner's expectation.

- A simple example of setting an expectation could be something like this: Suppose the partner on the Tootsie Roll audit expects sales growth of about 4%. Thus, if sales were \$100 last year, the partner would expect to see sales around \$104 this year.

- Naturally, the auditor generally does not tell the client his or her balance expectations.

Next, an overall threshold level needs to be selected:

- The threshold level is used to identify what deviations from the auditor's expectation should be investigated. Account balances that are above or below the threshold limits are considered to deviate from the auditor's expectation.

- For simplicity's sake, assume the auditor sets the threshold at 3% of the expectation. This means that an account balance that fluctuates less than plus or minus 3% from the expectation is not considered to deviate from the expectation.

- For example, suppose a firm's sales were expected to be \$200. The threshold amounts would be set at plus and minus \$6. Thus, if sales were reported below \$194 or above \$206 the account would be considered to deviate from the expectation.
- Since the partner typically sets the threshold based on a variety of risk factors, when you conduct analytical review today, you will be given the partner's threshold. You will also be given the upper and lower bounds of the threshold.
- For the same reasons the auditor does not disclose the expectation to the client, the auditor generally does not disclose the threshold to the client.

Then, accounts that exceed the threshold limits are identified:

- The auditor will investigate differences that exceed the threshold bounds since they indicate unexpected activity.
- Drawing upon the 3% threshold we established earlier, suppose the partner expects sales to be \$200. If client reported sales were \$207 sales would deviate from the expectation, but if they were \$204, sales would not be considered to deviate from the expectation.
- During an audit engagement, it usually falls to the audit senior to identify account balances and accounting ratios that exceed the threshold bounds. When an account is found to violate the threshold bounds, this is referred to as finding an "unexpected fluctuation."
- After the audit senior identifies an unexpected fluctuation, the actual investigation usually falls to a staff auditor. The staff auditor will be tasked with finding out why the unexpected fluctuation occurred.
- Since the senior usually hands the investigation of to a staff auditor, in today's study you will assume the role of a newly hired staff auditor asked to find out why an unexpected fluctuation occurred.

Once an unexpected fluctuation has been identified, the first step in explaining the fluctuation is to become knowledgeable about the client and the unexpected fluctuation. To do this, the staff auditor examines the information available.

- The auditor often looks at accounts related to the unexpected fluctuation for cues that might help explain why the unexpected fluctuation occurred.
 - If accounts related to the unexpected fluctuation are also out-of-threshold, the auditor may identify these accounts as potentially related to the unexpected fluctuation.

After the auditor becomes familiar with the information available, he or she uses this understanding to come up with potential causes of the unexpected fluctuation:

- The auditor will often refer to accounts related to the unexpected fluctuation to try to find a pattern among them that can explain why the unexpected fluctuation occurred.
- Since many things can potentially cause an unexpected fluctuation, the auditor generally comes up with more than one potential explanation.
- Although this stage may be a little unclear to you at this point, don't worry... it is one of those things that is easier to understand through example. I will go over a short sample case after we discuss the next step.

After the auditor comes up with as many potential explanations as possible, he or she evaluates each explanation and then chooses the one he or she believes best explains the unexpected fluctuation's cause.

- To do this, the auditor evaluates each of his or her proposed explanations against the information the auditor felt was related to the unexpected fluctuation. The auditor usually picks the explanation that best matches the cues the auditor feels are related to the unexpected fluctuation.

Now that I have given you a brief overview of the analytical review process, I want to go over a quick example.

- The purpose of the example is to give you a feel for conducting analytical review. Since I am giving the example, the speed with which I'll go over the case will be much faster than when you conduct analytical review on your own.

- Do not worry about this - it has nothing to do with your ability and everything to do with the fact that I have already worked the case and don't need any time to identify the patterns and potential explanations.

Analytical Review Case

<<Experimenter Note: Be sure to have the Training Application running here>>

Background: Bolt Co. is a manufacturer of scissor bolts based in Chicago, Illinois. The company sells only one product: Scissor bolts, which are used to fasten the two blades required to make a pair of scissors. Bolt Co.'s industry is classified as very mature with low annual growth of about 3% per year. The scissor bolt industry has experienced very slow growth over the past 30 years and experts predict the same steady, slow growth for the foreseeable future.

