UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

FACTORS INFLUENCING WORKING BACKWARDS AMONG AUDITORS IN THE APPLICATION OF NONSTATISTICAL DECISION AIDS

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

STEPHANIE FAREWELL
Norman, Oklahoma
2001

UMI Number: 3014525

UMI®

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

FACTORS INFLUENCING WORKING BACKWARDS AMONG AUDITORS
IN THE APPLICATION OF NONSTATISTICAL DECISION AIDS

A Dissertation APPROVED FOR THE
MICHAEL F. PRICE COLLEGE OF BUSINESS
(SCHOOL OF ACCOUNTING)

BY

ACKNOWLEDGEMENTS

I wish to express my gratitude to the faculty who have helped in the culmination of this dissertation. Each of you has contributed a unique perspective to this project. Especially the members of my committee: G. Lee Willinger (Chairman), Stephen Butler, Michael C. Knapp, Marlys G. Lipe and Scott Gronlund. Special thanks goes to my Chairman, G. Lee Willinger, for being willing to endure the trials and tribulations of a dissertation not within his area of interest. In addition, a huge thank you goes to Professor Stephen Butler for giving so freely of his time throughout my doctoral program. You have contributed more to my perspective on accounting research than any other single person.

I am also grateful to Grant Thornton, LLP for providing access to the auditors who participated in this research. George Krull and Emmett Day are owed individual thanks for their invaluable assistance in the revision and administration of the research instrument.

In addition, I want to thank my fellow Ph.D. students at the University of Oklahoma for their comments are encouragement throughout this project. Special thanks goes to Lori Mason who always knew we would make it through, even when I had doubts.

Thanks, also goes, to my entire family for supporting my efforts during the past eight years. To my mother, Hazel Warshum, for always having faith and confidence that through perseverance anything can be achieved. To my sons, Justin and Jacob, for enduring the highs and lows of a Ph.D. program even though they cannot yet understand what it means. Finally, thanks goes to my husband, Rick, for supporting my efforts and being willing to put his academic studies on hold while I completed mine.

A project such as this is not the work of the author alone; therefore, I again say thank you to everyone who has been involved.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract

For more than a decade, research has demonstrated decision makers' unwillingness to rely on decision aids. However, extant research has most often studied the use of unfamiliar decision aids. For example, Kachelmeier and Messier (1990) and Messier, Kachelmeier, and Jensen (2001) provide evidence of auditors working backwards in the use of an unfamiliar decision aid. The current study builds on extant research by using process tracing software in an experimental setting to test whether auditors have learned from a decision aid used in practice and how learning affects decision aid use in relation to the use of an unfamiliar decision aid. Auditors used either a familiar decision aid, a modified familiar decision aid, an unfamiliar decision aid or a modified unfamiliar decision aid to compute a nonstatistical sample size. Results indicate that decision aid familiarity does not affect the auditors decision to work backwards, approximately 30 percent of auditors in all conditions worked backwards to alter the decision aid sample size. Further, working backwards has the affect of eliminating the significant, negative relationship that should exist between the number of controls tested and the substantive sample size. It also appears that auditors acquire task specific knowledge from general audit experience rather than task specific experience with the decision aid.

1.0 Introduction

Auditors use substantive sampling to obtain evidence to support their audit opinion. Auditing standards [see for example SAS 39 (AICPA 1981) and SAS 47 (AICPA 1983)] give auditors guidance on designing audit samples and evaluating the results obtained from those samples. Additionally, the AICPA and some individual audit firms provide auditors with decision aids for computing nonstatistical sample sizes. Kachelmeier and Messier (1990) and Messier, Kachelmeier, and Jensen (2001) investigate the effect the AICPA decision aid has on the calculation of nonstatistical sample sizes. They provide experimental evidence that auditors work backwards, circumventing, what is to the auditors, an unfamiliar decision aid. A working backwards strategy, in this case, is characterized by changing the required decision aid input factors until the output (i.e., sample size) is consistent with the auditor's intuitive assessment of sample size.[1] If working backwards is pervasive among auditors, they may not be achieving the desirable effects that a decision aid may provide.[2]

For example, if auditors use their firm's decision aid inappropriately, they may be relying on a sample size that differs from the firm's definition of a

[1] Ellis and Hunt (1993, 270; see also, Matlin 1983; Newell, Shaw, and Simon 1958; and Wickelgren 1974) define working backwards as "seeing what the solution ought to look like and then working backwards from the solution to the current problem state." In a broader sense it consists of breaking the problem into subgoals, which facilitates problem solving. However, the path from the current state to the goal state in the AICPA decision aid is shorter (e.g., two steps) than most psychology applications; therefore, the operationalization of working backwards for this research is altering the decision aid inputs.

[2] Desirable effects that decision aids may provide include: (1) increased consistency in the combination of decision aid inputs (Einhorn 1972), and (2) increasing consensus across individuals as the decision aid focuses each user's attention on "normatively relevant information resulting from training in the use of the aid" (Libby and Libby 1989, 737).

rational sample size given the client's characteristics.[3] People, in general, are not good intuitive statisticians (Meehl 1957). Tversky and Kahneman (1971) point out that individuals often place too much reliance on small samples. If auditors circumvent the decision aid to rely on an intuitive sample size that is smaller than the decision aid sample size, then the audit firms are potentially bearing additional audit risk.[4] This increased audit risk exposes the audit firm to a heightened probability of litigation if an auditee's stakeholders discover a previously undisclosed error. Alternatively, an inefficient audit would occur if an auditor chooses to circumvent the decision aid and rely on an intuitive sample size that is larger than the decision aid sample size. Firms are facing competitive pressure to eliminate audit inefficiencies. Therefore, if auditors use a decision aid inappropriately, an ineffective or inefficient audit may result.

The working backwards behavior observed by Kachelmeier and Messier (1990) and Messier et al. (2001) may be due to their participants' unfamiliarity with the AICPA's decision aid. In practice, auditors use a decision aid on multiple occasions, rather than only a single instance, which may mitigate the working backwards behavior observed in these studies. By using the decision aid on multiple occasions, auditors gain experience with the decision aid, thereby acquiring knowledge for calculating nonstatistical sample sizes with the decision aid. This accumulated knowledge, referred to as a knowledge structure, could

---

[3] The firm has expended time and effort to develop a nonstatistical sample size decision aid that aids auditors in calculating what the firms believes to be a rational sample size given the client's characteristics. The decision aid also provides a potential defense in the event of litigation.
[4] Audit firms may have other ways, such as increased reliance on analytical procedures, to minimize audit risk.

contain information on the required decision aid parameters, the form of the input parameters, how the decision aid manipulates the inputs and what combination of inputs will lead to the desired sample size output.

This study uses process tracing software to experimentally investigate auditors' working backwards behavior. Audit seniors and managers analyzed the case used by Kachelmeier and Messier (1990) and Messier et al. (2001), which facilitates direct and analogous comparisons between their studies and the results of this investigation.[5] After analyzing the case, participants determined a nonstatistical sample size using a decision aid. Because decision aid familiarity may eliminate working backwards behavior, aid familiarity is manipulated in a between-participants design. The auditors received either their firm's nonstatistical sample size decision aid or the AICPA nonstatistical sample size decision aid contained in the 1999 *Audit Sampling* practice release used by Messier et al. (2001). In an effort to further our understanding of what an auditor may learn through the use of a decision aid, auditors are tested on their knowledge of a familiar decision aid component.

This study makes three contributions to the current body of knowledge on decision aids. Foremost, this is the first study to provide direct evidence, through the use of process tracing software, on the use of a working backwards strategy.[6] Second, this study tests whether aid familiarity influences an auditor's observable

---

[5] Appendix A contains the case and the two post-experimental questionnaires.

[6] This is in contrast to 'indirect' evidence provided by Kachelmeier and Messier (1990) and Messier et al. (2001). They compared the intuitive sample size of one group of auditors to the mechanically computed sample size of a group of auditors that provided only the decision aid inputs (i.e., the dollar amount of tolerable misstatement, the degree of reliance on other audit procedures to test the same objective, and the assessment of inherent and control risk) without knowledge of the decision aid model for combining the decision aid inputs.

processing strategy. Third, this study tests decision aid knowledge acquisition.

The remainder of the paper is organized as follows. The next section provides background. Section Three presents the theory and develops hypotheses. The experimental methodology is included in Section Four. Section Five presents the results, while Section Six provides a discussion of the conclusions and implications of this work. The final section discusses potential future research.

2.0 Background

The goal of the audit is to express an opinion on whether the client's financial statements reflect its financial position and the results of operations in accordance with GAAP. The third standard of fieldwork requires the auditor to examine sufficient evidence to provide a basis for the audit opinion (AICPA 1972). Evidence may be obtained by examining the entire population or by sampling. If the auditor decides that sampling is to be used, the auditor is then faced with the decision of how to determine the appropriate sample size. The auditor may choose to apply statistical or nonstatistical sampling. In statistical sampling the auditor decides the maximum level of acceptable sampling risk, which is the probability that the auditor will reach an incorrect conclusion because the population was sampled rather than examined in its entirety. The chosen level of sampling risk is mathematically incorporated into the determination of sample size. Alternatively, nonstatistical sampling relies on an auditor's intuitive assessment, rather than a normative model, to determine the sample size necessary to reach a conclusion about whether the population under examination satisfies the audit objectives. The auditor may apply nonstatistical (i.e., judgmental) sampling if it is decided that the costs of statistical sampling outweigh the benefits. Because nonstatistical sampling does not require auditors to explicitly quantify inputs, such as sampling risk, it is less time consuming and, hence, less costly than statistical sampling (AICPA 1999, 11).

The use of nonstatistical sampling presents possible problems in the audit environment. Because sampling risk is not explicitly incorporated into the sample size decision, auditors may be introducing an unacceptable level of audit risk into the audit process. Audit risk is the probability that the auditor will conclude that the financial statements are fairly stated when, in fact, they are not. An unacceptable audit risk level would occur if the auditor selected a sample size that was too small to obtain the desired degree of assurance. Assurance is a probability expressing how confident an auditor wants to be in the conclusion reached about the financial statements.[7] Conversely, if the sample size is too large, the auditor may be decreasing audit risk at the cost of an inefficient audit. Either way, the use of nonstatistical sampling poses potential problems.

The process of review may mitigate some of the potential for auditors to choose inappropriate nonstatistical sample sizes. For example, if a senior chooses a nonstatistical sample size that is too small (large), when the audit manager reviews the senior's sample size determination it may be suggested that the sample size be increased (decreased). In addition, peer review may enhance learning from decision aids. For example, Libby (1995) finds that supervisory review is effective because it promotes the learning of causal relations.

In addition to the review process another way that public accounting firms can constrain an individual's choices is through the development and use of decision aids.[8] Decision aids can improve the decision making process by minimizing the effects of judgmental biases (e.g., Dawes, Faust, and Meehl

---

[7] Assurance is the complement to audit risk.

[8] I use Rohrmann's (1986, 365) definition of a decision aid. Rohrmann defines a decision aid as "any explicit procedure for the generation, evaluation and selection of alternatives that is designed for practical application and multiple use."

1988).[9] An accountant's experiences, learning, and perceptions may contribute to judgmental biases (Gibbins 1984). Judgmental biases may result because the auditor, at any given time, is both subjectively perceiving the current environment and trying to assimilate those perceptions within the framework of an existing schema (i.e., knowledge structure) (Gibbins 1984). It is the subjective perceptions that often lead individuals to apply decision rules inconsistently (Gibbins 1984).

Decision aids can improve decision making by increasing consistency in the application of decision rules. Einhorn (1972) and Bonner, Libby, and Nelson (1996) demonstrated the benefit of permitting individuals to select the information cues and then using a mechanical model to combine the cues into a judgment. By using a mechanical model to combine cues consistency is increased.

While the benefits that decision aids provide have been demonstrated in a variety of situations, an equally large body of research has documented decision makers' unwillingness to rely on decision aids. Many factors may influence a decision maker's willingness to rely on a decision aid. Several factors have been empirically examined, for example, justification (Ashton 1990 and 1992), financial incentives (Arkes, Dawes, and Christensen 1986; Ashton 1990), performance feedback (Ashton 1990; Davis and Kotteman 1995), and decision maker task expertise (Arkes, Dawes, and Christensen 1986). In general,

[9] One way that decision aids may improve the decision process is by decomposing a global judgment into composite judgments. The sub-tasks focus the auditor's attention on normatively relevant information.

individuals have resisted the use of mechanical models because of a "perceived attack on decision makers' prized judgmental ability" (Libby 1981, 105).[10]

Most extant decision aid reliance research has studied the use of decision aids that were made available to participants to use if they so chose. However, within some audit firms (e.g., Grant Thornton 1996) the use of firm specific decision aids is required as they represent an integral part of the audit work papers. Although an auditor is reluctant to rely on the decision aid, it cannot simply be disregarded in favor of an intuitive sample size assessment. For example, Grant Thornton requires auditors who deviate from the decision aid sample size by 10 or more to write a memo justifying the deviation.[11] However, an auditor may choose to manipulate the decision aid inputs to achieve the desired sample size output.[12] The manipulation of decision aid inputs is indicative of an auditor's reluctance to rely on a required decision aid.

2.1 AICPA Nonstatistical Sampling Decision Aid

The AICPA has recently revised the nonstatistical decision aid contained in the 1983 *Audit Sampling* guide.[13] The AICPA 1999 nonstatistical sample size decision aid and its 1983 predecessor use Probability Proportional to recorded Size (PPS) sampling as their basis (AICPA 1999, 52; AICPA 1983, 74-76).[14]

---

[10] Over-reliance on decision aids has also been shown; however, over-reliance most often occurs with novice decision makers (for example, Glover, Prawitt, and Spilker 1997)

[11] This is not expected to bias the results because there is no reason for the firm's auditors to believe that this latitude would be unavailable with the AICPA decision aid.

[12] Cuccia, Hackenbrack, and Nelson (1995) find a similar effect in a tax setting.

[13] A copy of the revised decision aid is contained in Appendix B.

[14] Appendix C discusses Probability-Proportional-to-Size sampling in detail.

PPS sampling incorporates sampling risk into the determination of sample size by increasing the likelihood that larger items are sampled. If there are misstatements in larger items, they are more likely to result in a material misstatement of the financial statements.

The revised decision aid requires auditors to determine two decision aid inputs: tolerable misstatement and an assurance factor. The calculation of tolerable misstatement is primarily a mechanical task to determine the amount of account misstatement that can occur without indicating material misstatement in the financial statements. The 1999 revision incorporates a tolerable misstatement worksheet. In calculating tolerable misstatement the auditor first determines whether total assets or total revenues are greater, locates the appropriate percentage from the table in the tolerable misstatement worksheet, and then performs the specified mathematical computation.

The 1999 revision significantly affects the determination of the assurance factor. The assurance factor attempts to associate two audit parameters with the desired degree of sample assurance. The assurance factor is based on the auditor's subjective assessment of inherent and control risk and the auditor's reliance on other audit procedures to test the same audit objectives. Another revision in the 1999 Auditing Practice Release was the modification of the assurance factor values. The values previously ranged from 1.5 to 6 and in the revised Auditing Practice Release range from 1 to 3. This change in parameter values, in and of itself, results in smaller sample sizes when using the current decision aid. The change in parameter values may be the AICPA's attempt to recognize a decreased reliance on substantive test sampling in favor of increased reliance on analytical procedures and tests of controls during audit engagements.

2.1.1 Tests of the AICPA Decision Aid

Kachelmeier and Messier (1990) used a draft of the 1983 *Audit Sampling* guide and Messier et al. (2001) used a draft of the 1999 *Audit Sampling* Auditing Practice Release to examine the effect that the AICPA nonstatistical sampling decision aid has on the determination of sample size. They had audit seniors and managers analyze a case and determine a nonstatistical sample size. The auditors were assigned to one of several conditions. Auditors determined the sample size without mention of a model (i.e., an intuitive sample size), determined the sample size given the AICPA decision aid, or provided the parameters for the AICPA model without knowledge of how the parameters would be combined. Other variables manipulated are not relevant to this investigation (i.e., population size and exclusion of the AICPA tolerable misstatement worksheet).

A key finding of Kachelmeier and Messier (1990) and Messier et al. (2001) is that auditors appear to work backwards when using the AICPA decision aid. The evidence of working backwards comes from comparing the sample size given by auditors who used the AICPA decision aid to the mechanically computed sample size of auditors who provided only the AICPA decision aid model parameters (i.e., tolerable misstatement and the assurance factor). The mechanically computed sample size of the "parameters-only" condition was significantly larger than the sample size given by auditors in the decision aid condition. The authors interpreted this finding to mean that the auditors in the decision aid condition had worked backwards (i.e., manipulated the decision aid input parameters to achieve the desired sample size output).

The use of a specified model ensures that the same parameters are consistently combined; however, individual differences may exist in the selection of the decision aid parameter values from the tables contained in the AICPA decision aid (e.g., the AICPA decision aid assurance factor table has 16 possible values). Table 1 reproduces the AICPA assurance factor table. The selection of decision aid parameter values is based on the subjective evaluation of the audit client. Because the evaluation is subjective, different auditors may choose different descriptors to describe the same audit client. For example, in assessing the same audit client, one auditor may feel the reliance on other audit procedures is moderate, while another auditor feels the reliance is substantial. Differences in the selection of factors from the tables may contribute to working backwards. If the auditor believed the decision aid was contributing to under- or over-auditing, the decision aid inputs may be altered to change the resulting sample size output. Working backwards can be accomplished by altering the subjectively determined assurance factor. For example, ceterus paribus, changing the reliance on other audit procedures from substantial to moderate when the assessment of control risk is slightly below maximum will change the sample size from 36 to 47.

Table 1
Assurance Factors for Nonstatistical Sampling Formula

ASSESSMENT OF CONTROL RISK (AND INHERENT RISK)	RELIANCE ON OTHER RELEVANT AUDITING PROCEDURES			
	None	Little	Moderate	Substantial
Maximum	3.0	2.7	2.3	2.0
Slightly below maximum	2.7	2.4	2.0	1.6
Moderate	2.3	2.1	1.6	1.2
Low	2.0	1.6	1.2	1.0

Because the only extant results are indirect, the next step is to directly observe, through the use of process tracing, whether auditors work backwards.[15] Therefore, one of the goals of this paper is to provide direct evidence on the existence, or lack thereof, of the working backwards behavior suggested by Kachelmeier and Messier (1990) and Messier et al. (2001).

2.2 The Firm's Nonstatistical Sample Size Decision Aid

A national audit firm's nonstatistical sample size decision aid was used as the familiar decision aid for this research.[16] This decision aid was chosen because it is characteristically similar to the revised AICPA decision aid. Both decision

[15] It is also possible that auditors mentally calculate and manipulate sample size, which would be unobservable.
[16] The aid is familiar to the research participants.

aids require the auditor to select decision aid input factors from multiple tables and combine the inputs according to a pre-specified model.[17] As evidenced in Table 2, the firm decision aid is as complex, or arguably more complex, than the AICPA decision aid. The number of mathematical manipulations contributes to the firm decision aid's complexity. The firm decision aid's complexity may inhibit learning. However, the firm decision aid has more steps because the firm decision aid decomposes the task further than the AICPA decision aid. Task decomposition may facilitate learning. Any differential task complexity, in this research, is minimized through the use of computer-performed calculations and researcher choice in selecting the firm decision aid parameter factor for environmental assessment (Table 2, Step 2d). In addition, if the AICPA decision aid is easier to use, this would bias toward the null hypothesis, as explained later.

