AUDITORS' PLANNING STAGE MATERIALITY JUDGMENTS
AND THE MEDIATING EFFECTS OF LEVEL OF RESPONSIBILITY,
FIRM AFFILIATION, AND AUDIT TECHNOLOGY: AN EXPERIMENT

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the
Degree Doctor of Philosophy in the Graduate School of
The Ohio State University

By

Donald Eugene Tidrick, B.S., M.B.A.

* * * * *

The Ohio State University
1987

Reading Committee: Approved by:
Thomas J. Burns
Bruce R. Gaumnitz
Richard J. Murdock
Jane F. Mutchler
Dennis K. Pearl

Thomas J. Burns
Adviser
Department of Accounting
DEDICATION

to the memory of my beloved mother,
Hazel G. Tidrick, who died May 30, 1986,
and

to the memory of my friend and classmate,
Daniel R. Lenox, Ph.D. Candidate, who died
July 18, 1986
ACKNOWLEDGMENTS

I am indebted to the members of my reading committee for their thoughtful suggestions and helpful contributions: Bruce R. Gaumnitz, Richard J. Murdock, Jane F. Mutchler, and Dennis K. Pearl. David Fairley contributed greatly to the statistical analysis; Richard J. Klimoski was helpful in the design of the questionnaires. My program adviser and dissertation chairman, Professor Thomas J. Burns, deserves special acknowledgment. Without his example and positive influence, I would not have chosen an academic career. I appreciate having had the opportunity to pursue my doctorate at The Ohio State University.

I am grateful for the kind assistance that I received from the following Big Eight CPA firms: Arthur Andersen; Coopers & Lybrand; Deloitte, Haskins & Sells; Peat Marwick; Price Waterhouse; and Touche Ross. Without their substantial contributions, this dissertation would have remained "on the drawing board."

My family has been a constant source of encouragement. I have been profoundly fortunate to have had wonderful parents. Throughout my life, they will be an integral part of any success which may come to me.

Finally, I would like to express my appreciation, in general, to my good friends at Ohio State and elsewhere. Without these friendships, the accomplishment of this goal would be much less enjoyable.
VITA

May 7, 1953 ................ Born -- Coshocton, Ohio

1976. ...................... B.S., The Ohio State University, Columbus, Ohio

1977. ...................... MBA, University of Wisconsin, Madison, Wisconsin

1978-1980 .................. Staff and Senior Auditor, Peat, Marwick, Mitchell & Co., Columbus, Ohio

1981. ...................... Instructor, Department of Accounting, The Ohio State University, Columbus, Ohio

1982-1986 .................. Teaching Associate, Department of Accounting, The Ohio State University, Columbus, Ohio

1987. ...................... Assistant Professor, Department of Accountancy, University of Notre Dame, South Bend, Indiana

FIELDS OF STUDY

Major Field: Accounting

Minor Fields: Decision Theory (Engineering) Statistics
TABLE OF CONTENTS

DEDICATION .................................................. ii
ACKNOWLEDGMENTS ........................................ iii
VITA .......................................................... iv
LIST OF TABLES ............................................ vii
LIST OF FIGURES ........................................... ix