Your firm has audited Bolt Co. for the past ten years so the engagement partner has a very good understanding of the firm and the industry in which the company operates. Consistent with industry trends, the partner expected Bolt Co. to experience about 3% growth in the current year. Further, the partner set the threshold on the engagement to 3% of the expectation.

By analyzing the percentage change in the number of Bolt Co.'s new customers, which is an analytical review procedure, the senior on the engagement was very surprised to find that Bolt Co. gained 13 new customers this year. This represents a 13% increase and is especially surprising since the scissor industry is so stable and the overall industry demand for scissor bolts increased only 3% (as expected) this year. The senior has asked you to find out why the client gained so many new customers this year.

Let's turn to the financials to conduct analytical review!

<<PUT BOLT CO. FINANCIALS ON OVERHEAD HERE>>

Bolt Co. Financial Information
This Year vs. Projections vs. Last Year
Audit Workpapers

Reminder: Threshold = 3% of the expectation

Ratios	Last Year (Audited)	Expected Change	This Year (Expected)	This Year (Unaudited)	Threshold	
					Low	High
Cost of sales ratio (Cost of Goods Sold / Sales)	50.00%	+ 0%	50.00%	50.00%	48.50%	51.50%
% of customers with outstanding balances over 90 days old (Number of customers with balances over 90 days / All customers with outstanding balances)	2.00%	+ 0%	2.00%	6.00%	1.94%	2.06%

Account Balances	Last Year (Audited)	Expected Change	This Year (Expected)	This Year (Unaudited)	Threshold	
					Low	High
Sales	\$242,718	+3%	\$250,000	\$282,500	\$242,500	\$257,500
Accounts payable	\$100,000	+3%	\$103,000	\$103,000	\$99,910	\$106,090
Accounts receivable	\$52,000	+3%	\$53,560	\$65,000	\$51,953	\$55,167
Bad debt expense	\$7,282	+3%	\$7,500	\$16,950	\$7,275	\$7,725
Sales, general, and advertising expense	\$72,500	+0%	\$72,500	\$72,500	\$70,325	\$74,675
Net income	\$35,000	+3%	\$36,050	\$38,500	\$34,969	\$37,132

Other Information	Last Year (Audited)	Expected Change	This Year (Expected)	This Year (Unaudited)	Threshold	
					Low	High
Number of new customers placing orders	0	+ 0	0	13	0	0
Number of customers placing orders	100	+ 0	100	113	97	103
Average size of each order	\$485	+3%	\$500	\$500	\$485	\$515
Average number of orders placed by each customer	5	+ 0	5	5	4.85	5.15
Number of products offered	1	+ 0	1	1	0.97	1.03
Credit manager	John Malone	None	John Malone	John Malone	N/A	N/A
Overall scissor bolt industry growth	3.00%	+3%	3.00%	3.00%	2.91%	3.09%
Average price of each scissor bolt	5 cents	+ 0%	5 cents	5 cents	4.85 cents	5.15 cents

FIGURE C1. Training Case Materials: Financial Information Presented to All Participants During Training.

Explanation of Financial Terms

Before we delve too deeply into trying to figure out why Bolt Co. gained so many new customers compared to expectation, let's take a moment to familiarize ourselves with Bolt Co.'s financial information, which is very similar to the format of the information you will receive in the case you will be asked to work.

Notice that there are three categories of information: (1) ratios, (2) account balances, and (3) other information.

- Each category provides you with a different type of information that you may find useful in determining what caused the unexpected fluctuation (in Bolt Co.'s case, the unexpected fluctuation we are trying to explain is why Bolt Co. attracted so many new customers this year).

- The information provided by each of the information categories may help you come up with potential explanations for the unexpected fluctuation or may help you to rule out some of your other explanations.

You'll notice there are five columns in the spreadsheet. Let's discuss these five column headings:

Last Year (Audited): This information is the result of last year's audit and, therefore, can be considered completely reliable.

Expected Change: This represents the partner's expected change this year compared to last year. Since these expectations are generated by the partner, you can consider them completely reliable.

This Year (Expected): This is the account balance expectation developed by the engagement partner. This expectation can be considered completely reliable.

This Year (Unaudited): This is information provided to your accounting firm by the client. The information has not been audited, so there is no assurance regarding its accuracy or reliability.