To use the firm decision aid the auditor first uses a worksheet to determine tolerable misstatement. The firm decision aid examines all items that are individually material and samples the remainder of the population.[18] To determine the appropriate sample size the auditor selects parameters from three tables: a sample size adjustment factor, an environmental assessment factor, and an expected aggregate error factor. The sample size adjustment factor adjusts the base sample size for reliance on related substantive audit procedures and the assessed level of control risk. The environmental assessment factor adjusts the

[17] The minimum sample size of the firm and AICPA decision aid is 10 and 22, respectively. The maximum sample size of the firm and AICPA decision aid is 138 and 83, respectively. The minimum and maximum sample sizes are based on total assets of $50,783,000. Potential floor and ceiling effects are eliminated by non-extreme case characteristics. The mean sample size in Messier et al. (2001) is 30.

[18] To avoid confounding effects the auditors in the firm decision aid condition will be told that no individual items are material.

base sample size for the effect of environmental factors, significant audit evidence from other audit procedures when auditing nominal accounts, or if the audit objective is population completeness. The expected aggregate error factor adjusts the sample size for errors that are expected in the sampled population. The final step is to perform the calculations to arrive at the sample size.

Table 2
Steps in the AICPA and Firm Decision Aids

AICPA DECISION AID	FIRM DECISION AID
1. Determine tolerable misstatement **a. Determine the greater of total assets or total revenues.** b. Locate amount "a" in Materiality Table c. Subtract lower bound from "a". d. Multiply the result from step "c" by the factor in the Materiality Table. e. Add "d" to the minimum planning materiality. f. Multiply planning materiality determined in step "e" by 2/3 to determine tolerable misstatement. 2. Determine sample size a. Divide the population book value (given in case) by tolerable misstatement. **b. Determine the assurance factor based on the client's combined inherent and control risk and planned reliance on other audit procedures.** c. To determine the sample size, multiply "a" by "b".	1. Determine tolerable misstatement **a. Determine estimated total assets.** **b. Determine estimated total revenues, but not less than total assets.** c. Locate amount "b" in Touchstone Table. d. Subtract lower bound from "b". e. Multiply "d" by table factor. f. Add "e" to the minimum tolerable misstatement for the level of total assets or total revenues. 2. Determine sample size a. Divide population book value (given in case) by tolerable misstatement. b. Multiply step "a" by 5. **c. Determine the sample size adjustment factor based on the client's control risk and evidence provided from related substantive procedures.** d. Determine environmental assessment factor from the environmental assessment worksheet, based on environmental factors and evidence from other audit procedures. e. Subtract the environmental assessment factor from the sample size adjustment factor, but not less than zero. f. Multiply the result of step "b" by step "e". g. Determine expected aggregate error factor based on the expected aggregate error and the revised sample size adjustment factor in "e" h. To determine the sample size, multiply "g" by "f". i. Round the sample size up to a whole number, but not less than ten.

Note: Steps in bold are performed by the auditor. Steps "d" and "g" for determining sample size with the firm decision aid are determined by the researcher and the remaining steps are mathematically computed by the computer program.

3.0 Theory and Hypothesis Development

3.1 Decision Aid Reliance

Research has generally shown that individuals are often unwilling to rely on decision aids (for example, Ashton 1990, 1992; Boatsman et al. 1995; Dawes 1979; Dawes et al. 1989; Kleinmuntz 1990; Peterson and Pitz 1986; Whitecotton 1996). Arnold, Collier, Leech, and Sutton (1998) summarize the findings of several decision aid reliance studies, finding decision aid reliance by inexperienced individuals and mixed evidence for reliance by individuals with moderate and high task experience. While some of the research on decision aid reliance has used a repeated-measures design (e.g., Ashton 1990), extant research may not have captured the full potential of a participant's learning through the use of a decision aid. By not examining decision aid usage over an extended period of time, researchers are ignoring a potentially important aspect of the natural environment in which judgments are made. Eining, Jones, and Loebbecke (1997) suggest that "the influence of decision aids might vary with continued use" and that "it would be interesting to examine reliance over an extended time period" (1997, 17). Arnold and Sutton (1998) identify three criteria that are critical to achieve decision aid reliance by experienced individuals. One of the criteria is the user's familiarity with the decision aid through previous usage.[19] Familiarity with

[19] The other two criteria are task complexity and cognitive fit.

the decision aid provides the potential for the decision aid to become a part of the auditor's decision making process.

3.2 Knowledge Acquisition

Anderson (1980) indicates that individuals systematically acquire knowledge. Knowledge may be acquired intentionally or incidentally. Intentional knowledge acquisition begins when the individual is actively involved in learning/memorizing task rules. The process continues with the individual deliberately applying the rules to a task until it can be performed with little or no conscious effort (i.e., automatic processing). Incidental, or passive, knowledge acquisition occurs without conscious effort on the part of the learner (Postman 1964).

The accounting profession relies heavily on experiential learning to foster expertise development (Bonner and Walker 1994; Libby 1995). Decision aids represent one opportunity for experiential learning. Ashton and Willingham (1988) and Libby and Luft (1993) argue that a structured decision aid facilitates task-related learning by providing a model of the knowledge components necessary for task completion. However, Glover, Prawitt, and Spilker (1997) point out that if the decision aid is used passively, knowledge acquisition may be inhibited. Knowledge acquisition would be inhibited by the possible low level cognitive processing from passive decision aid use.

3.3 Hypothesis One

In practice, an auditor potentially acquires two types of knowledge by using a decision aid to determine nonstatistical sample sizes. The auditor acquires general task knowledge and decision aid specific knowledge, both of which may influence how an auditor uses a decision aid in practice. In general, through the use of the decision aid, the auditor learns about the task and the criteria the firm believes should be considered in the determination of nonstatistical sample sizes. In addition, by using a decision aid on multiple occasions, the auditor has the potential to learn the structure of the decision aid and to develop an understanding of how a particular decision aid works. Specifically, the auditor could be expected to learn the required decision aid inputs, the rules that the model applies to manipulate the decision aid inputs, and the effect that changing the input parameter values has on the decision aid sample size. When the decision aid is familiar the auditor may have to apply little or no conscious effort to use the decision aid. The ability to complete the task with little or no conscious effort reflects an automatic level of processing. The automatic processing of the task (i.e., determining a nonstatistical sample size with a decision aid) will lead to observable working forward, rather than backward, in the use of the decision aid.

A knowledge structure is not static, it is continually being refined. Feedback during the review process provides one basis for the refinement of a knowledge structure for the task of determining a nonstatistical sample size. For example, an auditor may refine his or her task knowledge structure if superiors indicate during review that an inappropriate sample size had been selected.

A knowledge structure for the task of determining a nonstatistical sample size is a prerequisite to working backwards. An existing task knowledge structure allows an auditor to calculate an intuitive nonstatistical sample size. The knowledge structure may be based on learning from the firm decision aid, feedback during review, firm training or other experiences.[20] An auditor's knowledge structure guides the search for information and influences how the information is combined into a sample size judgment.

When using a new, unfamiliar decision aid the sample size may be counterintuitive to the auditor's nonstatistical sample size based on his or her existing knowledge structure. A counterintuitive sample size may occur when using an unfamiliar decision aid because an auditor does not have knowledge about the decision aid. For example, an auditor would not know how the decision aid parameter values would impact the final sample size.

Two factors which may contribute to obtaining a counterintuitive sample size with an unfamiliar decision aid are: (1) a lack of familiarity with the verbal probability categories used to select the decision aid parameter values, and/or (2) a lack of knowledge about the model used to manipulate the decision aid parameters into a sample size. If the probability phrases are unfamiliar to the user they may not be interpreted as intended by the decision aid developer.[21] For example, the AICPA may have used the phrase moderate to reflect around 75%

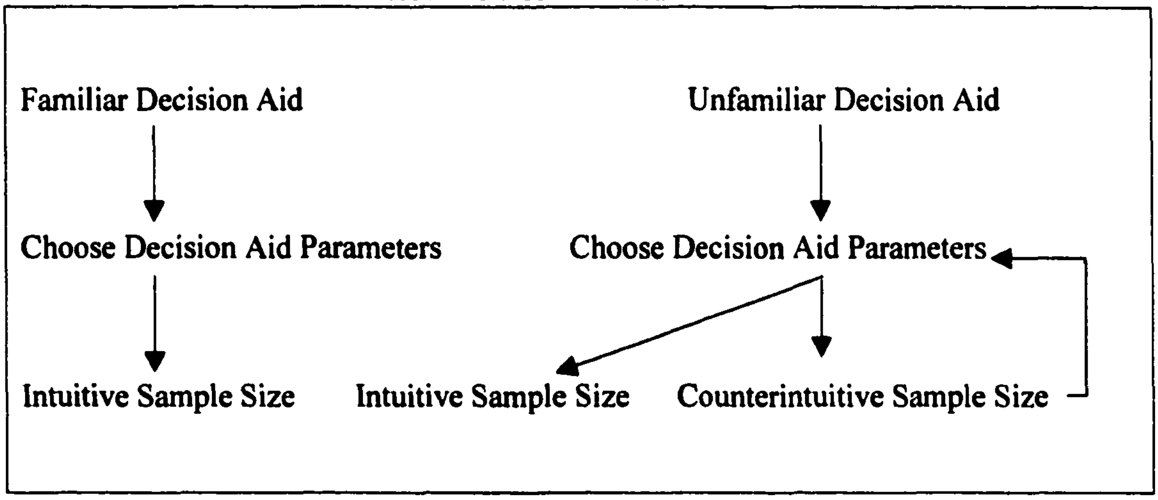
---

[20] Kachelmeier and Messier (1990) encountered a firm effect, whereby the difference between the intuitive and decision aid sample sizes was smaller for auditors from a firm that had its own nonstatistical sampling decision aid than the difference in the intuitive and decision aid sample sizes of auditors from a firm that did not have a nonstatistical sampling decision aid. They posited that this may have occurred because the auditors had internalized their firm's decision aid.

[21] Appendix D discusses prior research on verbal probability theory.

reliance on other audit procedures; however, as a result of firm training, the auditors may interpret moderate to mean 50%. Because the auditor's choice of verbal probability descriptors, from the assurance factor table, significantly impacts sample size the decision aid sample size may be counterintuitive.[22] If the auditor feels that the sample size is counterintuitive there may be an increased level of working backwards with an unfamiliar decision aid.[23] Figure 1 depicts the difference in how a familiar and an unfamiliar decision aid are used.

Figure 1
Patterns of Decision Aid Use[24]



[22] The sample size may be counterintuitive for several reasons, for example the auditor may not believe the sample size corresponds to the desired level of audit assurance.
[23] In addition to working backwards because the sample size is counterintuitive to an auditor's knowledge structure sample size, the hierarchical review structure of the audit process may influence working backwards. This may occur because the review process increases an auditor's accountability. It is only if an auditor chooses to **not** work backwards and eliminate a counterintuitive sample size that a problem would be discovered in the review process. By working backwards this does not imply that the decision aid is wrong, or that it is inappropriate for the task, just that the decision aid is inconsistent with the auditor's knowledge structure which is contingent on firm-specific training, experience and decision aids.
[24] For the auditor with no knowledge of the task the sample size would be accepted, not because it was intuitive but because there was no intuitive sample size with which to compare the decision aid sample size. The participant selection procedures ensure that all participants are familiar with the task.

As previously mentioned the acquisition of knowledge occurs over time as individuals learn the rules for task completion, apply the rules, and eventually are able to apply the rules without conscious effort. This discussion of the influences of knowledge acquisition and the existence of a knowledge structure on decision aid use leads to the following hypothesis presented in the null and alternative form:

$H1_0$: Auditors who use an unfamiliar nonstatistical sample size decision aid are less likely, or equally likely, to observably work backwards than auditors who use a familiar nonstatistical sample size decision aid.

$H1_a$: Auditors who use an unfamiliar nonstatistical sample size decision aid are more likely to observably work backwards than auditors who use a familiar nonstatistical sample size decision aid.

The preceding hypothesis examines observable differences in the use of familiar and unfamiliar decision aids. While we may expect to observe differences in the use of familiar and unfamiliar decision aids, the reasons for these differences have not been empirically examined. I have posited that a reason for observed differences in how familiar and unfamiliar decision aids are used can be attributed to learning.

Audit firms do not explicate what an auditor is expected to learn through the use of a decision aid. However, Glover et al. (1997) and Rose (1998) point out that audit firms expect auditors to learn from the decision aids they use in practice. Little research has empirically examined what an auditor does learn from a decision aid (Glover et al. 1997; Rose 1998). Glover et al. (1997) point out that any potential learning may be inhibited if the decision aid is used passively. This firm's decision aid appears to require active auditor involvement through multiple judgment-based decisions. For example, the auditor has to actively think about the client's internal controls and the evidence that will be provided by other substantive procedures in order to provide the assessments the decision aid requires. This deep level of processing may facilitate incidental learning of the decision aid.

3.4 Hypothesis Two

Waller and Felix (1984) suggest that the type of encounter (i.e., instruction or direct experience) may affect the knowledge actually acquired. In addition, Marchant (1990, 23) contends that experiences only provide opportunities for learning. As previously mentioned learning may occur either intentionally or incidentally. Therefore, an auditor's knowledge level may depend on the number of opportunities they have had to learn from the decision aid.[25] It is not possible to specify ex ante the number of times an auditor would need to use a decision aid to acquire knowledge. Because learning exists along a continuum, participants familiar with the firm decision aid may differ in the extent of their decision aid knowledge.[26] Hence, an auditor may have a global understanding of the decision aid without having specific knowledge of the decision aid components.

---

[25] The auditors will self-report the number of times they have used their firm's nonstatistical sample size decision aid.

[26] Although the auditors may differ in the extent of their decision aid knowledge, it was assumed that all auditors who participated in this research would have the requisite knowledge to calculate an intuitive sample size.

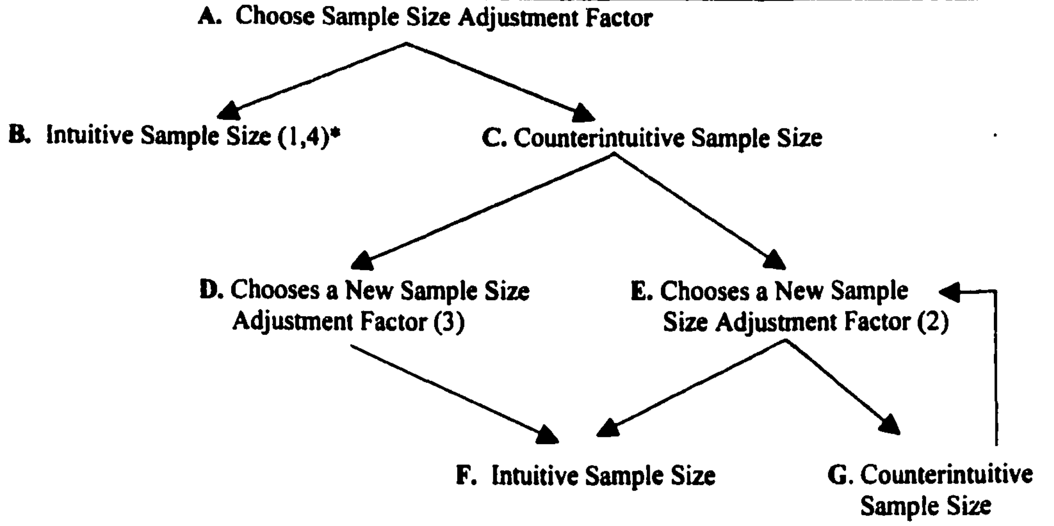
To further our understanding of what an auditor learns from a decision aid and how decision aid experience leads to knowledge acquisition we can test the auditor's learning of individual decision aid components. By changing an individual component of a familiar decision aid, making that component unfamiliar, we can test whether the auditor has learned that component of the decision aid.[27] One way to test whether the auditor has learned about the decision aid sample size adjustment factor table is to alter the presentation format of the table. For example, the rows and columns of the sample size adjustment factor table could be rotated.[28] Although the sample size adjustment factor table has been made less familiar, via its rotation, the remainder of the decision aid is familiar to the firm's auditors. Figure 2 indicates the potential paths auditors may take to arrive at their intuitive sample size and the implied level of their knowledge of the decision aid table knowledge. Figure 2 expands the number of decision aid paths an auditor may follow for the familiar decision aid in Figure 1 at the point of 'choose decision aid parameters.'

---

[27] By making a subtle change to the decision aid it is potentially made incongruent with the auditor's existing knowledge structure. Depending on how well developed the auditor's existing knowledge structure is, the auditor will or will not recognize the change.

[28] Metzler and Shepard (1974) presented subjects with pairs of rotated objects and had the subjects indicate whether the objects were the same. They found that the farther the degree of separation of the objects, the longer it took the subjects to form and rotate a mental image of the object and judge the objects the same. In the current research proposal the auditor will need to compare the physical decision aid table displayed on a computer screen to a decision aid table that is potentially stored mentally, within a knowledge structure.

Figure 2
Inset to Figure 1, Choosing Input Factors for a Familiar Decision Aid: Paths to Achieve an Intuitive Sample Size When the Sample Size Adjustment Factor Table has been Rotated



Level of Implied Knowledge of the Sample Size Adjustment Factor Table

(1) No:	These auditors read the row/column headings to find the appropriate factor.
(2) Low:	These auditors know the row/column headings; however, they do not know the factors. They do not recognize that the table is rotated.
(3) Medium:	These auditors know the row/column headings; however, they do not know the factors. When they do not achieve the expected sample size they work backwards. They would look at the headings, recognize the table has been rotated and find the correct factor.
(4) High:	These auditors know the row/column headings, when they do not find the expected factor they look at the headings, realize the table has been rotated and find the correct factor.

For the auditor with no knowledge of the task or the decision aid table, the sample size would be accepted, not because it was intuitive; but because, there was no intuitive sample size with which to compare the decision aid sample size.

In Figure 2, step A, 'choose a sample size adjustment factor,' an auditor with a high level of decision aid table knowledge may know that in row 3, column 2 the sample size adjustment factor is .34. If the sample size adjustment factor table had been rotated, when the auditor went to row 3, column 2 the factor would be .72. When the expected factor was not found the auditor's attention would be heightened, causing the auditor to go back and read the row and column headings and find the appropriate sample size adjustment factor. This would lead to the auditor's intuitive sample size (Figure 2, step B).[29] Because the auditor with high knowledge of the sample size adjustment factor table had to re-read the row and column headings, they would spend more time in choosing the sample size adjustment factor (Figure 2, step A) than auditors with less knowledge of the decision aid table.