CHAPTER

<table>
<thead>
<tr>
<th>I.  INTRODUCTION.</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Concept of Materiality</td>
<td>1</td>
</tr>
<tr>
<td>Motivation for the Study</td>
<td>2</td>
</tr>
<tr>
<td>Objectives of the Dissertation</td>
<td>5</td>
</tr>
<tr>
<td>Research Methodology</td>
<td>6</td>
</tr>
<tr>
<td>Contributions</td>
<td>7</td>
</tr>
<tr>
<td>Organization of the Dissertation</td>
<td>11</td>
</tr>
<tr>
<td>Chapter I Footnotes</td>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. LITERATURE REVIEW.</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>14</td>
</tr>
<tr>
<td>Standard-setting Bodies</td>
<td>14</td>
</tr>
<tr>
<td>Empirical Research Regarding Materiality</td>
<td>21</td>
</tr>
<tr>
<td>Other Relevant Research</td>
<td>65</td>
</tr>
<tr>
<td>Summary</td>
<td>82</td>
</tr>
<tr>
<td>Chapter II Footnotes</td>
<td>86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. HYPOTHESES AND METHODOLOGY.</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>91</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>95</td>
</tr>
<tr>
<td>Empirical Research Strategy</td>
<td>103</td>
</tr>
<tr>
<td>Analyses To Be Performed</td>
<td>118</td>
</tr>
<tr>
<td>Underlying Statistical Assumptions</td>
<td>145</td>
</tr>
<tr>
<td>Administration of the Experiment</td>
<td>149</td>
</tr>
<tr>
<td>Chapter III Footnotes</td>
<td>153</td>
</tr>
</tbody>
</table>
IV. RESULTS OF ANALYSES

    Introduction .................................. 161
    Experience Effects ............................. 165
    Firm Effects .................................. 176
    Audit Technology Effects ...................... 190
    Risk Attitude ................................ 200
    Additional Analyses ........................... 206
    Comments on Statistical Assumptions .......... 217
    Chapter IV Footnotes .......................... 220

V. REVIEW AND SUMMARY OF RESULTS, CONCLUSIONS,
   LIMITATIONS, AND EXTENSIONS .................. 228

    Review and Summary of Results ............... 228
    Conclusions .................................. 238
    Limitations of the Experiment ............... 245
    Extensions .................................. 250

APPENDICES ....................................... 253

    A. Instructions and Background Information .. 253
    B. Attitude Measurement Questionnaire ....... 263
    C. Survey Questionnaire ....................... 275
    D. Results of Analyses
       (Balance Sheet Judgments) ................. 283

BIBLIOGRAPHY .................................... 314
<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>127</td>
</tr>
<tr>
<td>5</td>
<td>131</td>
</tr>
<tr>
<td>6</td>
<td>134</td>
</tr>
<tr>
<td>7</td>
<td>136</td>
</tr>
<tr>
<td>8</td>
<td>166</td>
</tr>
<tr>
<td>9</td>
<td>170</td>
</tr>
<tr>
<td>10</td>
<td>175</td>
</tr>
<tr>
<td>11</td>
<td>180</td>
</tr>
<tr>
<td>12</td>
<td>185</td>
</tr>
<tr>
<td>13</td>
<td>188</td>
</tr>
<tr>
<td>14</td>
<td>193</td>
</tr>
<tr>
<td>15</td>
<td>198</td>
</tr>
<tr>
<td>16</td>
<td>199</td>
</tr>
<tr>
<td>17</td>
<td>205</td>
</tr>
</tbody>
</table>