Threshold (High/Low): These amounts represent the boundaries of the threshold. If an account's balance lies within the two, the account is not deemed to deviate from the expectation. If an account's balance is below the low or above the high, the account is considered to deviate from the partner's expectation.

Sample Case

Now that you understand the format and content of the information being provided to you, I will delve into finding out what caused the unexpected fluctuation.

Keep in mind that other members of the audit team have already set account balance expectations, set the threshold (at 3%), and identified the unexpected fluctuation that needs to be identified (the number of new clients increased by 13 this year which was surprising because the partner did not expect Bolt Co. to gain any new clients this year).

9. Let's examine the information to gain an understanding of Bolt Co. and the unexpected fluctuation:
 - Remember, we are trying to find cues that might give us some insight as to what is causing the unexpected fluctuation.
 - A good start is to consider accounts related to new customers to see if any of them exhibited unexpected behavior.

- If any accounts related to the unexpected fluctuation are out of tolerance, we probably want to identify them because they could help explain why Bolt Co. attracted so many new customers

10. Thus, the first thing I am going to do is make a note of those accounts that are related to the unexpected fluctuation and are out of threshold.

Looking over the information available:

- I know that sales are related to new customers, so I'm going to count the fact sales increased by 16% as relevant to explaining why Bolt Co. may have gained new customers.
 - Thus, I am going to make a note of this in the application
 - USING APPLICATION: record this cue by pressing the correct button and typing "Sales increased by 16%, much greater than the 3% expected increase"
- I know that the percent of customers with outstanding balances over 90 days may be related to new customers (since more customers means some of those new customers probably will not pay Bolt Co. in a timely manner). However, if Bolt Co. is attracting the same quality of customers, I wouldn't expect this figure to jump as much as it did (200%).
 - USING APPLICATION: record this cue by pressing the correct button and typing "Percentage of customers with balances over 90 days increased by 200%, much greater than the 2% expected increase"
- I would expect accounts receivable to increase with new customers, so I'm going to consider the fact that accounts receivable increased by 25% as relevant – especially since I would expect accounts receivable to increase by 13% if Bolt Co. gained 13% new customers... the fact that it is up 25% is a bit odd especially since the order size among customers has remained consistent.
 - USING APPLICATION: record this cue by typing "Accounts receivable increased by 25%, much greater than the 3% expected increase"
- I would expect bad debt expense to increase proportionally with new customers, so I'm surprised it is up 133%
 - USING APPLICATION: record this cue by typing "Bad debt expense was up 133%, much greater than the 3% expectation"

11. Now that I have identified the information cues I believe are related to the reason the number of customers increased, the next step is to generate explanations for why the number of customers increased so much. Because I know you are getting tired of hearing me talk I'm only going to propose 3 explanations, although there are many more possible:

- Explanation #1: USING APPLICATION, type in "Bolt Co. cut the price of its bolts causing customers to switch to Bolt Co. from other scissor bolt manufacturers, increasing sales"
- Explanation #2: USING APPLICATION, type in "Bolt Co. spent a lot more on advertising so it attracted new customers"
- Explanation #3: USING APPLICATION, type in "Bolt Co. loosened its credit policy, resulting in new, lower-credit quality customers buying from Bolt Co."

Now that I have identified some potential reasons Bolt Co. may have gained so many new customers, I am going to evaluate each proposed explanation and select the one that I believe best explains why Bolt Co. gained so many new customers:

1. My first proposed explanation was that Bolt Co. cut the price of its bolts, causing new customers to order from it. However, looking at the information provided, I see that the average price per bolt sold remained at 5 cents. Therefore, it does not seem that Bolt Co. cut its price, so I no longer believe this is the reason Bolt Co. obtained so many new customers.
2. My second proposed explanation was that Bolt Co. spent a lot more money on advertising, leading to the attraction of new customers. Although I don't really have any direct information on advertising expense, I do see that sales, general, and advertising expense did not increase. This indirectly suggests Bolt Co. did not spend more money on advertising. Thus, based on indirect evidence, I tend to believe more advertising is not the reason for the new customers.
3. My third proposed explanation is that Bolt Co. may have loosened its credit policy. Looking over the information and cues I believe to be relevant, I do see that bad debt expense jumped 133% and I would expect to see this happen if the credit policy was eased because a looser credit policy means weaker-credit quality customers are approved.