Auditors who have a medium level of knowledge will recognize that the table is rotated when they work backwards (Figure 2, step D) to resolve the difference between the decision aid sample size and their intuitive sample size. The auditor will read the row/column headings of the table and realize the table had been rotated. The unique combination of a medium knowledge auditor working backwards and recognizing the table rotation leads the auditor to systematically work backwards. I define systematically working backwards as the auditor finding the factor that corresponds to the original row/column choice (i.e., if the auditor had assessed control risk as moderate and the reliance on other audit procedures as moderate, as indicated by row 3 and column 2, they would now find the correct factor .34 in row 2 and column 3). Therefore, systematic working backwards implies that an auditor with a medium level of knowledge of the sample size adjustment factor table would work backwards only once.

---

[29] The high knowledge auditor presumably will have internalized the meaning of the terms used to select the sample size adjustment factor. Also, the high knowledge auditor may associate certain combinations of terms with a particular sample size output.

A low level of knowledge of the sample size adjustment factor table will cause the auditor to not recognize that the table has been rotated when they work backwards in Figure 2, step E. Because the auditor does not recognize that the table has been rotated they change the combination of verbal descriptors unsystematically to alter the decision aid sample size (Figure 2, step E). The auditor who has a low level of knowledge of the table may work backwards repeatedly (Figure 2, steps C and E) to achieve an intuitive sample size (Figure 2, step F).

Auditors who have no knowledge of the table may never recognize that the sample size adjustment factor table has been rotated. A lack of table knowledge will require the auditor to read the row/column headings (Figure 2, step A). Although the particular decision aid table is unfamiliar, the auditor, presumably, has had other experiences within the firm that will allow him or her the opportunity to internalize the firm's intended meaning of the probability phrases used to describe control risk and reliance on other substantive procedures. Without prior knowledge of the decision aid table, there is not a knowledge structure in place with which a comparison to the rotated table can be made. The auditor simply works forward through the decision aid and accepts the decision aid result.[30]

In summary, auditors who have high knowledge or no knowledge of the decision aid table will work forward; whereas, the auditors with medium or low table knowledge will work backwards. However, it is only the auditors who have

[30] This is consistent with Arnold et al. (1998) who find that inexperienced individuals rely on decision aids.

a medium or high level of knowledge who will correctly identify on the post-experimental questionnaire that the sample size adjustment factor table had been rotated. These two response variables will be used to classify the auditor's measured level of knowledge. Table 4 shows how the auditor's measured level of knowledge of the sample size adjustment factor table will lead to the four possible combinations of the two response variables.

Table 3
Auditor's Measured Knowledge Level
of the Familiar Decision Aid Sample Size Adjustment Factor Table

	Works Backwards with Rotated Table	**Doesn't Work Backwards with Rotated Table**
Identifies the Table Rotation	Medium	High
Fails to Identify the Table Rotation	Low	No

This discussion on knowledge acquisition and its impact on decision aid use leads to the following hypotheses in the null and alternative form:

$H2_0$: The auditor's self-reported level of experience with the decision aid will not be positively correlated with the auditor's measured knowledge level of the familiar decision aid sample size adjustment factor table.

$H2_a$: The auditor's self-reported level of experience with the decision aid will be positively correlated with the auditor's measured knowledge level of the familiar decision aid sample size adjustment factor table.

Rejecting the null hypothesis will provide insight into what an auditor learns from this firm's decision aid, how much experience with the decision aid is necessary for such learning to occur, and how the level of knowledge influences decision aid use. This hypothesis attempts to investigate the continuum of the familiarity dimension that is ignored in the dichotomous analysis of familiarity in Hypothesis One.

3.5 Hypothesis Three

Hypothesis Two investigates whether auditors have learned an individual component of the firm decision aid (i.e., the sample size adjustment factor table from the nonstatistical sampling workpaper). It is possible that an auditor has learned the entire process that the decision aid uses to combine the auditor's assessments into a sample size judgment. If this is true, the auditor may know how the parameter values combine to form a particular sample size (i.e., the auditor may associate particular decision aid inputs with a sample size output). An auditor who has such a complete understanding of the decision aid could be classified as an expert.

It has been shown that experts often solve problems via pattern recognition (Ericsson and Smith 1991). If the auditor is an expert in using the decision aid, it may be impossible for the auditor to recall the intermediate steps the decision aid uses to arrive at sample size. Therefore, the expert may **only** know which combination of decision aid inputs leads to a particular output sample size. The auditor may no longer be able to explicate the process that is used to transform the assessment of control risk, reliance on other substantive procedures, and expected aggregate error into a nonstatistical sample size. Because of this, an expert may appear in Figure 2 to have no knowledge of the sample size adjustment factor table (i.e., the auditor does not work backwards or identify the sample size adjustment factor table rotation). However, unlike an auditor who does not have knowledge of the decision aid table, the expert will be more likely to accurately specify the decision aid sample size given the client's control risk, the degree of reliance on other substantive procedures, and the expected aggregate error rate. Again, the auditor's level of knowledge presumably depends on experience with the decision aid. This leads to the following hypothesis in the null and alternative form:

> $H3_o$: The auditor's self-reported level of experience with the decision aid will not be positively correlated with the accurate selection of the familiar decision aid's sample size given the verbal descriptors for the client's control risk, the degree of reliance on other substantive procedures, and the expected aggregate error rate.

> $H3_a$: The auditor's self-reported level of experience with the decision aid will be positively correlated with the accurate selection of the familiar decision aid's sample size given the verbal descriptors for the client's control risk, the degree of reliance on other substantive procedures, and the expected aggregate error rate.

Hypothesis Three provides an increased understanding of the level of knowledge that auditors gain through the use of decision aids.

4.0 Method

This study uses an experiment, administered via computer, to test the hypotheses. The software allows for tracking a participant's decision processes in detail. For example, the program records the time spent on the task, in total, and the time spent on each instance of working backwards. The program also records which input(s) an auditor changes when working backwards, the number of times an auditor worked backwards and the direction of the sample size revision. The data set provided by the program allows for a more definitive look at the processes that auditors apply when using decision aids. The data set provides evidence for the extent of an auditor's decision aid knowledge.

The experiment has the objectives of determining the process that auditors use in applying a decision aid, determining the effect of decision aid familiarity on aid use and ascertaining what an auditor learns from a decision aid.

The remainder of this section will include a discussion of the participants, design, task, and procedures to be employed.

4.1 Participants

Audit seniors and managers from a single international accounting firm participated in this experiment during six regularly scheduled training sessions. The training sessions were held over four months approximately three weeks apart.[31] The use of audit seniors and managers facilitates comparisons with Kachelmeier and Messier (1990) and Messier et al. (2001). While the auditors were not randomly selected from all audit seniors and managers within the firm, they were randomly assigned to the treatment conditions. The proportion of case materials distributed at each of the six training sessions was held constant:

[31] Communication about the task among auditors does not appear to be a problem. The training session that an auditor participated in was not a significant determinant of working backwards ($F=1.56$, $p=.18$).

Condition One, 25%; Condition Two, 50%; Condition Three, 12.5%; and Condition Four, 12.5%.

One hundred thirty-four auditors participated in this research study. Twenty-four participants were deleted from the data analysis because they entered an incorrect amount for 'Total Assets' and 'Total Revenue' or 'Population Book Value.' Of the Twenty-four participants that were deleted, 7 were from Training Session One, 4 were from Training Session Two, 3 were from Training Session Three, 5 were from Training Session Four, 3 were from Training Session Five, and 2 were from Training Session Six. Thus, there were 110 useable respondents.

4.2 Research Design

The study uses a 2 X 2 between-participant design. The manipulated variables are aid familiarity and table rotation. Table 4 details the experimental design. The participants were randomly assigned to one of four possible conditions: (1) the firm's decision aid, (2) the firm's decision aid with the sample size adjustment factor table rotated, (3) the revised AICPA decision aid, or (4) the revised AICPA decision aid with the assurance factor table rotated.[32]

By design, an equal number of auditors were not assigned to each Conditions. The design was planned so that when Condition 3 and Condition 4 were combined they would be approximately equal with Condition 1 for the test of Hypothesis One. Condition 2 is larger than the other conditions to allow the

---

[32] There is no reason to expect a difference in working backwards when auditors use the unfamiliar AICPA decision aid with or without a rotated table. Condition 2, the revised AICPA decision aid with a rotated table, serves to rule out a plausible alternative explanation that the revised AICPA decision aid is not only different, but very different, from the firm decision aid which leads to differential working backwards. For example, the AICPA decision aid presents control risk in the rows; whereas, the firm decision aid presents control risk in the columns. If there is no difference in the frequency of working backwards or the sample size of auditors who use the revised AICPA decision aid with or without the table rotated the two conditions will be combined for the analyses.

maximum amount of variance in the auditors' decision aid experience and measured knowledge levels of Hypothesis Two.

Table 4
Experimental Design

	Familiar Decision Aid	**Revised AICPA Decision Aid**
Non-rotated Table	Condition 1 (27 participants)	Condition 3 (11 participants)
Rotated Table	Condition 2 (59 participants)	Condition 4 (13 participants)

4.2.1 Dependent Variables

4.2.1.1 Hypothesis One

Kachelmeier and Messier (1990) and Messier et al. (2001) found that auditors worked backwards in the use of an unfamiliar decision aid. However, the experimental design did not allow the authors to capture the number of times an auditor worked backwards or how many auditors worked backwards. Because it is not possible based on prior research to predict the number of times an auditor will work backwards the dependent variable for Hypothesis One is the existence of working backwards behavior. This is measured as a dichotomous variable, depending on whether the auditor does or does not revise the decision aid nonstatistical sample size. Hence, the number of times an individual auditor works backwards is excluded in the analyses of Hypothesis One.[33]

4.2.1.2 Hypothesis Two

Hypothesis Two is a test of the association between the auditor's

[33] However, Hypothesis Two does consider the number of times an auditor works backwards. The more times an auditor works backwards, the lower the implied level of knowledge of the adjustment factor table.

measured level of knowledge and the auditor's decision aid experience. The auditor's measured level of knowledge is a function of the presence or absence of (1) working backwards and (2) correct identification of the table rotation on the post-experimental questionnaire. The auditors will self-report decision aid experience on the post-experimental questionnaire.

In addition to the main association test, time will be used as a dependent variable in the additional analysis of Hypothesis Two. It is expected that the auditor's measured decision aid knowledge will significantly affect the amount of time the auditor spends looking at the sample size adjustment factor table screen, both (1) each time they look at the screen and (2) in total. Therefore, the median time is expected to significantly differ between the measured knowledge levels.[34]

4.2.1.3 Hypothesis Three

Hypothesis Three is also a test of association. The auditors were asked to choose, from four alternatives, the familiar decision aid sample size given the tolerable misstatement, the verbal descriptors of a client's control risk, the degree of reliance on other substantive procedures, and the expected aggregate error.[35] The auditor was asked to select the sample size for five different combinations of client descriptors. The auditors were scored according to the number of correct responses (i.e., 0 to 5). The score is correlated with the auditor's self-reported decision aid experience. A positive relation is expected between experience and the accurate specification of the decision aid sample size.

[34] The mean time is used in the parametric data analyses contained in Appendix E.

[35] A recognition task was chosen because the sample size must be exactly correct, it is not sufficient to be close to the decision aid sample size. However, it is recognized, as Libby and Lipe (1992) point out that knowledge differences are more critical in recall than recognition tasks.

4.3 Task

Appendix A contains the case developed by Kachelmeier and Messier (1990), as well as the two post-experimental questionnaires that were developed for the current study. The task was to determine a nonstatistical sample size for the case presented in Kachelmeier and Messier (1990) and Messier et al. (2001). Minor changes in terminology were made in the case to make it consistent with firm terminology. It is necessary to use the same case to facilitate comparisons between the current work and extant research. As previously mentioned, the current study presents the case via computer, whereas, Kachelmeier and Messier (1990) and Messier et al. (2001) presented the case in a pencil and paper format.

The case consists of background information and a set of financial statements from a manufacturer of small consumer appliances. The account of interest is supplies inventory. The auditor's task is to determine the nonstatistical sample size for supplies inventory. The choice of nonstatistical sampling is based on the audit manager's recommendation that a nonstatistical sample was desired for the subsidiary accounts.

4.4 Procedures

Participants were given a hard copy of the case. The task (i.e., calculating a nonstatistical sample size with a decision aid) was provided on a computer diskette.[36] The auditors were able to read through the information at their own

[36] Although the presentation format differs from that used in practice this is not expected to affect the task because the auditors routinely use other computerized decision aids.

pace. The software tracked the auditor's decision aid inputs throughout the task.

After the auditors had read through the background material at their own pace, they started the computer program. The program prompted the auditor for the input factors for either their firm's decision aid or the AICPA decision aid. The computer program performed all mathematical calculations. When all inputs had been entered the computer calculated a nonstatistical sample size based on either the AICPA decision aid model or the firm decision aid model. The auditor was then asked if this is the desired sample size. If it was not the desired sample size the auditor was allowed to make changes to the sample size inputs. If the auditor made changes he or she was presented with the new sample size computation. Again, changes could be made. The number of changes was not limited. When the participant indicated satisfaction with the sample size, the screen was cleared of the sample size information and the auditor was asked to complete the demographic data and answer post-experimental multiple choice questions.

5.0 Results

The goal of this study was to first document, through process tracing, the working backwards behavior indirectly observed by Kachelmeier and Messier (1990) and Messier et al. (2001) and to examine the role of experience in auditors' propensity to work backwards in the application of a nonstatistical sampling decision aid.

The initial plan for analyzing the results of this research was to use parametric statistical analysis techniques. However, preliminary analysis of the data indicated violations in the assumptions necessary for parametric tests. Therefore, the data analysis reported is based, in general, upon nonparametric statistical analysis. For the reader interested in the results based upon parametric statistics, see Appendix E. The results based upon parametric statistical methods are essentially the same as those calculated with nonparametric statistics.

5.1 Hypothesis One

The first hypothesis examined the effect of an auditor's decision aid familiarity on working backwards in the application of the decision aid.[37] The 2 X 2 research design was planned so that Condition 3, the revised AICPA decision aid, and Condition 4, the revised AICPA decision aid with a rotated sample size adjustment factor table, could be collapsed into a single condition for the data analysis. Table 5, Panel A shows the number of auditors who did and did not

[37] While it was expected that all participants who received the firm decision aid would be familiar with the decision aid subsequent analysis of hypothesis two indicated that a large number of the participants had never used the firm decision aid. Therefore, participants who indicated that they had no experience with the decision aid were eliminated from the analysis of hypothesis one. This resulted in the elimination of 40 participants, thirteen from Condition 1 and twenty-seven from Condition 2.

work backwards in each version of the decision aid. In order to combine Condition 3 and Condition 4, as planned, it was necessary to determine whether there was a difference in working backwards with the AICPA decision aid in its original form and the AICPA decision aid with the rotated sample size adjustment factor table. Panel B of Table 5 presents the results of the planned comparisons. The decision to work backwards was not significantly related to the table rotation for the AICPA decision aid (Z=.19, p=.42; Ksa=.10, p=.50).[38] Based upon visual observation of the data it appeared that there was no difference in the propensity of auditors to work backwards when using either version of the firm decision aid; therefore, this was tested prior to testing Hypothesis One. Similar to the results obtained for the two versions of the AICPA decision aid, the sample size adjustment factor table rotation did not significantly affect the auditor's decision to work backwards with the firm decision aid (Z=-.37, p=.36). Therefore, both versions of the firm decision aid and the AICPA decision aid were pooled for the test of Hypothesis One.

[38] The results are reported using the Wilcoxon 2-sample test with a normal approximation. All significance values are for one-tailed tests. The data include both large (firm decision aid) and small (AICPA decision aid) samples. While the Wilcoxon test is more efficient for large samples, the Kolmogorov-Smirnov test is more efficient for small sample sizes (Siegel and Castellan 1988, 151). However, for the sake of consistency, Wilcoxon values are reported in the Tables for both large and small sample sizes. The results of Kolmogorov-Smirnov tests are reported in the text when appropriate for small sample sizes.

Table 5
Observed Working Backwards by Condition

Panel A: Frequency of Working Backwards by Condition

	Firm Decision Aid (Condition 1)	Firm Decision Aid with Rotated Table (Condition 2)	AICPA Decision Aid (Condition 3)	AICPA Decision Aid with Rotated Table (Condition 4)	Total
Did Not Work Backwards	10	21	8	10	49
Did Work Backwards	4	11	3	3	21
Total	14	32	11	13	70

Panel B: Planned Comparisons

	Z	p-value
Firm Decision Aid Rotated Table versus Firm Decision Aid	-.37	.36
AICPA Decision Aid Rotated Table versus AICPA Decision Aid	.19	.42

Auditors, experienced in the determination of a nonstatistical sample size, who used an unfamiliar decision aid were expected to work backwards more often than auditors who used a familiar decision aid. Based upon the aforementioned finding that the sample size adjustment factor table rotation did not affect working backwards, Conditions 1 and 2 were compared with Conditions 3 and 4 for the test of Hypothesis One. Table 6, Panel A shows the frequency of working backwards in the unfamiliar (AICPA) and familiar (firm) decision aid conditions. Panel B of Table 6 shows the statistical results of the planned comparison. The results of Hypothesis One indicate that contrary to expectations, auditors who used the unfamiliar AICPA decision aid did not work backwards more often than the auditors who used the familiar firm decision aid ($Z=-.64$, $p=.26$). Therefore, Hypothesis One was not supported. Decision aid familiarity did not significantly impact the decision to work backwards.

On average, across all versions of the nonstatistical sampling decision aid, thirty percent of the participants worked backwards. While, the significance of thirty percent of the auditors working backwards cannot be empirically tested from the expected rate of no working backwards using a binomial test, it appears evident that this represents a significant proportion of working backwards.

Table 6

Observed Working Backwards between Familiar and Unfamiliar Decision Aid Conditions

Panel A: Frequency of Observed Working Backwards

	Familiar Decision Aid	Unfamiliar Decision Aid	Total
Did Not Work Backwards	31 (67%)	18 (75%)	49 (70%)
Did Work Backwards	15 (33%)	6 (25%)	21 (30%)
Total	46	24	70

Panel B: Planned Statistical Comparison

	Z	p-value
Familiar Decision Aid versus Unfamiliar Decision Aid	-.64	.26

5.1.1 Implications of working backwards on sample size

Given the propensity of auditors to work backwards it is important to assess the implications of the decision to work backwards. The decision to work backwards has implications for audit effectiveness and efficiency through the substantive sample size. Therefore, it is relevant to discuss the impact of working backwards on the decision aid output, sample size.