LIST OF TABLES
<table>
<thead>
<tr>
<th>Table No.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>All Significant Cues (p &lt; .05) By Level of Responsibility, By Firm Affiliation</td>
<td>208</td>
</tr>
<tr>
<td>19</td>
<td>Only One Significant Cue (p &lt; .05) By Level of Responsibility, By Firm Affiliation</td>
<td>210</td>
</tr>
<tr>
<td>20</td>
<td>Standardized Balance Sheet Materiality Scores By Level of Responsibility and By Firm Affiliation</td>
<td>288</td>
</tr>
<tr>
<td>21</td>
<td>Summary of Level of Responsibility Effects For the Balance Sheet Judgments</td>
<td>291</td>
</tr>
<tr>
<td>22</td>
<td>ANOVA Table For Firm Effects For the Balance Sheet Judgments</td>
<td>294</td>
</tr>
<tr>
<td>23</td>
<td>Average Variance By Firm and Level of Responsibility Based on Standardized (Balance Sheet) Materiality Scores</td>
<td>298</td>
</tr>
<tr>
<td>24</td>
<td>Summary of Firm Effects For the Balance Sheet Judgments</td>
<td>300</td>
</tr>
<tr>
<td>25</td>
<td>ANOVA Table For Audit Technology Effects For the Balance Sheet Judgments</td>
<td>303</td>
</tr>
<tr>
<td>26</td>
<td>Summary of Audit Technology Effects For the Balance Sheet Judgments</td>
<td>306</td>
</tr>
<tr>
<td>27</td>
<td>All Significant Cues (p &lt; .05) For the Balance Sheet Judgments; By Level of Responsibility, By Firm Affiliation</td>
<td>310</td>
</tr>
<tr>
<td>28</td>
<td>Only One Significant Cue (p &lt; .05) For the Balance Sheet Judgments; By Level of Responsibility, By Firm Affiliation</td>
<td>311</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>Description</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The Lens Model and the Relationships Among the Cues, the Criterion, and Subjects' Responses</td>
<td>66</td>
</tr>
<tr>
<td>2</td>
<td>Indices Related to Human Judgments</td>
<td>68</td>
</tr>
<tr>
<td>3</td>
<td>(Implicit) Trade-offs Inherent in Auditors Materiality Judgments</td>
<td>99</td>
</tr>
<tr>
<td>4</td>
<td>Variables Used</td>
<td>101</td>
</tr>
<tr>
<td>5</td>
<td>Sample of Materiality Judgment Questionnaire</td>
<td>110</td>
</tr>
<tr>
<td>6</td>
<td>Experimental Design</td>
<td>117</td>
</tr>
<tr>
<td>7</td>
<td>Notation Regarding Comparisons To Be Made</td>
<td>123</td>
</tr>
<tr>
<td>8</td>
<td>Standardized Scores for the Income Statement Materiality Judgments; Plotted by Level of Responsibility</td>
<td>169</td>
</tr>
<tr>
<td>9</td>
<td>Standardized Scores for the Income Statement Materiality Judgments; Plotted by Firm Affiliation</td>
<td>178</td>
</tr>
<tr>
<td>10</td>
<td>Standardized Scores for the Balance Sheet Materiality Judgments; Plotted by Level of Responsibility</td>
<td>286</td>
</tr>
<tr>
<td>11</td>
<td>Standardized Scores for the Balance Sheet Materiality Judgments; Plotted by Firm Affiliation</td>
<td>293</td>
</tr>
</tbody>
</table>
CHAPTER I -- INTRODUCTION

The purpose of this chapter is to introduce the research project along with a brief description of its motivation. In addition, this chapter identifies the organization of the remainder of the dissertation.

The Concept of Materiality

Reference to the professional accounting and auditing literature suggests that, definitionally, there is very little controversy about the meaning of "materiality" as a concept. For example, in its Statement of Financial Accounting Concepts (SFAC) No. 2, the Financial Accounting Standards Board (FASB) commented (FASB [1980, para. 132]):

The essence of the materiality concept is clear. The omission or misstatement of an item in a financial report is material if, in the light of surrounding circumstances, the magnitude of the item is such that it is probable that the judgment of a reasonable person relying upon the report would have been changed or influenced by the inclusion or correction of the item.

The Auditing Standards Board (ASB) of the American Institute of Certified Public Accountants (AICPA) described materiality similarly in its Statement of Auditing Standards (SAS) No. 47 (AICPA [1983, para. 3]):

The concept of materiality recognizes that some matters ... are important for fair presentation of financial statements in conformity with generally accepted accounting principles, while other matters are not important.
Motivation for the Study

Despite apparent agreement about the general meaning of the materiality concept, there has been considerably less agreement about the implementation of that concept in specific circumstances. Prior research has consistently reported a lack of consensus among auditors in their materiality judgments. However, these prior studies have tended only to conclude that a consensus does not exist among auditors in their materiality judgments. Very little has been done to identify systematic differences among auditors which may be associated with differences in their materiality judgments. In addition, prior research efforts have not explored reasons why auditors may establish different materiality thresholds. Rather, there has been an apparent presumption that individual differences in the materiality judgments under study should not exist. Consequently, there has been a long-standing debate about whether specific materiality guidelines or criteria should be adopted to achieve similar materiality thresholds across the profession.¹