I also see that the percentage of customers with balances over 90 days old increased, which I would expect to see if Bolt Co. was taking on customers with weaker credit. Further, I see that accounts receivable jumped 25%, something I would expect to see if companies with weaker credit were approved – those new

customers are probably more likely to buy products on credit, rather than pay cash – and since they are less credit worthy, they probably take longer to pay their bills, as well.

Looking over the rest of the information, it doesn't look like the credit manager changed, so that means its not the credit manager, so it could be the policy. There seems to be a lot of evidence suggesting this is the reason for the increase in the number of new customers.

- Thus, after reviewing all the information available and thinking matters over, I decide that I believe the reason Bolt Co. gained so many new customers this year was because it eased its credit policy. Now that I have selected a reason to explain the cause of the unexpected fluctuation, analytical review is complete.
 - USING APPLICATION, select this reason as the “best explanation” then show users how to exit out of application.

One thing that you will notice about analytical review is that there isn't necessarily a “smoking gun” that clearly indicates what caused the unexpected fluctuation. Instead, conducting analytical review requires you to use your knowledge of accounting and account relationships to come up with possible explanations.

APPENDIX D:
ONLINE POST-EXPERIMENT QUESTIONNAIRE

(All Interventions)

Thank you for completing my experiment. I'm now going to ask you a few questions about the task you just completed.

Please answer honestly - there are no right or wrong answers. Your responses are completely confidential and will greatly help advance accounting research.

To begin, please enter your User Name:

How easy or difficult was it for you to identify the pieces of information related to the cost of sales ratio decline?

	Very Easy	Easy	Somewhat Easy	Neither Easy or Difficult	Somewhat Difficult	Difficult	Very Difficult
.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How easy or difficult was it for you to use your knowledge of accounting relationships to identify information related to the cost of sales ratio decline?

	Very Easy	Easy	Somewhat Easy	Neither Easy or Difficult	Somewhat Difficult	Difficult	Very Difficult
.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How easy or difficult was it for you to see the accounting relationships associated with the cost of sales ratio decline?

	Very Easy	Easy	Somewhat Easy	Neither Easy or Difficult	Somewhat Difficult	Difficult	Very Difficult
.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How easy or difficult was it for you to come up with potential reasons to explain why the cost of sales ratio declined?

	Very Easy	Easy	Somewhat Easy	Neither Easy or Difficult	Somewhat Difficult	Difficult	Very Difficult
.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How easy or difficult was it for you to come up with THE correct reason the cost of sales ratio declined?

	Very Easy	Easy	Somewhat Easy	Neither Easy or Difficult	Somewhat Difficult	Difficult	Very Difficult
.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I would now like to get some feedback about the task. Please remember there are no right or wrong answers.

How important was it for you to find the correct reason Bean Co.'s cost of sales ratio declined?

	Not at all Important 1	2	3	Average Importance 4	5	6	Extremely Important 7
.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

(ARD Intervention and Combined-Aid Intervention)

The following questions relate to the diagram that was handed to you during the experiment

	Not at All 1	2	3	Somewhat 4	5	6	A Great Deal 7
Overall, how much did you rely on the diagram to conduct the task?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How much did the diagram help you to generate potential reasons to explain why the cost of sales ratio declined?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How much did the diagram help you to evaluate each of your proposed reasons?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How much did the diagram help you to select the best reason the cost of sales ratio declined?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall, how helpful was the diagram?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall, how much did you like the diagram?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

(PCA Intervention and Combined-Aid Intervention)

The following questions relate to specific aspects of the software program you used to conduct analytical review during the experiment

	Not at All 1	2	3	Somewhat 4	5	6	A Great Deal 7
Overall, how much did you rely on seeing the pieces of information you identified as relevant to conduct the task?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How much did seeing the pieces of information you identified as relevant help you to generate potential reasons to explain why the cost of sales ratio declined?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How much did the checkboxes help you to evaluate each of your proposed reasons?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How much did the checkboxes help you to select the best reason the cost of sales ratio declined?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

(PCA Intervention and Combined-Aid Intervention)

Overall, how helpful was seeing the pieces of information you identified as relevant during the task?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall, how helpful were the checkboxes?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall, how much did you like seeing the pieces of information you identified as relevant?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall, how much did you like the checkboxes?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

(All Interventions)

Which stage of the task did you like the most?