The mean initial sample sizes, prior to any working backwards, are shown in Table 7, Panel A. Visual observation of the data prompted a statistical comparison of the differences in the sample sizes across the decision aid conditions. Table 7, Panel B presents the results of the statistical comparisons. The initial sample size did not vary significantly between the two versions of the firm decision aid (Z=.08, p=.47) or the two versions of the AICPA decision aid (Z=.18, p=.43; KSa=.31, p=.50). However, the initial sample size did vary significantly between the firm and AICPA decision aids (Z=4.58, p<.0001).

Table 7

Initial Sample Size by Condition

Panel A: Descriptive Statistics for the Initial Decision Aid Sample Size by Condition

	Firm Decision Aid (Condition 1)	Firm Decision Aid with Rotated Table (Condition 2)	AICPA Decision Aid (Condition 3)	AICPA Decision Aid with Rotated Table (Condition 4)
Mean Sample Size	21.11	21.49	39	37.85
Standard Deviation	12.67	12.88	18.76	12.5
Minimum	10	10	0	23
Maximum	52	52	70	56
N	27	59	11	13

Panel B: Statistical Comparisons

	Z	p-value
Firm Decision Aid Rotated Table versus Firm Decision Aid	.08	.47
AICPA Decision Aid Rotated Table versus AICPA Decision Aid	.18	.43
Both versions of the Firm Decision Aid versus both versions of the AICPA Decision Aid	4.58	<.0001

Table 8

Final Sample Size by Condition

Panel A: Descriptive Statistics

	Firm Decision Aid (Condition 1)	Firm Decision Aid with Rotated Table (Condition 2)	AICPA Decision Aid (Condition 3)	AICPA Decision Aid with Rotated Table (Condition 4)
Mean Sample Size	19.59	20.27	35.45	37.85
Standard Deviation	10.66	12.21	22.45	12.50
Minimum	10	10	0	23
Maximum	44	52	70	56
N	27	59	11	13

Panel B: Statistical Comparisons

	Z	p-value
Firm Decision Aid Rotated Table versus Firm Decision Aid	.04	.48
AICPA Decision Aid Rotated Table versus AICPA Decision Aid	-.06	.48
Both versions of the Firm Decision Aid versus both versions of the AICPA Decision Aid	4.47	<.0001

In addition to the initial sample sizes, the final sample sizes, after any working backwards, were compared. Table 8, Panel A presents the mean final sample sizes by condition, while Panel B presents the results of the statistical comparisons. The results were similar to those observed in comparing the initial sample sizes. The final sample size did not vary significantly between the two versions of the firm decision aid (Z=.04, p=.48) or the two versions of the AICPA decision aid (Z=-.6, p=.48; KSa=.48 , p=.49). However, the final sample size did vary between the firm and AICPA decision aids (Z=4.47, p<.0001). Given the range of potential sample sizes of the firm and AICPA decision aids (see footnote 17), it was expected that the firm decision aid may result in a smaller mean sample size than the AICPA decision aid. This is consistent with a reduced dependence on substantive sampling by audit firms.

Table 9

Mean Change in Sample Size Due to Working Backwards by Condition

	Firm Decision Aid (Condition 1)	Firm Decision Aid with Rotated Table (Condition 2)	AICPA Decision Aid (Condition 3)	AICPA Decision Aid with Rotated Table (Condition 4)
Mean Initial Sample Size for Auditors Who Worked Backwards	25.00	23.05	47.00	31.67
Mean Final Sample Size for Auditors Who Worked Backwards	19.88	19.42	34.00	31.67
Mean Change in Sample Size Due to Working Backwards	<5.12>	<3.63>	<13.00>	0.00

Note: The mean change in sample size was not significantly different from zero for any of the Conditions.

Table 9 shows the effect that working backwards had on the mean sample size for each condition. Although the net change between the initial and final sample sizes across all conditions was a decrease, it is interesting that 9 auditors reported in the post-experimental questionnaire that they thought the initial sample size was too low. Four of the nine auditors successfully worked backwards to increase the sample size. Three of the four who successfully worked backwards were in Condition 2, while the fourth auditor was in Condition 3. Of the five auditors who were unsuccessful in increasing the initial sample size, three were in Condition 2 and their sample size remained unchanged. Of the remaining two, one was in Condition 1 and one was in Condition 2, the effect of working backwards for these two auditors was a reduction in the final sample size.

Table 10 shows the number of auditors by condition who worked backwards to increase and decrease the initial decision aid sample size. The total of 21 auditors is different from the total number of auditors who were observed to have worked backwards because 12 auditors who actually worked backwards reported in the post-experimental questionnaire that they were satisfied with the initial sample size.

Table 10

Direction of Working Backwards by Condition

	Firm Decision Aid (Condition 1)	Firm Decision Aid with Rotated Table (Condition 2)	AICPA Decision Aid (Condition 3)	AICPA Decision Aid with Rotated Table (Condition 4)	Total
Worked Backwards to Increase Sample Size	0	4	1	0	5
Worked Backwards to Decrease Sample Size	3	5	0	0	8
Worked Backwards, Sample Size Unchanged	0	5	1	2	8
Total	3	14	2	2	21

5.1.2 Sample Size Comparisons to Extant Research

The final AICPA decision aid sample size in this study can be compared to that observed by Messier et al. (2001). Table 11, Panel A provides descriptive statistics for the data in this study and the Messier et al. (2001) study. Panel B of Table 11 provides a statistical comparison of the results from this study and those reported in Table 1 of Messier et al. (2001). The results indicate no significant difference ($t=.3091$, $p=.7588$) in the final sample size of auditors who used the two versions of the AICPA decision aid in this study and the auditors who used the AICPA decision aid in the Messier et al. (2001) study.

Table 11

Comparative Sample Size Statistics for the Current Study and Messier et al. (2001)

Panel A: Descriptive Statistics

	Messier et al. (2001)	AICPA Decision Aid This Study	Firm Decision Aid This Study
Mean Sample Size	35.5	36.75	20.06
Standard Deviation	9.1	17.38	11.69
Minimum Sample Size	22	0	10
Maximum Sample Size	53	70	53
N	22	24	86

Panel B: Planned Comparison

	t-statistic	p-value
Messier et al. Versus Both Versions of AIPCA Decision Aid this Study	.3091	.7588

5.1.3 Implications of Working Backwards on Audit Effectiveness

One of the primary inputs into both the Firm and AICPA decision aid is the control risk assessment. The decision to work backwards has implications for audit effectiveness and efficiency through the reliance on tests of controls. The assessed level of control risk should be dependent on the number of key controls that auditors assumed had been tested. Therefore, the number of controls tested should be a significant determinant of the decision aid sample size. The two variables should be negatively correlated because testing controls reduces the need for substantive sampling if the controls are found to be operating effectively.[39] Therefore, the correlation between the final decision aid sample size and the number of controls that auditors indicated they had assumed were tested was computed. The correlation was negative and significant (Spearman correlation coefficient, -.181, p=.04, normal approximation). However, this is primarily the result of Condition 2, the modified firm decision aid (Spearman correlation coefficient, -.276, p=.028). While the correlation was negative in Condition 1, the firm decision aid, Condition 3, the AICPA decision aid, and Condition 4, the modified AICPA decision aid, the results were not significant (p>.20).

Because working backwards may have had an effect on the relation between the number of controls tested and sample size, similar analyses were conducted on the initial sample sizes. The results indicate that the initial decision aid sample size and the number of controls assumed tested were significantly negatively correlated (Spearman correlation coefficient, -.246, p=.0089, normal approximation). However, this result is primarily the result of Condition 1, the firm decision aid in original form (Spearman correlation coefficient, -.304, p=.06) and Condition 2, the modified firm decision aid (Spearman correlation coefficient, -.35, p=.01, normal approximation). The AICPA decision aid conditions were not significant. For Condition 3, the AICPA decision aid in original form, the correlation was positive and insignificant (Spearman correlation coefficient .063,

[39] The case indicated that the controls were operating as described.

p>.25). In Condition 4, the modified AICPA decision aid, the correlation was negative and insignificant (Spearman -.04, p>.10).

These results taken together with the primary test of Hypothesis One indicate that audit effectiveness may be harmed by the auditors decision to work backwards in the application of the nonstatistical sampling decision aid. Hypothesis One found that across all conditions thirty percent of the auditors worked backwards; therefore, working backwards is not eliminated through decision aid familiarity. The additional analysis found that by working backwards the auditors were reducing or eliminating the significant, negative relation that should exist between the number of controls tested and the decision aid sample size. Therefore, it is apparent that audit effectiveness may be compromised by the auditor's decision to work backwards in the application of the nonstatistical sampling decision aid.

5.2 Hypothesis Two

The second hypothesis tested whether auditors had acquired knowledge of the Firm's nonstatistical sampling decision aid by having auditors use a familiar (Firm) decision aid that had one key input, the sample size adjustment factor table, rotated and then having the auditors identify the potential difference in a multiple choice question contained in the post-experimental questionnaire. Recall from Table 4 that the auditor's measured level of knowledge was dependent on the presence or absence of the classification variables 'working backwards' and 'identification of the table rotation'. Each participant was assigned a knowledge score from 1, 'no knowledge,' to 4, 'high knowledge' based on whether he or she worked backwards and correctly reported the recognition of the sample size adjustment factor table rotation.[40] Hypothesis Two predicted that an auditor's measured level of knowledge was positively correlated with the number of experiences an auditor had had with the decision aid. Table 12 details the number of participants by knowledge score and the number of times they had used the Firm's manual nonstatistical sampling decision aid.

Table 12

Manual Decision Aid Experience by Knowledge Score

	Knowledge Score				
Number of Times participant had used Manual Decision Aid	No Knowledge	Low Knowledge	Medium Knowledge	High Knowledge	Total
0	13	4	3	5	25
1	6	1	0	2	9
2	1	0	1	1	3
3	1	0	0	0	1
4	1	1	0	0	2
5	2	0	1	1	4
10	2	1	0	0	3
11	0	1	0	0	1
12	0	1	0	0	1
20	2	1	1	0	4
25	0	1	0	0	1
30	0	0	0	1	1
55	0	1	0	0	1
Total	28	12	6	10	56

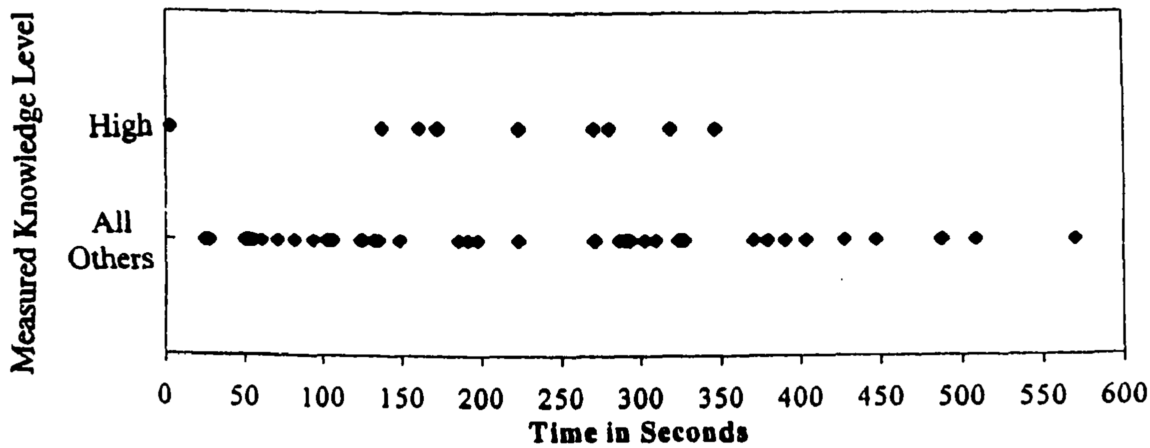
[40] Three participants were eliminated from this analysis because they failed to answer question 11 on the demographic questionnaire, "Was there anything different, other than the computerization, from the sample program (workpaper) used today and the Firm's manual sampling workpaper and/or CBEAM sampling program routinely used in practice?" Therefore, a knowledge score could not be computed for these participants.

Contrary to expectations, experience with the manual decision aid was not significantly related to the knowledge score (Kruskal-Wallis Test, Chi-Square approximation, CHISQ=6.15, p=.45). However, a significant limitation of this analysis was the large number of participants, 45%, who had never used the firm's manual nonstatistical sampling decision aid. Based on prior research it was expected that audit seniors and managers would be familiar with the task; however, for this sample of participants this does not appear to be a valid assumption. Therefore, I examined whether other experiences within the firm may contribute to an auditor's knowledge score. While not significant, the number of months that an auditor had held their current rank was more significant in explaining the participant's knowledge score (Kruskal-Wallis Test, Chi-Square Approximation, CHISQ=15.58, p=.28). In addition, while not significant, the number of months that an auditor had been employed with the firm was more significant in explaining the relation with the participant's knowledge score (Kruskal-Wallis Test, Chi-Square Approximation, CHISQ=20.99, p=.20).

As an additional test of Hypothesis Two the time that participants viewed the sample size adjustment factor table screen was analyzed for differences in the median processing time. The first time the screen was opened it was expected that only auditors with a high level of knowledge would recognize the table rotation. By recognizing the table rotation, high knowledge auditors would have the table screen open for longer than auditors with less knowledge of the table. Figure 3 displays the time that the participants viewed the sample size adjustment factor table the first time through. The expected outcome was not observed

(Median 2-sample test, Z=0, p=.50). The ten auditors classified as high knowledge viewed the screen for a median time (standard deviation) of 198.5 (101.57) seconds while the 46 auditors with medium, low or no knowledge scores viewed the screen for a median time (standard deviation) of 211 (150.33) seconds.[41]

Figure 3
Plot of Sample Size Adjustment Factor Table Viewing Time the First Time Through: High and All Other Knowledge Level Auditors



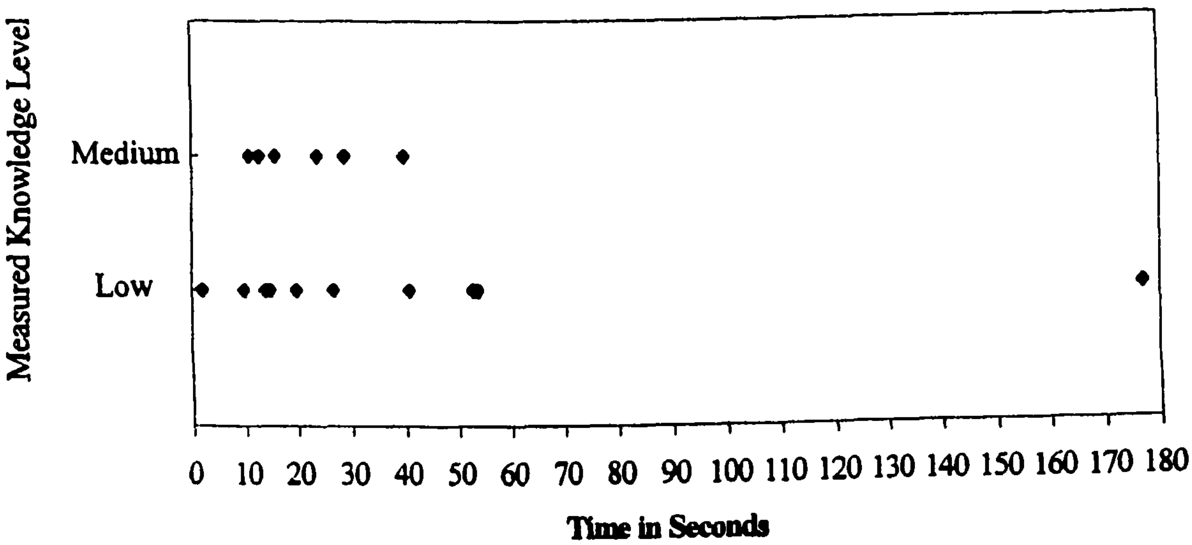
It was expected that auditors classified as having either moderate or low knowledge would work backwards, opening the sample size adjustment factor table screen a second time. The second time the screen was opened, auditors with moderate knowledge would have the screen open longer than low knowledge auditors as they searched to discover what led to a counterintuitive sample size. Figure 4 displays the time that medium and low knowledge participants viewed the sample size adjustment factor table the second time through. Although the

[41] The mean time that auditors classified as high knowledge spent viewing the sample size adjustment factor table screen the first time through was 208.9 seconds. The mean viewing time for all other knowledge levels was 228.61 seconds.

results of this test were not significant (Median 2-Sample test, Z=-.97, p=.166, normal approximation) the means were in the correct direction. The medium knowledge score auditors viewed the table for a median time (standard deviation) of 20.0 (48.78) seconds while the low knowledge score auditors viewed the table for a median time (standard deviation) of 20.0 (26.03) seconds.[42]

Figure 4

Plot of Sample Size Adjustment Factor Table Viewing Time the Second Time Through: Medium and Low Knowledge Level Auditors



In addition, the total time the sample size adjustment factor table screen was opened was analyzed using three planned comparisons for a difference in the

[42] The mean time that auditors classified as medium knowledge spent viewing the sample size adjustment factor table screen the second time through was 40 seconds. The mean viewing time for auditors classified as low knowledge was 26 seconds.

total processing time between the measured knowledge levels. It was expected that the total time would vary depending on the participants' measured knowledge levels. For example, moderate and low knowledge auditors would work backwards leading to a total task time that exceeded that of high and no knowledge auditors. In addition, because the low knowledge level auditors work backwards repeatedly, the total time would exceed the total time of auditors with a moderate knowledge level. Figure 5 displays the total time that participants viewed the sample size adjustment factor table screen by knowledge level. Figure 6, Panel A shows the mean total sample size adjustment factor table viewing time by knowledge score. The comparative statistics are given in Figure 6, Panel B. The results are reported using the Kolmogorov-Smirnov statistic, which is more powerful than the Median-Test (Siegel and Castellan 1988, 151). While the mean times were generally in the expected direction all of the results were not significant.