In addition to the continuing debate regarding the desirability of specific materiality guidelines, other factors have generated renewed interest in materiality as a research topic. For example, Holstrum and Messier [1982, p. 47] suggested that "... with the growing competition among public accounting firms and the increased emphasis on audit efficiency, the study of materiality in auditing may take on a more important role." One consequence of this emphasis on
efficiency may be to increase the importance which CPA firms place on planning stage activities. Yet, prior research has scarcely addressed the auditor's materiality judgment within a planning context.

This research focuses on auditors' materiality judgments in a planning context. The planning stage materiality judgment (termed "the preliminary judgment about materiality" by SAS No. 47) was chosen for three reasons. First, prior studies have almost exclusively examined auditors' "final" (i.e., evaluation stage) materiality judgments. Yet the professional auditing standards (SAS Nos. 22 and 47) indicate that materiality judgments are made at both the planning and evaluation stages. Hence, it is a legitimate topic for investigation, and the materiality judgment at the planning stage chronologically precedes the judgment at the evaluation stage.2 Second, insights gained with respect to auditors' planning stage materiality judgments may extend to the evaluation stage. Third, there should be actual auditing ramifications to these judgments at the planning stage. Preliminary judgments about materiality which are, in some sense, "too high" may lead to the gathering of insufficient audit evidence and audits which are not in accordance with GAAS. Preliminary judgments about materiality which are, in some sense, "too low" may lead to audits which are inefficient (i.e., "over-auditing").

In view of prior research involving audit judgment, it seems likely that the participants will differ in their specified planning
stage materiality thresholds. Ashton [1974, p. 153] observed, "Additional research is needed in order to determine why the judgments of particular auditors are inconsistent. For example, do individual factors, situational factors, or some combination of both account for judgment inconsistency?" This research focuses on selected "individual factors." Specifically, differences in the participants' judgments will be examined for systematic effects associated with level of responsibility (i.e., senior, manager, or partner), firm affiliation, and audit technology (i.e., whether the firms are regarded as relatively "structured" or "unstructured").

Based on the expectation that significant individual differences will occur, this study is designed to identify selected factors associated with the participants which may be related to differences in their materiality judgments. To the extent that there are systematic differences among auditors in their materiality judgments, it may be misleading for prior studies to refer to the variation in these judgments wholly as a "lack of consensus."

This study primarily emphasizes the effects of level of responsibility, firm affiliation, and audit technology on consensus among auditors in their planning stage materiality judgments. Consensus is examined in terms of location (regarding the central tendency of the different groups' materiality judgments) and variation (regarding the intra-group variability of the participants' judgments).
A secondary issue explores the extent to which differences among auditors in their planning stage materiality judgments may be attributable to differing "risk attitudes." This concept of risk attitude deals with the tradeoff of Type I and Type II errors (i.e., issuing a qualified audit opinion with respect to a client's financial statements that contain no "material" misstatements and issuing an unqualified audit opinion with respect to a client's financial statements that do contain "material" misstatements, respectively) that may be viewed as implicit in rendering materiality judgments at the aggregate financial statement level. The purpose is to identify factors associated with differing materiality judgments in order to explore reasons why the participants may differ in their stated materiality thresholds. If auditors differ systematically in their judgments due to different perceptions of the various consequences associated with these so-called alpha and beta risks, then evidence may be obtained that different materiality thresholds are consistent with the notion of auditors' utility maximization.