- The training session provided before the task began
- Learning about Bean Co.'s background and operations
- Identifying which items contributed to the decline in Bean Co.'s cost of sales ratio
- Coming up with reasons that could explain why Bean Co.'s cost of sales ratio declined
- Evaluating each of your proposed reasons to determine how likely each one was the actual cause of the cost of sales ratio decline
- Choosing the proposed reason that best explained why Bean Co.'s cost of sales ratio declined

Which stage of the task did you like the least?

- The training session provided before the task began
- Learning about Bean Co.'s background and operations
- Identifying which items contributed to the decline in Bean Co.'s cost of sales ratio
- Coming up with reasons that could explain why Bean Co.'s cost of sales ratio declined
- Evaluating each of your proposed reasons to determine how likely each one was the actual cause of the cost of sales ratio decline
- Choosing the proposed reason that best explained why Bean Co.'s cost of sales ratio declined

How much did you like the Excel program you used to conduct analytical review?

	Not at All 1	2	3	Somewhat 4	5	6	A Great Deal 7
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Overall, how much did you enjoy today's task?

	Not at All 1	2	3	Somewhat 4	5	6	A Great Deal 7
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please indicate what materials were available to you while conducting analytical review today (Check all that apply):

- Background information about Bean Co.
- A flowchart depicting common accounting relationships
- An Excel application / workbook
- An electronic check-box screen that helped you evaluate your proposed explanations

Okay, now I would like to ask you five (5) accounting-specific questions. Please answer the questions below:

When a company uses a periodic inventory system, how is cost of goods sold (COGS) calculated at the end of the year?

- COGS = Ending inventory + inventory purchases - beginning inventory
- COGS = Beginning inventory + inventory purchases - ending inventory
- COGS = Ending inventory - inventory purchases + beginning inventory
- COGS = Beginning inventory - inventory purchases + ending inventory

Which of the following transactions would make accounts payable decrease?

Please select all that apply:

- Goods are sold to a client on account
- Inventory is purchased from a vendor with cash
- Cash is paid to a vendor for services received a month ago
- Inventory is purchased from a vendor on account
- A customer returns goods they purchased on account before they paid for the goods

In a periodic inventory system, how is the balance of inventory usually determined at year-end?

- A physical count of the inventory remaining on hand is taken at the end of the year
- All the inventory purchases for the year are added up to determine year-end inventory
- All the inventory sales for the year are added up to determine how much inventory remains on hand
- The accountants check the inventory ledger for the year end balance since the inventory records are updated every time an inventory purchase or inventory sale occurs

Which of the following would occur when inventory is purchased from a vendor on account?

Please select all that apply:

- Accounts receivable increases
- Accounts receivable decreases
- Accounts payable increases
- Accounts payable decreases
- Cost of goods sold increases
- Cost of goods sold increases after paying for the inventory

How do inventory purchases relate to cost of goods sold expense?

Please select all that apply:

- Directly: Every time inventory is purchased cost of goods sold expense increases
- Directly: Every time inventory is purchased and paid for cost of goods sold increases
- Indirectly: Every time a sale is made the cost of goods sold increases
- Indirectly: Every time a sale is made and payment is received the cost of goods sold increases

I'm now going to ask you for some information about yourself. Please answer honestly - your responses are completely confidential and will greatly help advance accounting research.

Again, there are no right or wrong answers.

Please indicate your academic status

- I am an undergraduate junior
- I am an undergraduate senior
- I am in the accounting masters program

Please select all accounting courses you have completed (check all that apply):

- | | |
|--|--|
| <input type="checkbox"/> Intermediate Financial Accounting I | <input type="checkbox"/> Audit I |
| <input type="checkbox"/> Intermediate Financial Accounting II | <input type="checkbox"/> Audit II |
| <input type="checkbox"/> Intermediate Financial Accounting III | <input type="checkbox"/> Accounting Information Systems |
| <input type="checkbox"/> Advanced Financial Accounting | <input type="checkbox"/> Advanced Accounting Information Systems |
| <input type="checkbox"/> Cost Accounting and Control I | <input type="checkbox"/> Contemporary Issues in Auditing |
| <input type="checkbox"/> Cost Accounting and Control II | <input type="checkbox"/> Integrative Accounting Seminar |