Figure 5

Plot of Total Time Viewing the Sample Size Adjustment Factor Table by Knowledge Level

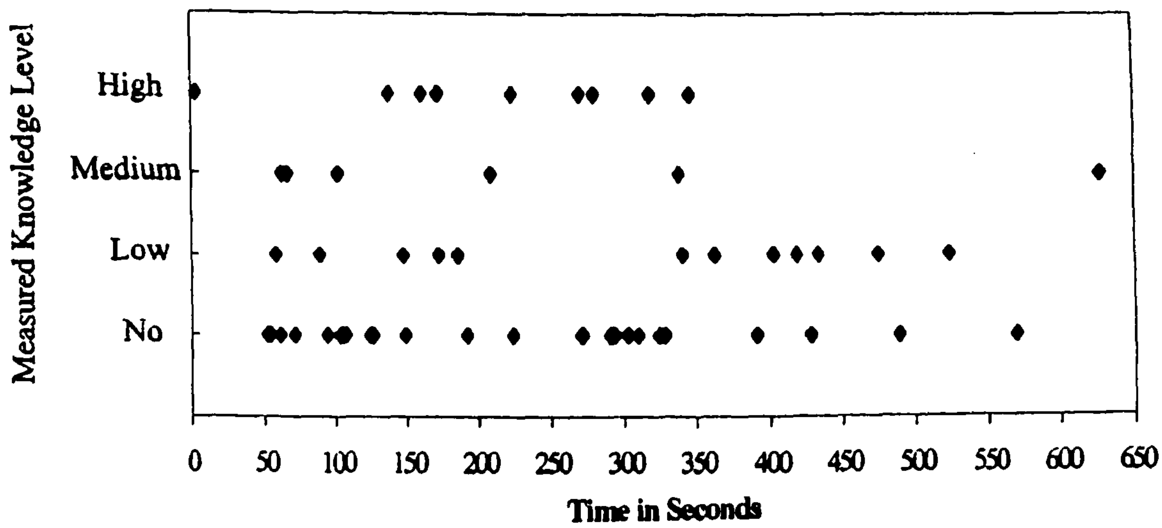
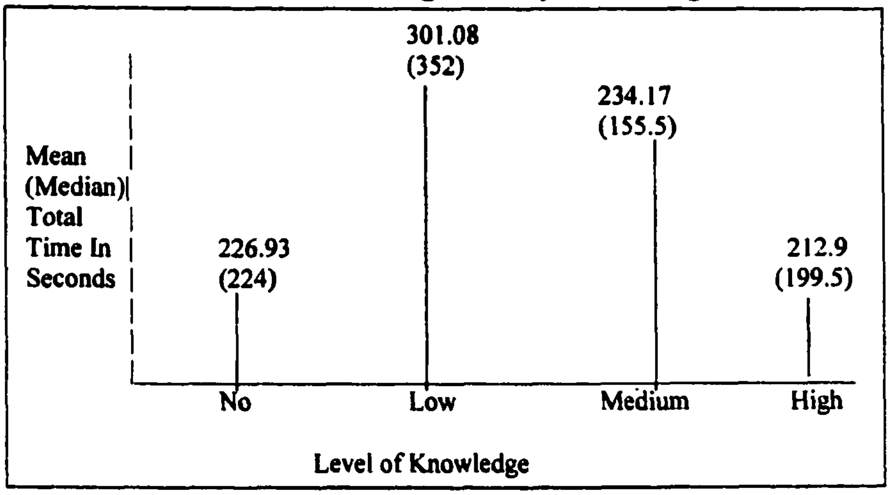


Figure 6
Relationship Between the Mean Total Sample Size Adjustment Factor Table Viewing Time and the Measured Knowledge Levels

Panel A: Mean (Median) Total Table Viewing Time by Knowledge Level



Panel B: Planned Statistics

	KSa-Value	p-value
No versus High Knowledge	.89	.20
Low versus Medium Knowledge	.83	.25
No and High versus Low and Medium Knowledge	1.28	.04

5.3 Hypothesis Three

The third hypothesis examined the extent of an auditor's knowledge of the firm decision aid by comparing the auditor's selected sample size based on tolerable misstatement, the verbal descriptors of the client's control risk, the degree of reliance on other substantive audit procedures, and the expected error rate, to the sample size implied by the familiar decision aid based on the same verbal descriptors. Therefore, Hypothesis Three tested the auditor's knowledge of the complete decision aid. This is different from Hypothesis Two, which examined the effect of knowledge acquired on one component of the familiar decision aid. In Hypothesis Two, the auditor selected only the sample size adjustment factor. On the other hand, Hypothesis Three had auditors select the decision aid sample size using a multiple choice question, without actually using the decision aid. For example, the first question on Part II of the Post-Experimental Questionnaire was:

1. If the control risk is assessed at moderate and there is moderate evidence provided by other substantive procedures, the sample size is
   A. 10
   B. 16
   C. 22
   D. 41
   E. Other ______

It was expected that experience with the decision aid, either in manual or computerized form, would allow subjects to choose the same sample size as would be provided by the decision aid based on the same verbal descriptors of the control risk and reliance on other audit procedures. Therefore, participants with more decision aid experience should match the decision aid sample size more often than participants with less decision aid experience. However, contrary to expectations, manual decision aid (Kruskal-Wallis Test, Chi-Square Approximation, CHISQ=12.49, p=.32) or computer decision aid (Kruskal-Wallis Test, Chi-Square Approximation, CHISQ=19.8, p=.27) experience were not significant in explaining the number of correct sample size matches. Therefore, I examined whether other experiences within the firm were significantly related to the number of correct responses an auditor achieved. The number of months that an auditor had been with the firm and the number of months that an auditor had held their current rank were not significant in explaining the number of correct responses (p=.27, Kruskal-Wallis Test).

Although the post-experimental questions were designed to test for an exact match between the correct sample size and the auditor's chosen sample size, it appears relevant to also test whether decision aid experience allows the auditor to closely approximate the decision aid sample size. Therefore, a difference score was calculated for the net difference between the decision aid sample size and the auditor's sample size summed across the five multiple choice questions. The net difference was then correlated with manual decision aid experience. As would be expected more experience with the decision aid leads to a lower total difference (Spearman correlation coefficient -.19, p=.03).

6.0 Conclusions

A significant body of research has amassed in the development of audit decision aids. Research has documented an unwillingness of knowledgeable individuals to subjugate their professional judgment and decision making responsibility to a decision aid. Working backwards is one way that auditors can demonstrate an unwillingness to rely on a decision aid. If auditors routinely work backwards to circumvent the decision aid this may mitigate the potential benefits of decision aids. However, one potential benefit of decision aids is increased consensus in decision making and the auditors in this study were able to reduce the variance in the decision aid sample size by working backwards.

It is important to understand how auditors use decision aids to improve audit practice. This study contributes to the study of audit decision aids in three ways. First, the results from this research contribute to our understanding of how auditors use decision aids. Previous research has not directly demonstrated how auditors use decision aids that are a required part of the audit workpapers. This research finds that approximately thirty percent of auditors work backwards to alter the decision aid sample size. Therefore, the firm may need to assess the potential effect of this behavior on audit risk. For example, this study finds that when the auditors worked backwards there was no longer a significant negative relation between the number of controls assumed tested and the nonstatistical sample size. If the firm is aware of working backwards in the use of this decision aid, steps can be taken to either change the way the aid is used or additional ways can be sought to minimize the potential increased level of audit risk.[43]

Second, the results provide evidence that familiarity with the decision aid does not affect working backwards. Auditors were equally likely to work backwards when they used their own firm's decision aid and when they used the unfamiliar AICPA decision aid. However, a number of auditors indicated they had little or no experience using the nonstatistical sampling workpaper in either its paper or computerized form. Therefore, it is difficult to conclude that the firm decision aid was familiar to this sample of auditors.

Third, the study provides evidence on what auditors learn through the use of decision aids. Given that thirty percent of the auditors worked backwards they presumably felt that the decision aid sample size was not congruent with their intuitive sample size. Because so few of the auditors had experience with the manual or computer nonstatistical sampling decision aid they must have had other experiences that allowed them to form an intuitive sample size. It is possible that the auditor acquires this knowledge through general audit experience within the firm. The association between time at rank and the measured knowledge score was as strong or stronger than the association of decision aid experience and measured knowledge score, Hypothesis Two.[44] In addition, the correlation between time at rank and time with the firm and the total number of correct answers to the multiple choice questions in part 2 of the post-experimental

---

[43] The firm could attempt to change the way the decision aid is used in a number of ways. For example, the match between the decision aid and the decision aid user could be improved by altering the decision aid or by providing the auditors with additional training.
[44] The value for the Mantel-Haenszel association between time at rank and the measured knowledge level was 2.032, p=.154; while the value for the Mantel-Haenszel association of decision aid experience and the measured knowledge score was .127, p=.722 (manual decision aid experience) and .30, p=.584 (computer decision aid experience).

questionnaire was as strong or stronger than the association of decision aid experience and the total number of correct answers to the multiple choice questions in part 2 of the post-experimental questionnaire, Hypothesis Three.[45] It is apparent from the results of this research that a great deal of what auditors learn comes from general audit experience rather than from specific experiences with the decision aid.

In conclusion, the results of this investigation provide the audit firm with tangible evidence with which they can make informed decisions about the continued use of the firm's nonstatistical sampling decision aid. The first step to such an evaluation would be to assess the costs to the firm of working backwards and the potential benefits of eliminating working backwards. The nonstatistical sample size decision aid currently available within the firm is part of the computerized audit software. Therefore, it would be relatively simple to modify the program to require additional justification if the auditor works backwards.

There are two potential caveats to keep in mind when interpreting the results of this research. First, the auditors that participated in this research were from a single international audit firm. Auditors from other firms may or may not exhibit the same propensity to work backwards. Second, it is not possible to say that working backwards, in and of itself, is good or bad. For example, an auditor may work backwards to correct an error in the initial calculation of sample size.

[45] The value for the Mantel-Haenszel association between time at rank and total correct was .053, p=.819, and the value for the Mantel-Haenszel association between time with the Firm and total correct was 1.262, p=.261; while the value for the Mantel-Haenszel association of decision aid experience and total correct was .003, p=.955 (manual decision aid experience) and 1.476, p=.224 (computer decision aid experience).

Working backwards becomes detrimental when it adversely affects audit risk or results in audit inefficiencies.

7.0 Future Research

Future research should continue to focus on other factors that may contribute to an unwillingness of knowledgeable decision makers to rely on a decision aid and how those influences can be mitigated. There are three key directions for future research. First, method changes could be made. Second, examining the design of the decision aid may prove fruitful. Finally, the focus of the research could change to determine the costs of working backwards.

The method could be changed in several ways. The sample could be expanded to include other firms that have a nonstatistical sampling workpaper. The propensity to work backwards with a familiar decision aid may be firm-specific. Second, the focus could be changed to examine a cross-section of firm-specific decision aids. Working backwards may be a specific bias within the nonstatistical sampling decision aid or it may be an individual bias that affects the use of multiple familiar decision aids. Third, a field study could be conducted. Given that the firm decision aid is computerized within this firm, with the addition of a process tracing program, working backwards could be traced across a larger sample of auditors working on actual audit clients. Also, several auditors indicated that they did not work with manufacturing clients or had little or no experience with the decision aid. Future researchers could improve the match between the task and the participant's experience through improved pre-screening participant selection procedures rather than relying on post-experimental classification. Finally, a limitation of this study was that auditors were only asked how many key controls were tested in the post-experimental questionnaire.

Future research could be enhanced by asking participants which controls they had assumed were tested prior to selecting a sample size adjustment factor.[46] By doing so participants would need to justify deviations from the initial sample size adjustment factor when working backwards. In addition, participants could be asked to list which key controls had been tested and the assumed results of testing those controls.[47]

The specific design of the decision aid could be examined for ways in which it contributes to working backwards. First, both the firm and AICPA nonstatistical sampling decision aids make use of verbal probability phrases to characterize the control risk, inherent risk and the reliance on other substantive procedures. Verbal probability phrases have been shown to exhibit a great deal of variability in interpretation. Therefore, future research may need to consider the role that this previously untested linguistic variable may play in working backwards. Second, the decision aid used in the current study, as well as the firm decision aid used in practice is computerized. By nature of the computerization it may be easier for the auditors to work backwards, because it requires only a click or two of the mouse. Therefore, it may be interesting to discover whether individuals who work backwards with a computerized decision aid are equally likely to work backwards with a decision aid that requires the auditor to manually

---

[46] Auditors may or may not test the same controls for a particular audit assertion. It may be that auditors differ in the weight they apply to each test of control for its ability to reduce control risk. Therefore, it may not be how many controls are tested but which controls are tested that reduces control risk. In addition, auditors may have difficulty translating the results of testing controls into a verbal probability phrase describing the control risk assessment.

[47] The current case simply stated that all controls appeared to be operating as described.

make the required computations. Computerization may also be a factor that inhibits learning because much of the task is unobservable.

Finally, future research should focus on determining the costs of working backwards. The auditors are expending time to work backwards. Therefore, it would be interesting to determine whether the time spent working backwards is offset by a reduced cost for sampling. Second, several auditors in this study worked backwards to increase the sample size. If the increased sample size results in audit inefficiency it is necessary to determine the costs of this inefficiency. Lastly, this study found that the final sample size was not significantly related to the number of controls tested, which the case indicated were operating as described. This finding may indicate that the firm is exposing itself to an increased audit risk.

References

American Institute of Certified Public Accountants. 1972. *Third Standard of Fieldwork.* Generally Accepted Auditing Standards. New York, NY: AICPA.

American Institute of Certified Public Accountants. 1981. *Audit Sampling.* Statement on Auditing Standards No. 39. New York, NY: AICPA.

American Institute of Certified Public Accountants. 1983. *Audit Risk and Materiality in Conducting an Audit.* Statement on Auditing Standards No. 47. New York, NY: AICPA.

American Institute of Certified Public Accountants, Statistical Sampling Subcommittee, Audit and Accounting Guide. 1983. Audit Sampling. AICPA.

American Institute of Certified Public Accountants, Statistical Sampling Subcommittee, Audit and Accounting Guide. 1999. Audit Sampling. AICPA.

Amer, T., K. Hackenbrack, and M. Nelson. 1994. Between-Auditor Differences in the Interpretation of Probability Phrases. *Auditing: A Journal of Practice & Theory* 13(1): 126-136.

Anderson, J.R. 1980. *Cognitive psychology and its implications.* San Francisco: W.H. Freeman.

Arkes, H.R., R.M. Dawes, and C. Christensen. 1986. Factors Influencing the Use of a Decision Rule in a Probabilistic Task. *Organizational Behavior and Human Decision Processes* 37: 93-110.

Arnold, V., P.A. Collier, S.A. Leech, and S.G. Sutton. 1998. The Impact of Intelligent Decision Aids on Experienced and Novice Decision Makers' Judgments. Working Paper, Texas Tech University.

----, and S.G. Sutton. 1998. The Theory of Technology Dominance: Understanding the Impact of Intelligent Decision Aids on Decision Makers' Judgments. *Advances in Accounting Behavioral Research* 1: 175-194.

Ashton, R.H. 1990. Pressure and Performance in Accounting Decision Settings: Paradoxical Effects of Incentives, Feedback, and Justification. *Journal of Accounting Research* 28 (Supplement): 148-180.

----- 1992. Effects of Justification and a Mechanical Aid on Judgment Performance. *Organizational Behavior and Human Decision Processes* 52: 292-306.

-----, and J.J. Willingham. 1988. Using and Evaluating Audit Decision Aids. In R.P. Srivastava and J.E. Rebele (eds.). *Auditing Symposium IX: Proceedings of the 1988 Touche Ross/University of Kansas Symposium on Auditing Problems*. Lawrence, KS: University of Kansas Press.

Beyth-Marom, R. 1982. How Probable is Probable? A Numerical Translation of Verbal Probability Expressions. *Journal of Forecasting* 1: 257-269.

Boatsman, J.R., C. Moeckel, and B.K. Pei. 1995. The Effects of Decision Consequence on Auditors Reliance on Decision Aids in Audit Planning. Working Paper, Arizona State University.

Boettcher III, W. A. 1995. Contexts, Methods, Numbers, and Words. *Journal of Conflict Resolution* 39(3): 561-83.

Bonner, S.E., R. Libby, and M. Nelson. 1996. Using Decision Aids to Improve Auditors' Conditional Probability Judgments. *The Accounting Review* 71(2): 221-240.

Bonner, S.E. and P. Walker. 1994. The Effects of Instruction and Experience on the Acquisition of Auditing Knowledge. *The Accounting Review* 69(1): 157-178.

Brun, W., and K.H. Teigen. 1988. Verbal Probabilities: Ambiguous, Context-Dependent, or Both? *Organizational Behavior and Human Decision Processes* 41(3): 394-404.

Bryant, G.D., and G.R. Norman. 1980. Expressions of Probability: Words and Numbers. *The New England Journal of Medicine* 302: 411.

Budescu, D.V. and T.S. Wallsten. 1985. Consistency in Interpretation of Probabilistic Phrases. *Organizational Behavior and Human Decision Processes* 36(3): 392-405.

-----, S. Weinberg, and T.S. Wallsten. 1988. Decisions Based on Numerically and Verbally Expressed Uncertainties. *Journal of Experimental Psychology* 14(2): 281-294.

Cuccia, A.D., K. Hackenbrack, and M.W. Nelson. 1995. The Ability of Profession Standards to Mitigate Aggressive Reporting. *The Accounting Review* 70 (2): 227-248.

Davis, F.D. and J.E. Kotteman. 1995. Determinants of Decision Rule Use in a Production Planning Task. *Organizational Behavior and Human Decision Processes* 63(2): 145-157.

Dawes, R.M. 1979. The Robust Beauty of Improper Linear Models in Decision Making. *American Psychologist* 34: 571-582.

-----, D. Faust, and P.E. Meehl. 1989. Clinical versus Actuarial Judgment. *Science* 243: 1668-1674.

Einhorn, H.J. 1972. Expert Measurement and Mechanical Combination. *Organizational Behavior and Human Performance* 7 (1): 86-106.

Eining, M.M., 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 and Theory* 16 (4): 1-19.

Ellis, H.C. and R.R. Hunt. 1993. *Fundamentals of Cognitive Psychology.* Madison, WI: Brown and Benchmark Publishers.

Erev, I. and B. Cohen. 1990. Verbal versus Numerical Probabilities: Efficiency, Biases, and the Preference Paradox. *Organizational Behavior and Human Decision Processes* 45: 1-18.

Ericsson, K.A. and J. Smith 1991. Prospects and Limits of the Empirical Study of Expertise: an Introduction. In Ericsson, K.A. and J. Smith (Eds.). *Toward a General Theory of Expertise.* Cambridge, MA: Cambridge University Press.

Gibbins, M. 1984. Propositions about the Psychology of Professional Judgment in Public Accounting. *Journal of Accounting Research.* 22 (1): 103-125.

Grant Thornton. 1996. *Accounting and Auditing Manual.* Grant Thornton.

Glover, S.M., D.F. Prawittt and B.C. Spilker. 1997. The Influence of Decision Aids on User Behavior: Implications for Knowledge Acquisition and Inappropriate Reliance. *Organizational Behavior and Human Decision Processes* 72(2): 232-255.

Hakel, M.D. 1968. How Often is Often? *American Psychologist* 23: 533-534.

Hamm, R.M. 1991. Selection of Verbal Probabilities: A Solution for Some Problems of Verbal Probability Expression. *Organizational Behavior and Human Decision Processes* 48(2): 193-223.

Kachelmeier, S.J. and W.F. Messier, Jr. 1990. An Investigation of the Influence of a Nonstatistical Decision Aid on Auditor Sample Size Decisions. *The Accounting Review*. 65 (1): 209-226.

Kleinmuntz, B. 1990. Why We Still Use Our Heads Instead of Formulas: Toward an Integrative Approach. *Psychological Bulletin* 107(3): 296-310.

Libby, R. 1981. *Accounting and Human Information Processing: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall.

------. 1995. The Role of Knowledge and Memory in Audit Judgment. In Ashton, R. and A. Ashton (Eds.). *Judgment and Decision Making Research in Accounting and Auditing*. Cambridge, MA: Cambridge University Press.