Objectives of the Dissertation

The three principal objectives of this dissertation involve the identification of factors hypothesized to be associated with systematic differences in consensus among auditors' planning stage materiality judgments. (As mentioned, consensus is examined in terms
of differences as to location and variation.) Specifically, these objectives are as follows:

1. to identify systematic differences in location and variation of auditors' planning stage materiality thresholds associated with their levels of responsibility;

2. to identify systematic differences in location and variation of auditors' planning stage materiality thresholds associated with their Big Eight firm affiliations;

3. to identify systematic differences in location and variation of auditors' planning stage materiality thresholds associated with the audit technologies of their Big Eight affiliations.

A subordinate research issue of interest is whether different risk attitudes regarding the tradeoff of Type I and Type II errors are associated with the levels of auditors' planning stage materiality thresholds. The purpose is exploratory and attempts to identify reasons why auditors may establish different materiality thresholds.

Research Methodology

In order to achieve the above research objectives, a judgment-capture experiment was conducted. As described in detail in Chapter II, such an approach is well-established in the accounting literature, in general, and with respect to auditors' materiality judgments, in particular. This research may be viewed essentially as a "lens" model study, although no criterion variable exists with which to assess the accuracy of the auditors' materiality judgments. The lens model framework was developed by Egon Brunswik in 1955 and has been applied extensively in the psychology and accounting literatures. In
particular, the lens model applies to human judgments which are based on a set of cues related to the environment; these cues are related in probabilistic terms to some criterion variable. Because no criterion exists with which to assess the accuracy of auditors' materiality judgments (for either the planning or evaluation stage), consensus was emphasized.6

Six of the Big Eight firms participated in this study. Three firms were viewed as highly structured (and were classified as "structured" in this study) and three were viewed as less structured (and were classified as "unstructured" in this study) based on Kinney [1986]. Each firm provided twelve auditors as participants—four each at the senior, manager, and partner levels. Thus, a total of 72 subjects participated. Each subject was asked to specify the appropriate planning stage materiality threshold at the aggregate financial statement level, separately for the income statement and balance sheet, for each of the 24 unique combinations of income (loss) before taxes, net sales, and total assets. Four cases were repeated in order to measure test-retest consistency.

Contributions

A major contribution of this research concerns the measurement of consensus. This research focuses on two aspects of consensus in order to assess the agreement among auditors in their materiality judgments: (1) tests of the significance of the differences in location
(regarding central tendency); and (2) tests of the significance of the differences in variation (regarding dispersion) of the participants' judgments. This is an improvement over the prior materiality studies which generally have measured agreement only in terms of pairwise correlations.

A second major contribution of this study concerns the judgment task. This research focuses on auditors' planning stage materiality judgments. Prior research has almost exclusively examined evaluation stage materiality judgments. The one published study which dealt with planning stage materiality judgments (i.e., Moriaruty and Barron [1979]) occurred prior to the issuance of any professional literature related to these preliminary judgments about materiality. Hence, Moriaruty and Barron acknowledged that the task was unfamiliar to their eight subjects. However, following the issuance of SAS No. 47, which discusses materiality judgments at both the planning and evaluation stages, it is appropriate to ask auditors to quantify their planning stage materiality thresholds at the aggregate financial statement level. Asking auditors to specify a dollar magnitude as a planning stage materiality threshold may be a more challenging task than asking them to consider a hypothetical financial reporting issue and then make a dichotomous judgment (i.e., "material" or "immaterial") or to express a materiality judgment on, for example, a seven-point scale. In that sense, one may argue that the planning stage materiality judgment is more complicated than the judgment at
the evaluation stage; and, as a result, it may be a more interesting task with which to explore auditors' judgments.

In addition, prior research has reported unclear results about the effect of experience on the consensus among auditors in their judgments, in general, and in their materiality judgments, in particular. This research examines the impact of level of responsibility (presumably related to experience) on the consensus among practicing auditors in the context of their planning stage materiality judgments. Further, insights may be gained about the effect of level of responsibility on these materiality judgments under different audit technologies.