Please select all accounting courses you are currently enrolled in (check all that apply):

- | | |
|--|--|
| <input type="checkbox"/> Intermediate Financial Accounting I | <input type="checkbox"/> Audit I |
| <input type="checkbox"/> Intermediate Financial Accounting II | <input type="checkbox"/> Audit II |
| <input type="checkbox"/> Intermediate Financial Accounting III | <input type="checkbox"/> Accounting Information Systems |
| <input type="checkbox"/> Advanced Financial Accounting | <input type="checkbox"/> Advanced Accounting Information Systems |
| <input type="checkbox"/> Cost Accounting and Control I | <input type="checkbox"/> Contemporary Issues in Auditing |
| <input type="checkbox"/> Cost Accounting and Control II | <input type="checkbox"/> Integrative Accounting Seminar |

Please tell me about any internships you have had:

- I do not have any internship experience
- I have Big Four public accounting internship experience
- I have non-Big Four public accounting internship experience
- I have internship experience, but it was not at a public accounting firm

If you have internship experience, how many months of experience do you have (in total)?

Have you ever conducted analytical review before?

- Yes

- No

Are you interested in pursuing a career as an auditor in public accounting?

- Yes
- No

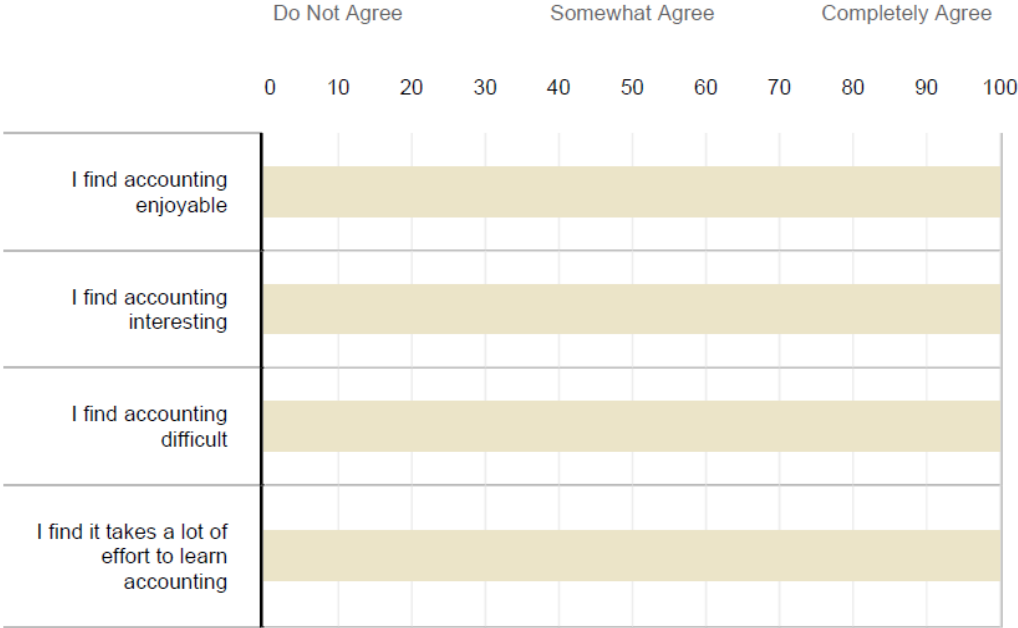
Are you a male or female?

- Male
- Female

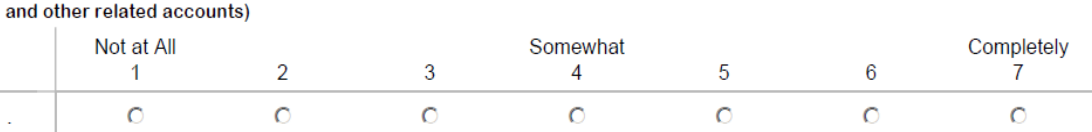
If you had to guess, how many different pieces of information can you think about simultaneously? (In other words, how many different pieces of information can you hold in your mind at one time?)