------, and P.A. Libby. 1989. Expert Measurement and Mechanical Combination in Control Reliance Decisions. *The Accounting Review* 64 (4): 729-747.

------, and M.G. Lipe. 1992. Incentives, effort, and the cognitive processes involved in accounting related judgments. *Journal of Accounting Research* 30: 249-273.

------, and J. Luft. 1993. Determinants of Judgment Performance in Accounting Settings: Ability, Knowledge, Motivation, and Environment. *Accounting, Organizations, and Society* 18: 425-450.

Lichtenstein, S. and J.R. Newman. 1967. Empirical Scaling of Common Verbal Phrases Associated with Numerical Probabilities. *Psychonomic Science* 9: 563-564.

Marchant, G. 1990. Discussion of determinants of audit expertise. *Journal of Accounting Research* 28 (Supplement): 21-28.

Matlin, M.W. 1983. *Cognition*. New York: Holt, Rinehart and Winston.

Meehl, P.E. 1957. When shall we use our heads instead of a formula? *Journal of Counseling Psychology* 4: 268-273.

Murphy, A.H., S. Lichtenstein, B. Fischhoff, and R.L. Winkler. 1980. Misinterpretation of Precipitation Probability Forecasts. *Bulletin of the American Meteorological Sciences* 9: 695-701.

Messier, W.F., Jr., S.J. Kachelmeier, and K. Jensen. 2001. An Experimental Assessment of Recent Professional Developments in Nonstatistical Audit Sampling Guidance. *Auditing: A Journal of Practice and Theory* 20 (1): 81-96.

Metzler, J. and R.N. Shepard. 1974.Transformational studies of internal representations of three dimensional objects. In R.L. Solso (Ed.) *Theories of cognitive psychology: The Loyola symposium*. Hillsdale, NJ: Erlbaum, 147-201.

Newell, A., J.C. Shaw, and H.A. Simon. 1958. Elements of a theory of human problem solving. *Psychological Review* 65: 151-166.

Pepper, S. 1981. Problems in Quantification of Frequency Expressions. In D.W. Fiske (Ed.). *New Directions for Methodology of Social and Behavioral Science: Problems with Language Imprecision*. New York: Jossey-Bass.

Peterson, D.K. and G.F. Pitz. 1986. Effect of Input from a Mechanical Model of Clinical Judgment. *Journal of Applied Psychology* 71: 163-167.

Postman, L. 1964. Short-term memory and incidental learning. In A.W. Melton (Ed.). *Categories of human learning*. New York: Academic Press.

Reagan, R.T., F. Mosteller, and C. Youtz. 1989. Quantitative Meanings of Verbal Probability Expressions. *Journal of Applied Psychology* 74(3): 433-442.

Reimers, J. 1992. Additional Evidence on the Need for Disclosure Reform. *Accounting Horizons* March: 36-41.

Rohrmann, B. 1986. Evaluating the Usefulness of Decision Aids: A Methodological Perspective. In Brehmer, B., H. Jungermann, P. Lourens, and G. Sevon (Eds.). *New Directions in Research on Decision Making*. North-Holland: Elsevier Science Publishers.

Rose, J. 1998. The Effects of Cognitive Load on Problem-Type Schema Acquisition and Effort: An Experiment in a Decision Aid Environment. Dissertation Proposal, Texas A&M University (January).

Siegel, S. and N.J. Castellan, Jr. 1988. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw Hill.

Teigen, K.H. and W. Brun. 1999. The Directionality of Verbal Probability Expressions: Effects on Decisions, Predictions, and Probabilistic Reasoning. *Organizational Behavior and Human Decision Processes* 80(2): 155-190.

Tversky, A. and D. Kahneman. 1971. Belief in the law of small numbers. *Psychological Bulletin* (August): 105-110.

Waller, W.S. and W.L. Felix (1984) The auditor and learning from experience: Some Conjectures. *Accounting, Organizations and Society* 9: 383-406.

Whitecotton, S. 1996. The Effects of Experience and Confidence on Decision Aid Reliance: A Causal Model. *Behavioral Research in Accounting* 8: 194-216

Wickelgren, W.A. 1974. How to solve problems: elements of theory of problems and problem solving. San Francisco: Freeman

Appendix A
Case Materials
MODERN PRODUCTS CORPORATION

You have been assigned to the audit of Modern Products Corporation (Modern), a manufacturer of various parts for small consumer appliances. Unadjusted amounts for Modern's balance sheet and income statement for the audit period, calendar year 1999, are as follows:

Modern Products Corporation
Balance Sheet
(in thousands)

Assets		Liabilities & Stockholders' Equity	
Current assets:		Liabilities:	
Cash	$ 3,402	Current liabilities	$10,571
Accounts receivable	10,231	Long-term debt	15,222
Inventory of manufactured goods	10,264	Total liabilities	$25,793
Inventory of supplies	3,758		
Total current assets	$27,655		
		Stockholders' Equity:	
Property, plant & equipment, net of accumulated depreciation	20,072	Capital stock	8,032
		Retained earnings	16,958
Other assets	3,056	Total stockholders' equity	$24,990
Total assets	$50,783	Total liabilities & stockholders' equity	$50,783

Modern Products Corporation
Income Statement
(in thousands)

Sales	$25,494
Cost of sales	15,341
Gross Profit	$10,153
Selling & Administrative expenses	5,032
Net Income before taxes	5,121
Provision for taxes	2,356
Net income	$ 2,765

Net income has been relatively stable in the range of $2 - $4 million for the past several years.

For purposes of this case, your attention is directed to the $3,758,000 balance of supplies inventory. It is considered material to the financial statements taken as a whole. This amount represents the total of a series of similar sized subsidiary accounts representing supplies on hand for any needed repairs of equipment used in manufacturing operations. Corporate management maintains a significant inventory of such supplies in order to minimize the risk of lengthy unproductive "down time" attributable to equipment failure. The supplies inventory consists of approximately 6,000 different parts and components.

For audit purposes, the inventory of supplies is considered material to the financial statements taken as a whole, but is not so critical to warrant formal statistical analysis. Accordingly, the audit manager considers it sufficient to apply **nonstatistical** (judgmental) sampling to supporting subsidiary records. For each sample item selected, the following substantive audit procedures will be performed to primarily test the **existence** and **valuation** assertions:

1. Physically inspect the actual inventory of the sample item comparing both identifying information and the quantity on hand to perpetual inventory records.
2. Ascertain that sample items are not obsolete or otherwise useless.
3. Reconcile recorded balances of sample items to supporting purchase invoices and disbursement vouchers, verifying accurate pricing.
4. Ascertain that inventory has been accurately deducted from inventory ledgers on a first-in, first-out (FIFO) basis, which is the company's inventory policy.

Since many supply inventory items are at remote locations, it is anticipated that physical inspection will be fairly time consuming. Further, pricing tests for some high volume parts and components could involve considerable audit effort. It is therefore desirable to keep the sample size at a minimum, while still obtaining reasonable assurance as to the accuracy of recorded balances. Thus, you want to use an efficient, highly stratified approach that focuses on high volume, high dollar amount parts and components.

This study requests your judgment as to certain characteristics which would be used in determining the extent of the above audit tests of the supplies inventory account. Accordingly, you should consider the following background information.

Internal Control Structure

Modern Products maintains computerized perpetual records for its supplies inventory accounts. The perpetual records can be accessed through computer terminals or personal computers connected to the company's network. The following exhibit shows an example of the information displayed.

Part Number: 473368
Description: Orifice For Stratman Pump #907b
Location: Warehouse E

				Quantity			Amount		
DATE	**P.O.#**	**REQ.#**	**UNIT COST**	**ADD**	**DEL**	**BAL**	**DEBIT**	**CREDIT**	**BALANCE**
4/3/94	Balance Forward		7 @ 43.92 16 @ 47.02			23			1,059.76
4/22/94		4-505	5 @ 43.92		5	18		219.60	840.16
5/9/94		5-219	2 @ 43.92 1 @ 47.02		3	15		134.86	705.30
6/29/94		6-872	5 @ 47.02		5	10		235.10	470.20
7/15/94		7-61	20 @ 49.80	20		30	996.00		1,466.20
9/30/94		9-850	5 @ 47.02		5	25		235.10	1,231.10

Purchases of supplies are entered (added) into the client's perpetual records at cost, with reference to the supporting purchase order number. Purchase orders are maintained in sequential files, must be signed by the person requesting the purchase, and approved by both the appropriate line manager and the corporate treasurer's office. When supplies are used, a requisition order (also sequentially numbered) is completed by the person requesting the supplies, and entered (credited) into the client's perpetual records on a FIFO basis after approval by an authorized line manager. Inventory is never released unless a requisition form has been completed by a line manager.

Physical controls over inventory warehouses include the company that only authorized warehouse personnel are allowed inside the warehouses; supervisors requesting supplies must obtain the requested items at designating receiving gates.

Finally, perpetual inventory records are summarized into account totals at the end of each month, and ten percent of all subsidiary ledger postings are traced back to the supporting purchase orders and requisitions by accounting staff (separate from those who maintain the inventory records) to verify accuracy and proper authorization.

New Controls

As the result of immaterial misstatements discovered during last year's audit (see below), the client has implemented two control procedures recommended in our prior-year letter to client management.

First, client staff now performs periodic physical counts of selected inventory items throughout the year as time permits, verifying quantities shown in the perpetual records. Such counts are performed by administrative personnel independent of the inventory recordkeeping department, and any noted discrepancies are promptly investigated and corrected.

Second, the client now supplements the sample checking of perpetual inventory postings to requisition forms and purchase orders with a "reverse" check from the files of requisition and purchase orders to the perpetual inventory records for ten percent of such documents, randomly selected. This supplemental procedure is performed to help ensure that all requisitions and purchase orders are posted.

Client staff documents both of the above controls by preparing a report of the records checked.

Tests of Controls

Our tests of controls indicate that control procedures appear to be operating as described.

Nature of Other Substantive Tests

Detail accounts maintained for each type of supply inventory will enable fairly comprehensive analytical procedures to test the reasonableness of recorded balances. In addition, year-end cut-off procedures should ensure that all purchases made around the year-end date have been properly recorded.

Expectation of Misstatements

Results from last year's similar audit test of supplies inventory showed a moderate number of physical count errors, usually understatements, which amounted to 1-1.5% of the account balance. It was speculated that some requisition forms were not being posted to the perpetual inventory ledgers. However, client management has stressed to us that the new control procedures implemented this year should have mitigated the problem.

Post-Experimental Questionnaire-Part I

1. Job Code:_______

2. Gender:
   A. Male ________
   B. Female ______

3. How long have you been with Grant Thornton? ______ years ______ months

4. How long have you held your current rank? _______ years ______ months

5. How many courses have you had in sampling?

   __University courses __Firm courses (National and local)__Non-Firm courses

6. Approximately how many times have you used your firm's nonstatistical substantive sample workpaper(i.e., the actual paper document)? **[Please provide a numerical quantity rather than a verbal descriptor such as 'many' or 'few.']** ________ times

7. Approximately how many times have you used your firm's nonstatistical substantive sample computer program included in CBEAM? **[Please provide a numerical quantity rather than a verbal descriptor such as 'many' or 'few.']** ________ times

8. What are the key steps for calculating a nonstatistical substantive sample?

   ____
   ____
   ____
   ____

9. How effective do you believe the Firm's manual sampling workpaper and/or CBEAM sampling program are in calculating an appropriate nonstatistical sample size? [Please put a slash (/) on the scale below.]

   |---------|---------|---------|---------|---------|---------|
   Not Effective — Very Effective

10. How confident are you in your ability to judge the appropriateness of the nonstatistical substantive sample size determined by either the Firm's manual sampling workpaper and/or the CBEAM sampling program? [Please put a slash (/) on the scale below.]

|---------|---------|---------|---------|---------|---------|
Not Confident — Very Confident

11. Was there anything different, other than the computerization, from the sample program (workpaper) used today and the Firm's manual sampling workpaper and/or CBEAM sampling program routinely used in practice? (Circle all that apply.)
    A. yes, the dollar value of tolerable misstatement
    B. yes, the sample size adjustment factor table was rotated
    C. yes, the sample size adjustment factor table values
    D. yes, the entire workpaper was different
    E. no, there was nothing different
    F. other ____________________________

12. Please provide a **numerical** quantity of the number of key controls you assumed had been tested. ________ key controls

13. Were you satisfied with the initial sample size that you calculated with the work paper?

Yes (skip to question 15) — No

14. If you were not satisfied with the initial sample size, was the sample size
    a. too small
    b. too large

15. How often have you changed you risk assessments to recompute the initial nonstatistical sample size determined using either the Firm's manual sampling workpaper or CBEAM sample program? [Please put a slash (/) on the scale below.]

|---------|---------|---------|---------|---------|---------|
Not Often — Very Often

16. How similar was Modern Products Corporation, the case client you evaluated today to those you routinely encounter in practice? [Please put a slash (/) on the scale below.]

|---------|---------|---------|---------|---------|---------|
Not Similar — Very Similar

17. How often do you encounter clients with characteristics similar to Modern Products Corporation, the case client? [Please put a slash (/) on the scale below.]

|---------|---------|---------|---------|---------|---------|
Not Often — Very Often

18. Where you interrupted while completing the task? [Circle one.]
Yes, ______ minutes No

Post-Experimental Questionnaire-Part II

The questions on the following page are intended to assess your knowledge of the Firm's nonstatistical substantive sample workpaper under various combinations of control risk and the amount of evidence provided by related substantive procedures. Because these questions are being asked to assess your knowledge of the Firm's workpaper you have not been provided with the Firm's workpaper. You **should not** use any materials (e.g., the Firm's workpaper, the Firm's CBEAM sampling program, or the sample program used with today's case) other than a calculator to complete these questions.

For each of the questions, assume that you are responsible for determining the nonstatistical sample size for the supplies inventory account. The book value of the supplies inventory account is $3,758,000.

The following additional information is available to aid in your sample size determination:

1. There are no individually significant items.
2. As in the past, a stratified will be computed this year.
3. Based on last year's audit, the expected level of aggregate error is 1.5% of Touchstone.
4. Touchstone is $280,000.
5. The environmental assessment factor is .11.

You should circle the sample size that you believe would be computed using the Firm's manual sampling workpaper of CBEAM sample program under the given combination of control risk and the amount of evidence provided by related substantive procedures. In addition, all Firm policies with respect to the minimum sample size requirements are in effect as you answer the following five questions.

1. If the control risk is assessed at moderate and there is moderate evidence provided by other substantive procedures, the sample size is
   A. 10
   B. 16
   C. 22
   D. 41
   E. Other ______

2. If the control risk is assessed at moderate and there is limited evidence provided by other substantive procedures, the sample size is
   A. 10
   B. 16
   C. 22
   D. 28
   E. Other ______

3. If the control risk is assessed at limited and there is maximum evidence provided by other substantive procedures, the sample size is
   A. 10
   B. 16
   C. 22
   D. 28
   E. Other ______

4. If the control risk is assessed at slightly below maximum and there is limited evidence provided by other substantive procedures, the sample size is
   A. 28
   B. 37
   C. 44
   D. 50
   E. Other ______

5. If the control risk is assessed at slightly below maximum and there is significant evidence provided by other substantive procedures, the sample size is
   A. 10
   B. 22
   C. 28
   D. 37
   E. Other ______

Appendix B
AICPA Decision Aid

Compute the required nonstatistical sample size for the detailed test of the 6,000 different parts and components that comprise the supplies inventory just described. Your computation should be based on the following nonstatistical sample size model suggested by the AICPA.

$$\textit{Sample Size} = \frac{\textit{Population Book Value}}{\textit{Tolerable Misstatement}} \; X \; \textit{Assurance Factor}$$

Since the population's recorded amount is known to be $3,758,000, you only need to assess the tolerable misstatement and the assurance factor. These terms are defined by the AICPA as follows:

<u>Tolerable Misstatement</u>

Tolerable misstatement is the monetary misstatement in the account balance or class of transactions that may exist without causing the specific financial statement assertion to be materially misstated. Please use the AICPA's materiality worksheet to calculate tolerable misstatement.

WORKSHEET FOR CALCULATING PLANNING MATERIALITY AND TOLERABLE MISSTATEMENT

Client:___________________
Balance Sheet Date:________

1. **IDENTIFY AND ESTIMATE BASE**

Base Amount: $___________(greater of total assets or total revenues)

2. **CALCULATING PLANNING MATERIALITY**

Use the following table to calculate planning materiality.

Larger of Total Revenues Or Total Assets Is:						
Over	But Not Over	Planning Materiality Is:	+	Factor	X	Excess Over
$0	$30 thousand	$0	+	.0593	X	$0
30 thousand	100 thousand	1,780	+	.0312	X	30 thousand
100 thousand	300 thousand	3,960	+	.0215	X	100 thousand
300 thousand	1 million	8,260	+	.0145	X	300 thousand
1 million	3 million	18,400	+	.00995	X	1 million
3 million	10 million	38,300	+	.00674	X	3 million
10 million	30 million	85,500	+	.00461	X	10 million
30 million	100 million	178,000	+	.00312	X	30 million
100 million	300 million	396,000	+	.00215	X	100 million
300 million	1 billion	826,000	+	.00145	X	300 million
1 billion	3 billion	1,840,000	+	.000995	X	1 billion
3 billion	10 billion	3,830,000	+	.000674	X	3 billion
10 billion	30 billion	8,550,000	+	.000461	X	10 billion
30 billion	100 billion	17,800,000	+	.000312	X	30 billion
100 billion	300 billion	39,600,000	+	.000215	X	100 billion
300 billion	-----------	82,600,000	+	.000148	X	300 billion

3. **CALCULATE TOLERABLE MISSTATEMENT**

Tolerable misstatement can be calculated using the following "rule of thumb:"

Tolerable Misstatement = Planning Materiality X 2/3

Assurance Factor

An assurance factor is one of twelve values, depending on your assessment of the combination of inherent/control risk, and on your assessment that other substantive procedures (e.g., analytical procedures) designed to test the same assertions will **fail** to detect a material misstatement in a particular assertion. The AICPA's instructions for determining the assurance factor are as follows:

1. Classify your assessment of the combination of inherent and control risk.

   a) **Maximum** – the auditor has assessed control risk at the maximum for the particular assertions.
   b) **Slightly below maximum** – the auditor has support for the belief that internal control structure policies and procedures are **somewhat** effective at preventing or detecting material misstatements in the particular assertions.
   c) **Moderate** – the auditor has support for the belief that internal control structure policies and procedures are **moderately** effective at preventing or detecting material misstatements in the particular assertions.
   d) **Low** – the auditor has support for the belief that internal control structure policies and procedures are **highly** effective at preventing or detecting material misstatements in the particular assertions.

1. Classify your assessment that the risk that other substantive procedures designed to test the same assertions will **fail** to detect a material misstatement in the particular assertions.

   a) **Maximum** – no other substantive procedures are performed that are designed to test the same assertions.

   b) **Slightly below Maximum**

   c) **Moderate** – other substantive procedures designed to test the same assertions are performed that are expected to be **moderately** effective in detecting material misstatements in those assertions.

   d) **Low** – other substantive procedures designed to test the same assertions are performed that are expected to be **highly** effective in detecting material misstatements in those assertions.