This study will identify whether there are firm differences among the participating Big Eight CPA firms associated with their planning stage materiality judgments. In addition, any effects attributable to different audit technologies (regarding highly structured versus less structured firms) on the consensus of auditors' preliminary judgments about materiality will be identified. Although prior research has reported somewhat mixed results about Big Eight versus non-Big Eight or national versus non-national firm differences, relatively little is known about whether differences exist across Big Eight firms. Similarly, little is known about the effect which different audit technologies may have on auditors' judgments.

Another potential contribution of this research relates to the acknowledgment that auditors may rationally differ in their
materiality judgments. Prior research has reported a lack of consensus among auditors in their materiality judgments, but reasons for the divergence have not been explored. Instead, there has been a tendency for researchers to interpret these individual differences as evidence of a need for specific quantitative materiality guidelines from the standard-setting bodies. As a secondary issue, this research explores the extent to which the variation in auditors' materiality judgments may be attributed to differing risk attitudes. Auditors who are relatively more concerned with Type II errors (e.g., litigation by financial statement users) are expected to specify lower materiality thresholds than auditors who are relatively more concerned with Type I errors (e.g., client retention and practice development). Differences in auditors' materiality thresholds attributable to such differing risk attitudes would be rational (i.e., consistent with the notion of utility maximization).

Although this research cannot hope to resolve the debate about the desirability of establishing specific criteria to guide auditors in their materiality judgments, it does have implications with respect to that issue. If a fairly high level of consensus is discovered, then the position of the proponents of specific materiality criteria at the planning stage would appear to be weakened. Conversely, if a fairly low level of consensus is found, then their position would appear to be strengthened. However, the reasons auditors select different materiality thresholds would be germane in either case. For
example, if materiality thresholds are significantly affected by risk attitudes, then any guidelines would have to provide enough latitude to allow for such differences. Otherwise, one could expect strong opposition to any proposed guidelines.

Organization of the Dissertation

The remainder of the dissertation is organized in four chapters. Chapter II reviews the applicable literature. The major relevant pronouncements of the standard-setting bodies are presented first. The empirical studies of auditors' materiality judgments are discussed second. Other relevant research is reviewed third, including the applicable literature from psychology relative to the lens model and the role of attitudes. Chapter III identifies the particular research interests and hypotheses to be investigated. Further, Chapter III discusses the planned research approach, including the general methodology as well as the specific analyses to achieve the stated research objectives. Chapter IV reports the results of the statistical analyses. Finally, Chapter V offers a brief summary and some conclusions, identifies the limitations of the research, and presents suggestions for future research.
CHAPTER I FOOTNOTES

1. Woolsey [1973a] summarized several arguments supporting such materiality guidelines. First, a greater degree of uniformity and consistency in materiality judgments might enhance the credibility of financial statements to various users. Second, the existence of authoritative guidelines might clarify the auditor's legal liability and influence the outcome of litigation favorably for the auditing profession. Third, such standards might reduce the likelihood of disagreements between the client and the auditor, and strengthen the auditor's position in the event of disagreement. Pattillo and Siebei [1973] cautioned that, in the absence of specific criteria, government overseers may intercede and, thereby, displace the auditor's judgment.

Opponents of specific materiality guidelines argue that the materiality judgment is so complex and context-specific that the guidelines could not possibly specify all the factors relevant to the judgment. Those opposed to such guidelines cite SFAC No. 2 that "...no general standards of materiality could be formulated to take into account all the considerations that enter into an experienced human judgment" (FASB [1980, paragraph 131]). Therefore, such guidelines would tend to be arbitrary. Establishing materiality guidelines might substitute the rigidity of a rule-book in place of the flexibility of judgment based on the circumstances.