	1 piece	2 pieces	3 pieces	4 pieces	5 pieces	6 pieces	7 pieces	8 pieces	9 pieces	10 pieces	11+ pieces
.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Using the scale below, please tell me how much you agree with the following statements:



How well do you understand the inter-relationships that exist between accounts? (In other words, how well do you understand the various ways an account like sales can impact other accounts such as cash, accounts receivable,



What is your overall grade point average (GPA) in the accounting classes you have taken? (If you do not know, just give me your best estimate):

- 1.0 - 1.4
- 1.5 - 1.9
- 2.0 - 2.4
- 2.5 - 2.9
- 3.0 - 3.4
- 3.5 - 4.0

Please select the unexpected fluctuation you investigated today:

- The unexpected fluctuation involved Bean Co.'s bad debt expense
- The unexpected fluctuation involved Bean Co.'s fixed assets
- The unexpected fluctuation involved Bean Co.'s cost of sales ratio
- The unexpected fluctuation involved Bean Co.'s quick ratio

This is the last part of the questionnaire. Now you have the opportunity to give me some comments about this study. I'd love to hear whatever you would like to say.

To enter a comment, just type it in the text box below. All your comments are completely confidential.

Thank you again for your participation.

Please feel free to tell me anything you would like below:

APPENDIX E:

ORAL SCRIPT DELIVERED TO PARTICIPANTS DURING THE MAIN TASK

Stage 1: Case Familiarization and Pattern Recognition Stage (All treatments)

Oral Instructions (All treatments)

This study examines how new auditors perform analytical review. You will be asked to read a short case. Then, you will conduct analytical review to find the cause of an unexpected fluctuation. Please read and evaluate the information to the best of your ability.

- I don't want to ruin the surprise, but you all have been hired by a Big Four public accounting firm.
- In the yellow information package, you will see that you have two pages of background information, a page of financial information, and a page containing your assignment from the audit senior.
- The financial information that you have received in the package is also in electronic format in the Excel application
- Let's all open the Excel application together. Please click on the Excel file on your desktop.
- [AFTER LOGIN SCREEN POP UP] The login screen will pop up. Please enter your user name and password. You'll see a screen with a few buttons and some financial information.

Stage 1: Your Assignment

(All treatments, **all treatments except the no-aid condition**, no-aid condition)

Your senior has asked you to spend 15 minutes looking over the audit workpapers to become familiar with Bean Co.

As you look over the workpapers, keep in mind the company's cost of sales ratio was lower than expected. During this stage, your senior has asked you to record any information cues you feel were related to the cause of the decline in the cost of sales ratio by clicking the button labeled "Click here to enter a piece of information associated with the decrease in the cost of sales ratio." [SHOW BUTTON AND DEMONSTRATE HOW TO ENTER PIECE OF INFORMATION].

Please enter each information cue that you believe is related to the unexpected fluctuation separately. Thus, each information cue should be entered alone by clicking the "Click here to enter a piece of information associated with the decrease in the cost of sales ratio" button.

After you enter information cues, you can also edit them by clicking the button labeled "Click here to review / edit the pieces of information associated with the decline in the cost of sales ratio." [SHOW BUTTON AND DEMONSTRATE HOW TO EDIT / DELETE]'

As you look over the workpapers, keep in mind the company's cost of sales ratio was lower than expected. Please identify and record the information cues (account balances,

ratios, or other information) that you believe are related to the decline of the cost of sales ratio.

Please take 15 minutes to get familiar with the case information and identify the information cues you feel are related to explaining why the cost of sales ratio declined.

- Remember, in this stage you are not trying to come up with explanations about why the unexpected fluctuation occurred, rather you are just trying to get familiar with Bean Co. and identify information cues you feel are associated with the cause of the cost of sales ratio decline.

<<At this point the proctor waits 15 minutes. After 15 minutes, Stage 1 ends.>>

Stage 2: Explanation Generation Stage

(All treatments, *activity relationship diagram treatment*, pattern-consideration aid treatment)

Please click the button labeled “Stage 2” at the top of your application. The password is: “accounting.”

Now that you are familiar with Bean Co., your senior has asked you to take at least 15 minutes to come up with as many potential explanations as you can to explain why Bean Co.’s cost of sales ratio declined. Once you have come up with a potential explanation, please type it into the application by pressing the button labeled “Please click here to enter a potential explanation for the decrease in the cost of sales ratio.” [SHOW BUTTON AND HOW TO DO THIS]

You will notice that when you press this button you are shown the cues you identified as associated with the cost of sales ratio decline. Because you felt these cues are related to the unexpected fluctuation, you might find it helpful to refer to these cues when coming up with potential explanations for the cause of the cost of sales ratio decline.