1. Based on these judgments, determine the appropriate assurance factor from the following table:

Assurance Factors for Nonstatistical Sampling Formula

ASSESSMENT OFCONTROL RISK (AND INHERENT RISK)	RELIANCE ON OTHER RELEVANT AUDITING PROCEDURES			
	Maximum	**SBM**	**Moderate**	**Low**
Maximum	3.0	2.7	2.3	2.0
Slightly below maximum	2.7	2.4	2.0	1.6
Moderate	2.3	2.1	1.6	1.2
Low	2.0	1.6	1.2	1.0

Appendix C
Probability-Proportional-to-Size (PPS) Sampling

Probability-Proportional-to-Size (PPS) Sampling is a statistical sampling technique that is used to reach conclusions about the dollar amount of the population. The theory underlying PPS sampling is attribute sampling theory; whereby, conclusions about the population are made based on a rate of occurrence. PPS sampling gives every dollar in the population an equal chance of being sampled. However, an auditor would not examine just a single dollar of an account balance or a transaction. For example, the auditor would examine the entire account balance for the supply subsidiary ledger for part number 473368, "Orifice for Stratman Pump #907b" or the auditor would examine the entire transaction to purchase a new building. Therefore, the larger the account balance of the subsidiary ledger or the transaction amount, the more likely it is to be selected for examination by sampling.

PPS sampling offers a number of advantages over the other common statistical sampling technique classical variables sampling. A key advantage of PPS sampling is that it does not require the auditor to directly consider the variation in the dollar amounts of the population. This is an advantage because classical variables sampling (CVS) requires the auditor to either know or estimate the standard deviation of the population; therefore, CVS may require a pilot sample. A second advantage to PPS is that it automatically stratifies the sample, therefore larger dollar items (which if misstated are more likely to cause a material misstatement) have a higher probability of being examined. In addition, all items that exceed the sampling interval will be examined. Third, when no misstatements are expected PPS results in a smaller sample size than CVS. Finally, a sample examined with PPS sampling can begin before the entire population is available. Because the variance in the population does not need to be known and any items added to the population will be examined according to the same sampling interval.

There are also several disadvantages or special considerations that an auditor must make in using PPS sampling. First, PPS assumes that balances are positive. Therefore, if the auditor feels it is important to test zero or negative account balances PPS is not an appropriate method because these balances will not be subject to selection for examination. Second, if the auditor expects or identifies understatements in the population, the evaluation of the sampling results will require special consideration. Third, if misstatements are discovered PPS sampling evaluation may overstate the risk of a Type II error, rejecting the population value when it was actually acceptable. Finally, the PPS sample size may exceed that of the CVS sample size depending on the expected amount of misstatement in the population.

To illustrate the PPS sampling method, the following abbreviated example is taken from the AICPA (1999) *Audit Sampling* Auditing Practice Release page 75. The auditor was designing a test to examine the commercial loans receivable reported balance of $5 million. Based on the assessment of the client the auditor determined tolerable misstatement to be $55,000. Further, the auditor assessed control risk at maximum and determined the appropriate risk of incorrect acceptance was 10 percent. In designing the sample the auditor used an expected misstatement of $10,000. Therefore, the sampling interval was determined as follows:

Tolerable misstatement		$55,000
Expected misstatement	$10,000	
Multiplied by expansion factor for a 10 percent risk of incorrect acceptance (Table D.2, p. 110)	x 1.5	
Less expected effect of misstatements		15,000
Tolerable misstatement adjusted for expected misstatements		$40,000
Divided by reliability factor for no expected misstatements for a 10 percent risk of incorrect acceptance (Table D.1, p. 109)		2.31
Sampling interval		$17,316

The sample size is then calculated by dividing the population value of $5 million by the sampling interval of $17,316; therefore, the sample is 289.[48] The auditor then manually selected the sample using an adding machine in the following manner:

1. He cleared the adding machine.
2. He subtracted a random start between $1 and $17,316, inclusive.
3. He began adding the recorded amounts of logical units in the population, obtaining a subtotal after the addition of each succeeding logical unit. The first logical unit that made the subtotal zero or positive was selected as part of the sample.
4. After each selection, he subtracted the sampling interval of $17,316 as many times as necessary to make the subtotal negative again.
5. He continued adding the logical units as before, selecting all items that caused the subtotal to become zero or positive.

[48] However, because 3 account balances exceed the sampling interval the actual sample size was 281 customer balances.

The primary difference between statistical and nonstatistical sampling is that the former requires the quantification of sampling risk and the latter does not. The Nonstatistical Sampling decision aid contained on pages 47-60 and 105–108 of the AICPA *Audit Sampling* Auditing Practice Release is based on the attributes sampling theory underlying the PPS statistical sampling method (AICPA 1999, 52). Both PPS and nonstatistical sampling do not directly consider variation in the population. While PPS sampling indirectly considers variation by sampling all items that exceed the sampling interval, nonstatistical sampling relies on the auditor's judgment to appropriately stratify the sample. In both methods the auditor must consider the tolerable misstatement[49] and frequency of expected misstatements. The PPS sampling method statistically adjusts for the risk of incorrect rejection, prior to determining the sampling interval and sample size. The nonstatistical sample size decision aid indirectly incorporates sampling risk in the 'Assurance Factor,' which is the auditor's professional judgment with regard to the combined inherent risk and control risk, as well as, "the risk that other substantive procedures designed to test the same assertion will fail to detect a material misstatement" (AICPA 1999, 53).

[49] A decision aid for the calculation of tolerable misstatement is contained on pages 105-108 of the 1999 AICPA *Audit Sampling* Auditing Practice Release.

Appendix D
Verbal Probability Theory

For over 30 years psychologists and others have been studying verbal probability theory (see, Hakel 1968; Lichtenstein and Newmann 1967). Verbal probabilities are intended to convey a range rather than a precise meaning.[50] A great deal of research has amassed demonstrating that individuals use verbal phrases imprecisely to convey quantitative information (see for example, Amer, Hackenbrack, and Nelson 1994; Brun and Teigen 1988; Budescu and Wallsten 1985; Reimers 1992)[51]. It has been found that between-subject variability exceeds within-subject variability (Beyth-Marom 1982; Budescu and Wallsten 1985). While most research has simply asked subjects to apply numerical ranges to multiple verbal phrases other research has tried to pinpoint the underlying cause of the variability in the numerical assessment of verbal phrases. The effect of elicitation method has been studied (Hamm 1991; Teigen and Brun 1999). Also, the effect of context has been tested in basic and applied settings (Boettcher 1995; Brun and Teigen 1988; Bryant and Norman 1980; Reagan, Mosteller, and Youtz 1989; for an overview, see Pepper 1981).

Different contexts may invoke different schema leading to differences in the interpretation of verbal probability phrases. For example, Brun and Teigen (1988) examined the effect of context and found that the subject's familiarity with the context and the desirability of the outcome influenced the interpretation of the

---

[50] This is in contrast to a numerical probability that conveys a precise meaning. For example, if something is said to occur 50% of the time, then over an extended number of observations we can expect the event to occur one-half of the time.

[51] Verbal phrases may be used to express probability (i.e., 'possibly,' 'probably,' or 'perhaps') or to express frequency information (i.e., 'sometimes,' 'often,' or 'frequently'). Empirical research has examined probability and frequency phrases both together and separately with similar results.

probability terms. In addition, context may "lead to higher between subject variability than when the terms are judged in isolation, presumably because the interpretation of probability terms tends to be correlated with the judges' personal opinion on the topic" (390).

A paradoxical finding of several studies is that individuals prefer to receive information numerically (Erev and Cohen 1990; Murphy, Lichtenstein, Fischoff, and Winkler 1980) while they prefer to provide quantitative information with verbal phrases (Erev and Cohen 1990). One reason that has been given for the latter finding is accountability. "(N)umbers are perceived as conveying a level of precision and authority that people do not want to associate with their opinions" (Budescu, Weinberg and Wallsten 1988, 281). As Erev and Cohen (1990) further explain, the vagueness of verbal probability phrases reduces the chance of being punished for being wrong.

Amer et al. (1994) provide an accounting example on the effect of numerical interpretation of verbal probability phrases in an audit task. They had audit managers interpret probability phrases used by a senior concerning the collectability of accounts' receivable. While their results provide evidence of between-person variability in the interpretation of verbal probabilities, they do not pinpoint the exact nature of the variability. The variability may be the result of the interpretation of the verbal probability descriptors or it may reflect an allowance for the difference in meaning another person attaches to a probability phrase. Amer et al. (1994) recognize that their verbal probability interpretation task may result in numerical assessments that are different from one's own numerical meaning of a specific probability phrase.

Physicians, political analysts, meteorologists and accountants are just a few of the many professionals that must deal with uncertainty in communicating information to others. For these and other professionals variability in the interpretation of verbal phrases does not pose a significant problem unless it results in different decisions among individuals. Very little research has focused on the effect of verbal probability phrases on decision making (Budescu, Weinberg and Wallsten 1988).

Decision aids are one means of improving the consistency of decision making across individuals (Bonner et al. 1996; Einhorn 1972). However, audit decision aids provided by the audit firms and the AICPA often make use of verbal probability descriptors to convey quantitative information, this may mitigate the desired increase in consistency. The interpretation of these verbal descriptors is most often left to professional judgment.[52] However, a problem occurs when a difference in the interpretation of the verbal descriptors results in a different decision.

Both the Firm and AICPA nonstatistical sampling decision aid use verbal probability descriptors to characterize the reliance on other audit procedure, control risk and inherent risk. These verbal descriptors, such as 'moderate' and 'slightly below maximum' are intended to convey a range of numerical

---

[52] While it is most often not the case, SFAS #109 paragraph 17 provides a numerical quantification of 'more likely than not,' this represents one way to eliminate the effect of variance in the interpretation of the verbal probability phrases.

probabilities. However, it is not clear that these descriptors are intended to cover the entire range from 0-100%; nor is it clear that they are intended to convey equal ranges of numerical probabilities.

Variability in the interpretation of the verbal probability descriptors may lead auditors, who assess the same audit client, to assign different verbal probability descriptors to the same numerical probability (i.e., 85% may be within the 'moderate' range for one auditor and within the range conveyed by 'slightly below maximum' to another auditor). An individual's professional judgment, with regard to sample size, may be inconsistent with the probability ranges intended by the decision aid creators. Therefore, the imprecision of the verbal probability descriptors may be a contributory factor to working backwards. Consider the following example:

> An auditor determines control risk to be 5% and classifies this as 'maximum.' The auditor also determines that there is a 15% risk that other audit procedures, testing the same assertion, will fail to detect a material misstatement and classifies this as 'low.' Using the AICPA decision aid the indicated sample size is 34. However, if the auditor's intuitive sample size is 40, the auditor may work backwards to increase the indicated sample size. By changing the risk of other substantive procedures failing to detect a material misstatement to 'moderate,' the indicated sample size is increased to 39 which is closer to the auditor's intuitive sample size. This scenario is in contrast to an auditor who assessed the same client similarly, but initially applied the verbal probability label 'moderate' to the risk that other substantive procedures would fail to detect a material misstatement.

Appendix E
Results of Parametric Data Analysis

Hypothesis One

The first hypothesis examined the effect of an auditor's decision aid familiarity on working backwards in the application of the nonstatistical sampling decision aid.[53] The 2 X 2 research design was planned so that Condition 3, the revised AICPA decision aid, and Condition 4, the revised AICPA decision aid with the rotated sample size adjustment factor table, could be collapsed into a single condition for the data analysis. Table 13, Panel A shows the number of auditors who did and did not work backwards in each version of the decision aid. In order to combine Condition 3 and Condition 4, as planned, it was necessary to determine whether there was a difference in working backwards with the AICPA decision aid in its original form and the AICPA decision aid with the rotated sample size adjustment factor table. Panel B of Table 13 presents the results of the planned comparisons. The decision to work backwards was not significantly related to the table rotation for the AICPA decision aid (F=.05, p=.8281). Based upon visual observation of the data it appeared that there was no difference in the propensity of auditors to work backwards when using either version of the firm decision aid; therefore, this was tested prior to testing Hypothesis One. Similar to the results obtained for the two versions of the AICPA decision aid, the sample size adjustment factor table rotation did not significantly affect the auditor's decision to work backwards with the firm decision aid (F=.15, p=.7011). Therefore, both of the versions of the firm decision aid and the AICPA decision

[53] While it was expected that all participants who received the firm decision aid would be familiar with the decision aid subsequent analysis of hypothesis two indicated that a large number of the participants had never used the firm decision aid. Therefore, participants who indicated that they had no experience with the decision aid were eliminated from the analysis of hypothesis one. This resulted in the elimination of 40 participants, thirteen from Condition 1 and twenty-seven from Condition 2.

aid were pooled for the test of Hypothesis One.

Table 13

Observed Working Backwards by Condition

Panel A: Frequency of working backwards by Condition

	Firm Decision Aid (Condition 1)	Firm Decision Aid with Rotated Table (Condition 2)	AICPA Decision Aid (Condition 3)	AICPA Decision Aid with Rotated Table (Condition 4)	Total
Did Not Work Backwards	10	21	8	10	49
Did Work Backwards	4	11	3	3	21
Total	14	32	11	13	70

Panel B: Planned Comparisons

	F-value	p-value
Firm Decision Aid Rotated Table versus Firm Decision Aid	.15	.7011
AICPA Decision Aid Rotated Table versus AICPA Decision Aid	.05	.8281

Auditors, experienced in the determination of a nonstatistical sample size, who used an unfamiliar decision aid were expected to work backwards more often than auditors who used a familiar decision aid. Based upon the aforementioned finding that the sample size adjustment factor table rotation did not affect working backwards, Conditions 1 and 2 were compared with Conditions 3 and 4 for the test of Hypothesis One. Table 14, Panel A shows the frequency of working backwards in the unfamiliar (AICPA) and familiar (firm) decision aid conditions. Panel B of Table 14 shows the statistical results of the planned comparison. The results of Hypothesis One indicate that contrary to expectations, auditors who used the unfamiliar AICPA decision aid did not work backwards more often than the auditors who used the firm's decision aid (Two-tail Fisher's Exact Test .974). On average, across all versions of the decision aid, 30% of the participants worked backwards. Therefore, Hypothesis One was not supported. Decision aid familiarity did not significantly impact the decision to work backwards.

Table 14

Observed Working Backwards between Familiar and Unfamiliar Decision Aid Conditions

Panel A: Frequency of Observed Working Backwards

	Familiar Decision Aid	Unfamiliar Decision Aid	Total
Did Not Work Backwards	31 (67%)	18 (75%)	49 (70%)
Did Work Backwards	15 (33%)	6 (25%)	21 (30%)
Total	46	24	70

Panel B: Planned Statistical Comparison

	F-value	p-value
Familiar Decision Aid versus Unfamiliar Decision Aid	.42	.52

Implications of working backwards on sample size

Given the propensity of auditors to work backwards it is important to assess the implications of the decision to work backwards. The decision to work backwards has implications for audit effectiveness and efficiency through the substantive sample size. Therefore, it is relevant to discuss the impact of working backwards on the decision aid output, sample size.

The mean initial sample sizes, prior to any working backwards, are shown in Table 15, Panel A. Visual observation of the data prompted a statistical comparison of the differences in the sample sizes across the decision aid conditions. Table 15, Panel B presents the results of the statistical comparisons. The initial sample size did not vary significantly between the two versions of the firm decision aid (F=.01, p=.9034) or the two versions of the AICPA decision aid (F=.04, p=.8346). However, the initial sample size did vary significantly between the firm and AICPA decision aids (F=29.21, p=.0001).

Table 15
Initial Sample Size by Condition

Panel A: Descriptive Statistics for the Initial Sample Size by Condition

	Firm Decision Aid (Condition 1)	Firm Decision Aid with Rotated Table (Condition 2)	AICPA Decision Aid (Condition 3)	AICPA Decision Aid with Rotated Table (Condition 4)
Mean Sample Size	21.11	21.49	39	37.85
Standard Deviation	12.67	12.88	18.76	12.5
Minimum	10	10	0	23
Maximum	52	52	70	56
N	27	59	11	13

Panel B: Planned Comparisons

	F-value	p-value
Firm Decision Aid Rotated Table versus Firm Decision Aid	.01	.9034
AICPA Decision Aid Rotated Table versus AICPA Decision Aid	.04	.8346
Both versions of the Firm Decision Aid versus both versions of the AICPA Decision Aid	29.21	.0001

In addition to the initial samples, the final sample sizes, after any working backwards, were compared. Table 16, Panel A presents the mean final sample sizes by version, while Panel B presents results of the statistical comparisons. The results were similar to those observed in comparing the initial sample sizes. The final sample size did not vary significantly between the two versions of the firm decision aid (F=.05, p=.8261) or the two versions of the AICPA decision aid (F=.20, p=.6594). However, the final sample size did vary between the firm and AICPA decision aids (F=28.94, p=.0001). Given the range of potential sample sizes of the firm and AICPA decision aids (see footnote 17), it was expected that the firm decision aid may result in a smaller mean sample size than the AICPA decision aid. This is consistent with a reduced dependence on substantive sampling by audit firms.

Table 16

Final Sample Size by Condition

Panel A: Descriptive Statistics

	Firm Decision Aid (Condition 1)	Firm Decision Aid with Rotated Table (Condition 2)	AICPA Decision Aid (Condition 3)	AICPA Decision Aid with Rotated Table (Condition 4)
Mean Sample Size	19.59	20.27	35.45	37.85
Standard Deviation	10.66	12.21	22.45	12.50
Minimum	10	10	0	23
Maximum	44	52	70	56
N	27	59	11	13

Panel B: Planned Comparisons

	F-value	p-value
Firm Decision Aid Rotated Table versus Firm Decision Aid	.05	.8261
AICPA Decision Aid Rotated Table versus AICPA Decision Aid	.20	.6594
Both versions of the Firm Decision Aid versus both versions of the AICPA Decision Aid	28.91	.0001

Table 17 shows the effect that working backwards had on the mean sample size for each condition. Although the net change between the initial and final sample sizes across all conditions was a decrease, it is interesting that 9 auditors reported in the post-experimental questionnaire that they thought the initial sample size was too low. Four of those auditors successfully worked backwards to increase the sample size. Three of the four who successfully worked backwards were in Condition 2, while the fourth auditor was in Condition

3. Of the five auditors who were unsuccessful in increasing the initial sample size, three were in Condition 2 and their sample size remained unchanged. Of the remaining two, one was in Condition 1 and one was in Condition 2, the effect of working backwards for these two auditors was a reduction in the final sample size.