2. The distinction between the materiality judgments at the planning stage versus the evaluation stage will be discussed in Chapter II. Although the professional standards do not require quantification (i.e., specifying a dollar amount as a materiality threshold) of the planning stage materiality judgment, as a practical matter, quantification is the most reasonable way in which to incorporate the concept of materiality into audit planning (Zuber, et al. [1983, p. 42]). The evaluation stage materiality judgment may be viewed as a dichotomous judgment to disclose (not disclose) an item in the financial statements or notes, or to issue a qualified (unqualified) audit opinion if a judgment item under consideration is deemed to be material (inmaterial).
3. Audit technology will be discussed in more depth in Chapters II and III. This dissertation investigates the differences between the highly structured firms and the less structured firms, based on Kinney [1986]. Kinney classified the Big Eight firms as "structured," "intermediate," or "unstructured." This dissertation dichotomizes the six participating Big Eight firms along these lines, except that one firm classified as "intermediate" by Kinney was treated as "unstructured." Review of the participating firms' audit manuals suggested that, with respect to these planning stage materiality judgments, this particular firm was similar to the other firms classified as unstructured.

4. The role of attitudes is discussed in detail in Chapter II. The methodology employed with respect to attitudes has support in the psychology and accounting literatures. Kida [1980] provided the basis for this extension of attitudinal differences into the domain of auditors' materiality judgments. Kida focused on auditors' going concern qualifications and showed that different attitudes with respect to Type I and Type II errors could affect participants' decisions to issue such qualifications in an experimental context.

5. As stated in Chapter II, the descriptions of the lens model in this dissertation rely on the discussions provided by Ashton [1982b] and Dudycha and Naylor [1966].

6. As discussed in Chapter III, Ashton [1982a, p.75] indicates that consensus, self-insight, and judgment stability are often used when a criterion (such as accuracy or normativeness) is unavailable. Prior research efforts, discussed in Chapter II (e.g., Messier [1983] and Krogstad, et al. [1984]), have reported relatively low consensus, but relatively high self-insight and stability on materiality tasks. Since this study emphasizes comparisons among groups of auditors, rather than within individual auditors, consensus is of primary interest.
CHAPTER II -- LITERATURE REVIEW

Introduction

The purpose of this chapter is to summarize the relevant literature in order to position this research with respect to previous work. The literature is presented in three parts. First, the major materiality-related pronouncements and research efforts by the U.S. standard-setting bodies [i.e., the Financial Accounting Standards Board (FASB), and the Auditing Standards Board (ASB) of the American Institute of Certified Public Accountants (AICPA)] are presented. Reference is also made to pronouncements and research studies by the professional bodies of several other major industrial countries, since the issues under consideration are not strictly indigenous to the auditing profession in the United States. Second, the primarily empirical studies of auditors' materiality judgments are summarized. Third, some additional relevant research is reviewed.

Standard-setting Bodies

The assertion that the materiality judgments of auditors and accountants are of continuing concern to the profession can be substantiated by examining the recent professional literature. Both the FASB and the ASB have issued pronouncements which explicitly discuss materiality.
The general nature of the materiality judgment and the related issue of specific criteria or guidelines were among the first issues placed on the FASB's agenda. The FASB [1975, paragraph 41] stated, "In placing the topic of materiality on its agenda, the Board tentatively accepted the view that it would be desirable to establish criteria for the determination of materiality in financial accounting and reporting."

The outcome of the discussion memorandum was ten paragraphs pertaining to materiality in the text of Statement of Financial Concepts (SFAC) No. 2 (FASB [1980]). However, SFAC No. 2 contained no quantitative guidelines. Despite their observation (paragraph 163) that "The courts have stepped in to fill the gap," the Board concluded (paragraph 131) that their "... present position is that no general standards of materiality could be formulated to take into account all the considerations that enter into an experienced human judgment."

The Board conceded that it may reconsider the appropriateness of issuing general quantitative materiality guidelines or, alternatively, guidelines applicable to specific standards on a piecemeal basis.¹

In 1978, the ASB issued Statement on Auditing Standards (SAS) No. 22, entitled "Planning and Supervision." SAS No. 22 identified several items to be considered when planning the audit (AICPA [1978, paragraph 3]). Of particular interest here was their reference to "preliminary estimates of materiality levels for audit purposes."
However, no further discussion was offered, nor were any guidelines provided.