To help you come up with potential explanations for the cause of the cost of sales ratio decline, you might find it helpful to refer to the diagram handed to you during this stage. The diagram might help you come up with potential explanations by making it easier to see the links between the unexpected fluctuation and the related accounts.

The diagram might also help you identify information related to the unexpected fluctuation that you did not record in the last stage. Please remember to record information you believe may be related to the unexpected fluctuation by pressing the button labeled “Click here to enter a piece of information associated with the decrease in the cost of sales ratio.” [SHOW BUTTON AND DEMONSTRATE HOW TO ENTER PIECE OF INFORMATION].

Please note that during this stage you may still add information cues you believe are associated with the reason the cost of sales ratio declined if you would like to do so. Your senior also encourages you to save your file a few times during the engagement.

After 15 minutes has passed, I will give you a handout containing the instructions for moving to Stage 3. However, this does not mean you must immediately proceed to Stage 3. Please take your time and move to Stage 3 only when you are ready.

APPENDIX F:
WRITTEN INSTRUCTIONS GIVEN TO PARTICIPANTS IN STAGE THREE

Stage 3: Explanation Evaluation and Selection Stage

(All treatments, *pattern-consideration aid*, activity relationship diagram)

Now that you have finished developing a list of potential explanations for the decline in the cost of sales ratio, your senior has asked you to evaluate your explanations and choose the one that best describes why the cost of sales ratio declined.

- To do this, please click the button labeled “Click here to select the best explanation for the decrease in the cost of sales ratio.”
- To help you evaluate each of your potential explanations for the cost of sales ratio decline, you might find it helpful to refer to the diagram handed to you during the last stage. Because the diagram displays the relationships between Bean Co.’s operating activities and accounts, it might make it easier to see how well each of your proposed explanations addresses the cost of sales ratio decline.
- *After clicking the button, you will notice the screen displays your proposed explanations, a column labeled “Supporting Pieces,” and a column labeled “Best Explanation” (the last 2 columns will be blank, for now)*
- *Before you can select the best explanation you need to evaluate **each** of your proposed explanations against the information cues you identified as related. To do this:*
 - Double click on a proposed explanation
 - *This will take you to another screen where you are asked to “check” the box of each piece of information the proposed explanation you are evaluating addresses.*
 - *Please check the box next to each piece of information that supports your proposed explanation.*
 - *After you check the appropriate boxes please hit the green “Save” button.*
 - *You will see that you are taken back to the explanation evaluation screen, but now the column labeled “Supporting Pieces” tells you how many pieces of information support the explanation you just evaluated.*
 - *Now, please do the same thing for your other proposed explanations.*
 - *After each of your proposed explanations has a number in the “Supporting Pieces” column, you can select the explanation you believe is the actual cause of the cost of sales ratio decline.*
 - *Before choosing an explanation that describes why the cost of sales ratio declined, feel free to double click any of your proposed explanations to review the information cues you felt supported it.*
 - *When you are ready to choose an explanation double click on it.*
 - *The check box screen will pop up again.*

- *Click the button labeled “Select as Best Explanation.” A pop up box will ask you to confirm your selection.*
- *You will then be taken back to the main screen*

After choosing the explanation you feel best describes why the cost of sales ratio declined, your senior has asked you to rate each of your explanations in terms of how likely you feel each explanation is the actual cause of the cost of sales ratio decline.

- To do this, please click the button labeled “Click here to assess the probability your explanations have identified the reason the cost of sales ratio decreased.”
- You will see all your proposed explanations, including the one you selected.
 - *You will also see the number of information cues you felt supported each proposed explanation.*
 - Type in a probability for each of your proposed explanations. The probability you enter is how likely you think the proposed explanation is the actual cause of the cost of sales ratio decline.
 - Please note that the probabilities you assign do not have to sum to 100%.
 - Thus, you could assign one of your explanations an 80% chance it was as the actual cause and give another explanation a 75% chance it was the actual cause of the cost of sales ratio decline.
 -
- When you have finished, please click the “End” button. Excel will close.
- Please raise your hand once the application has closed and I will direct you to the last part of the experiment.