Table 17

Mean Change in Sample Size Due to Working Backwards by Condition

	Firm Decision Aid (Condition 1)	Firm Decision Aid with Rotated Table (Condition 2)	AICPA Decision Aid (Condition 3)	AICPA Decision Aid with Rotated Table (Condition 4)
Mean Initial Sample Size for Auditors Who Worked Backwards	25.00	23.05	47.00	31.67
Mean Final Sample Size for Auditors Who Worked Backwards	19.88	19.42	34.00	31.67
Mean Change in Sample Size Due to Working Backwards backwards	<5.12>	<3.63>	<13.00>	0.00

Note: The mean change in sample size was not significant for any of the Conditions.

Table 18 shows the number of auditors by condition who worked backwards to increase and decrease the initial decision aid sample size. The total of 21 auditors is different from the total number of auditors who were observed to have worked backwards because 12 auditors who actually worked backwards reported in the post-experimental questionnaire that they were satisfied with the initial sample size.

Table 18

Direction of Working Backwards by Condition

	Firm Decision Aid (Condition 1)	Firm Decision Aid with Rotated Table (Condition 2)	AICPA Decision Aid (Condition 3)	AICPA Decision Aid with Rotated Table (Condition 4)	Total
Worked Backwards to Increase Sample Size	0	4	1	0	5
Worked Backwards to Decrease Sample Size	3	5	0	0	8
Worked Backwards, Sample Size Unchanged	0	5	1	2	8
Total	3	14	2	2	21

<u>Sample Size Comparisons to Extant Research</u>

The final AICPA decision aid sample size in this study can be compared to that observed by Messier et al. (2001). Table 19, Panel A provides descriptive statistics for the data in this study and the Messier et al. (2001) study. Panel B of Table 19 provides a statistical comparison of the results from this study and those reported in Table 1 of Messier et al. (2001). The results indicate no significant difference (t=.3091, p=.7588) in the final sample size of auditors who used the two versions of the AICPA decision aid in this study and the auditors who used the AICPA decision aid in the Messier et al. (2001) study.

Table 19

Comparative Sample Size Statistics for the Current Study and Messier et al. (2001)

Panel A: Descriptive Statistics

	Messier et al. (2001)	AICPA Decision Aid This Study	Firm Decision Aid This Study
Mean Sample Size	35.5	36.75	20.06
Standard Deviation	9.1	17.38	11.69
Minimum Sample Size	22	0	10
Maximum Sample Size	53	70	53
N	22	24	86

Panel B: Planned Comparisons

	t-statistic	p-value
Messier et al. Versus Both Versions of AIPCA Decision Aid this study	.3091	.7588

Implications of Working Backwards on Audit Effectiveness

One of the primary inputs into both the firm and AICPA decision aid is the control risk assessment. The decision to work backwards has implications for audit effectiveness and efficiency through the reliance on tests of controls. The assessed level of control risk should be dependent on the number of key controls that auditors assumed had been tested. Therefore, the number of controls tested should be a significant determinant of the decision aid sample size. The two variables should be negatively correlated because testing controls reduces the need for substantive sampling if the controls are found to be operating effectively.[54] Therefore, the correlation between the final decision aid sample size and the number of controls that auditors indicated they had assumed tested was computed. While there was a negative correlation (Pearson correlation coefficient, -.12551) the relationship was not significant (p=.2281). The correlation of each condition was tested as well. While each of the conditions exhibited a negative correlation between the final sample size and number of controls tested, none of the results were significant (p>.16).

Because working backwards may have had an effect on the relation between the number of controls tested and sample size, similar analyses were conducted on the initial sample sizes. The results indicate that the initial decision aid sample size and the number of controls assumed tested were significantly negatively correlated (Pearson correlation coefficient, -.2262, p=.0284). However, this result is primarily the result of Condition 1, the firm decision aid in original form (Pearson correlation coefficient, -.33854, p=.1056) and Condition 2, the modified firm decision aid (Pearson correlation coefficient, -.23911, p=.1017). In Conditions 3 and 4, the AICPA decision aid, the initial sample size and the number of controls tested were negatively correlated the results were not significant (p>.6055).

[54] The case indicated that the controls were operating as described.

These results taken together with the primary test of Hypothesis One indicate that audit effectiveness may be harmed by the auditors decision to work backwards in the application of the nonstatistical sampling decision aid.

Hypothesis Two

The second hypothesis tested whether auditors had acquired knowledge of the firm's nonstatistical sampling decision aid by having auditors use a familiar (firm) decision aid that had one key input, the sample size adjustment factor table, rotated and then having the auditors identify the potential difference in a multiple choice question contained in the post-experimental questionnaire. Recall from Table 4 that the auditor's measured knowledge level was dependent on the presence or absence of the classification variables 'working backwards' and 'identification of the table rotation.' Each participant was assigned a knowledge score from 1, 'no knowledge,' to 4, 'high knowledge' based on whether he or she worked backwards and correctly reported the recognition of the sample size adjustment factor table rotation.[55] Hypothesis Two predicted that an auditor's measured level of knowledge was positively correlated with the number of experiences an auditor had had with the decision aid. Table 20 details the number of participants by knowledge score and the number of times they had used the firm's manual nonstatistical sampling decision aid.

---

[55] Three participants were eliminated from this analysis because they failed to answer question 11 on the demographic questionnaire, "Was there anything different, other than the computerization, from the sample program (workpaper) used today and the Firm's manual sampling workpaper and/or CBEAM sampling program routinely used in practice?" Therefore, a knowledge score could not be computed for these participants.

Contrary to expectations, experience with the manual decision aid was not significantly related to the knowledge score ($p=.662$, Two-tail Fisher's Exact Test). A Fisher's Exact Test rather than a Chi-Square was used because 96% of the cells have expected counts of less than 5. However, a significant limitation of this analysis was the large number of participants, 45%, who have never used the firm's manual decision aid. Based on prior research it was expected that audit seniors and managers would be familiar with the task; however, for this sample of participants this does not appear to be a valid assumption. Therefore, I examined whether other experiences within the firm may contribute to an auditor's knowledge score. While not significant, the number of months that an auditor had held their current rank was more significantly correlated with the participant's knowledge score ($p=.124$, Two-tail Fisher's Exact Test) than manual decision aid experience.

Table 20

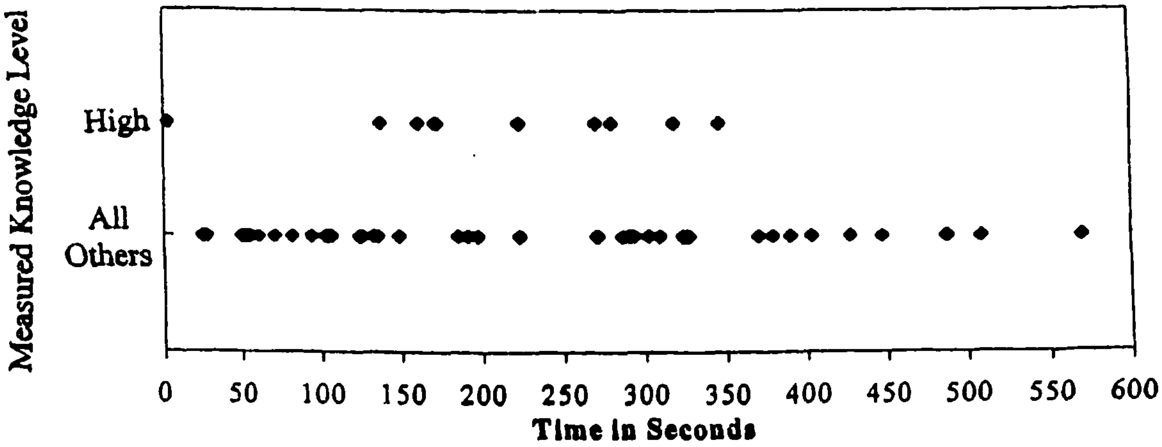
Manual Decision Aid Experience by Knowledge Score

	Knowledge Score				
Number of Times participant had used Manual Decision Aid	No Knowledge	Low Knowledge	Medium Knowledge	High Knowledge	Total
0	13	4	3	5	25
1	6	1	0	2	9
2	1	0	1	1	3
3	1	0	0	0	1
4	1	1	0	0	2
5	2	0	1	1	4
10	2	1	0	0	3
11	0	1	0	0	1
12	0	1	0	0	1
20	2	1	1	0	4
25	0	1	0	0	1
30	0	0	0	1	1
55	0	1	0	0	1
Total	28	12	6	10	56

As an additional test of Hypothesis Two the time that participants viewed the sample size adjustment factor table screen was analyzed for differences in the mean processing time. The first time the screen was opened it was expected that only auditors with a high level of knowledge would recognize the table rotation. By recognizing the table rotation, high knowledge auditors would have the table screen open for longer than auditors with less knowledge of the table. The expected outcome was not observed ($F=.394$, $p=.6951$). Figure 7 displays the time that the participants viewed the sample size adjustment factor table the first time through. The ten auditors classified as high knowledge viewed the screen for a mean (standard deviation) of 208.9 (101.57) seconds while the 46 auditors with medium, low or no knowledge scores viewed the screen for a mean (standard deviation) of 228.61 (150.33) seconds.

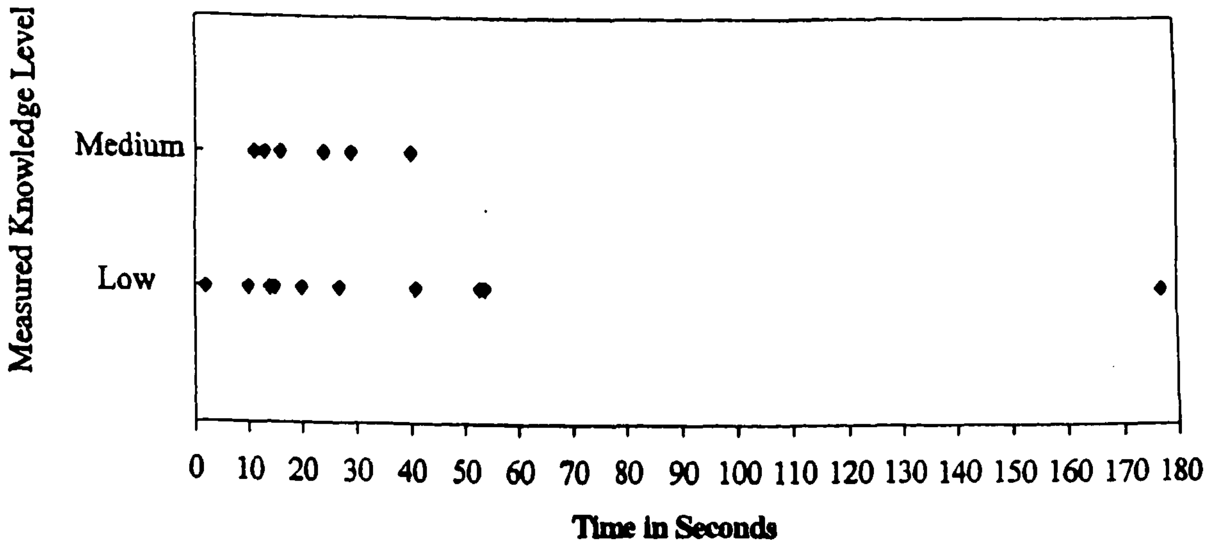
Figure 7

Plot of Sample Size Adjustment Factor Table Viewing Time the First Time Through: High and All Other Knowledge Level Auditors



It was expected that auditors classified as having either moderate or low knowledge would work backwards, opening the sample size adjustment factor table screen a second time. The second time the screen was opened, auditors with moderate knowledge would have the screen open longer than low knowledge auditors as they searched to discover what led to a counterintuitive sample size. Figure 8 displays the time that medium and low knowledge participants viewed the sample size adjustment factor table the second time through. Although the results of this test were not significant (t=.6939, p=.4977) the means were in the correct direction. The medium knowledge score auditors viewed the table for a mean (standard deviation) of 40.0 (48.78) seconds while the low knowledge score auditors viewed the table for a mean (standard deviation) of 26.00 (26.03) seconds.

Figure 8
Plot of Sample Size Adjustment Factor Table Viewing Time the Second Time Through: Medium and Low Knowledge Level Auditors



In addition, the total time the sample size adjustment factor table screen was opened was analyzed using an ANOVA with three planned comparisons for a difference in the mean processing time between the measured knowledge levels. It was expected that the total time would vary depending on the participants' measured knowledge levels. For example, moderate and low knowledge auditors would work backwards leading to a total task time that exceeded that of high and no knowledge auditors. In addition, because the low knowledge level auditors work backwards repeatedly, the total time would exceed the total time of auditors with a moderate knowledge level. Figure 9 displays the total time that participants viewed the sample size adjustment factor table screen by knowledge level. Figure 10, Panel A shows the mean total sample size adjustment factor table viewing time by knowledge score. The comparative statistics are given in Figure 10, Panel B. While the results were generally in the expected direction they were not significant.

Figure 9

Plot of Total Time Viewing the Sample Size Adjustment Factor Table by Knowledge Level

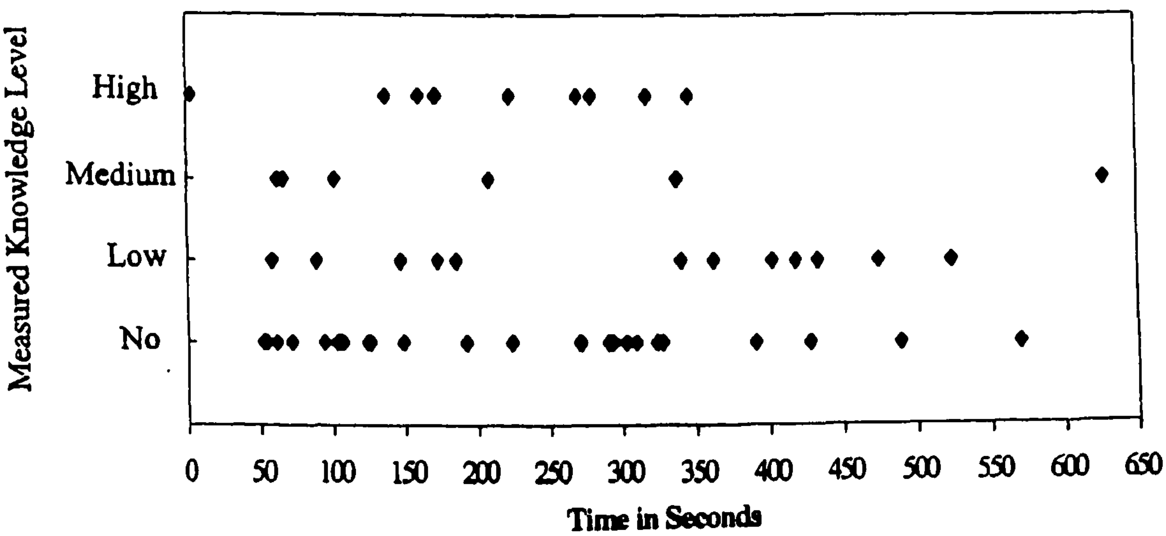
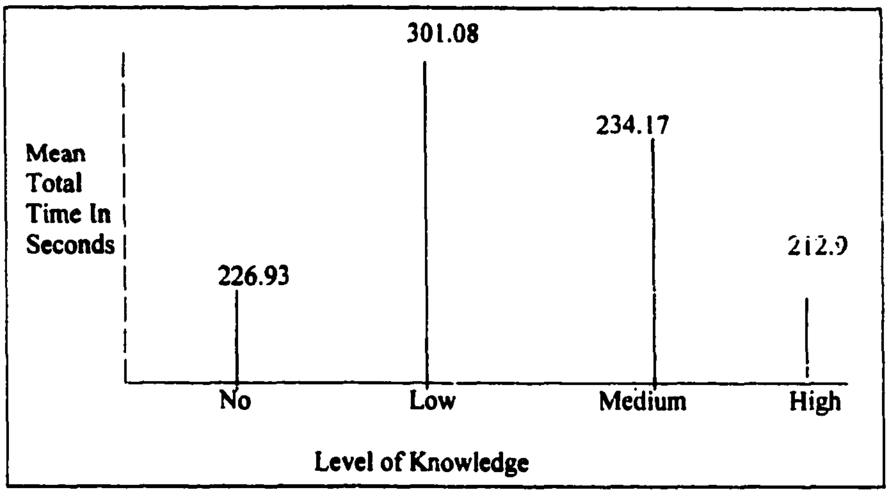


Figure 10

Relationship Between the Mean Total Sample Size Adjustment Factor Table Viewing Time and the Measured Knowledge Levels

Panel A: Mean (Median) Total Table Viewing Time by Knowledge Level



Panel B: Comparative Statistics

ANOVA

	df	F Value	p-value
Knowledge Score	3,52	.87	.4620

Planned Comparisons

	F Value	p-value
No versus High Knowledge	.07	.7976
Low versus Moderate Knowledge	.82	.3691
Low and High versus Low and Moderate Knowledge	1.08	.3031

Hypothesis Three

The third hypothesis examined the extent of an auditor's knowledge of the firm decision aid by comparing the auditor's selected sample size based on tolerable misstatement, the verbal descriptors of the client's control risk, the degree of reliance on other substantive audit procedures, and the expected error rate, to the sample size implied by the familiar decision aid based on the same verbal descriptors. Therefore, Hypothesis Three tested the auditor's knowledge of the complete decision aid. This is different from Hypothesis Two, which examined the effect of knowledge of one component of the familiar decision aid. In Hypothesis Two, the auditor was selecting the sample adjustment factor. On the other hand, Hypothesis Three had auditors select the decision aid sample size without actually using the decision aid. For example, the first question on Part II of the Post-Experimental Questionnaire was:

1. If the control risk is assessed at moderate and there is moderate evidence provide by other substantive procedures, the sample size is

A. 10
B. 16
C. 22
D. 41
E. Other ________

It was expected that experience with the decision aid, either in manual or computerized form, would allow subjects to choose the same sample size as would be provided by the decision aid based on the same verbal descriptors of the control risk and reliance on other audit procedures. Therefore, participants with more decision aid experience should match the decision aid sample size more often than participants with less decision aid experience. It was not possible to use a Chi-Square because all cells had expected frequencies of less than 5. In addition, a Fisher's Exact Test could not be run because of the size of the contingency table. Therefore, decision aid experience was regressed on the total number of correct matches between the auditor's sample size and the decision aid sample size. However, contrary to expectation, manual decision aid or computer decision aid experience were not significant in explaining the number of correct sample size matches (F=.68, p=.8751).

Although the post-experimental questions were designed to test for an exact match between the correct sample size and the auditor's chosen sample size, it appears relevant to also test whether decision aid experience allows the auditor to closely approximate the decision aid sample size. Therefore, a difference score was calculated for the net difference between the decision aid sample size and the auditor's sample size summed across the five multiple choice questions. The net difference was then correlated with manual decision aid experience. As would be expected more experience with the decision aid leads to a lower total difference (Pearson correlation coefficient, -.23, p=.03). In addition, manual decision aid experience was significant is explaining the total net difference between the decision aid sample size and the auditor's sample size (F=1.85, p=.04).