Following four years of study by an AICPA taskforce, the ASB issued an exposure draft entitled "Materiality and Audit Risk in Conducting an Audit" in 1982. The related SAS No. 47, entitled "Audit Risk and Materiality in Conducting an Audit" (AICPA [1983]), was issued the following year.²

SAS No. 47 recognizes that materiality judgments occur both at the planning stage of an engagement and at the evaluation stage where the audit evidence is evaluated. Although the auditor should consider materiality for planning purposes, this "preliminary judgment about materiality" need not necessarily be quantified (paragraph 10). For planning purposes, the SAS indicates that it is appropriate to consider materiality with respect to the individual financial statements in the aggregate (paragraph 12):

Materiality levels include an overall level for each statement; however, because the statements are interrelated, and for reasons of efficiency, the auditor ordinarily considers materiality for planning purposes in terms of the smallest aggregate level of errors that could be considered material to any one of the financial statements. For example, if [the auditor] believes that errors aggregating approximately $100,000 would have a material effect on income but that such errors would have to aggregate approximately $200,000 to materially affect financial position, it would not be appropriate ... to design auditing procedures that would be expected to detect errors only if they aggregate approximately $200,000.

Commenting on the interrelationship between the "preliminary" and the "final" materiality judgments, the SAS (paragraph 13) states,
"Although the auditor should be alert for errors that could be qualitatively material, it ordinarily is not practical to design procedures to detect them." The planning stage materiality judgment hinges primarily on quantitative factors, whereas the evaluation stage materiality judgment depends to a greater extent on qualitative factors. Even though the two materiality judgments theoretically would be the same if the information at the evaluation stage were available for planning, as a practical matter, the judgments will differ.

In summary, the SAS requires auditors to explicitly consider materiality in audit planning. Although materiality is not required to be quantified, Zuber, et al. [1983, p. 42] point out that "... quantification is the most practical way to consider such an estimate in audit planning." SAS No. 47 provides a broad conceptual discussion of the impact of materiality on the conduct of an audit, and clarifies the earlier reference to "preliminary estimates of materiality" in SAS No. 22. However, it contains no quantitative criteria or guidelines to assist practitioners in their materiality judgments at either the planning or evaluation stages.

Recently, a practicing CPA expressed his firm's desire for specific materiality criteria in "Letters to the Editor" of the Journal of Accountancy (Godick [1983, p. 101]). He observed that "... the nitty-gritty dollar determination of materiality hasn't yet been addressed." The lack of agreement about such guidelines can be seen
in the two published responses. Robert Berliner, chairman of the AICPA's materiality and audit risk task force, replied that the SAS does not require quantification of the preliminary judgment about materiality. Robert K. Elliott, a partner in Peat Marwick, agreed that more guidance would be desirable and published a tabular approach to assist practitioners in quantifying materiality at the planning stage.

In a separate article, Berliner [1983, p. 19] commented, "Some will say that the ASB 'copped out' by failing to provide quantitative guidelines for practitioners to follow. But this issue is ... under continuing consideration." It appears, therefore, that auditors' materiality judgments (at both the planning and evaluation stages), in general, and quantitative guidelines, in particular, are issues of continuing practical importance to U.S. practitioners.

The issue of quantitative materiality guidelines has also been addressed by professional bodies in several other countries, including Canada, the United Kingdom, Australia, and New Zealand. In 1965, the Canadian Study Group on Audit Techniques of the Canadian Institute of Chartered Accountants (CICA) published Materiality in Auditing. In 1968, the Institute of Chartered Accountants in England and Wales issued its Statement V10, "The Interpretation of 'Material' in Relation to Accounts." In 1974, the two authoritative accounting bodies in Australia, the National Council of The Institute of Chartered Accountants in Australia and the General Council of the
Australian Society of Accountants, jointly issued a Statement of Accounting Standards on "Materiality in Financial Statements." A similar pronouncement was issued by the New Zealand Society of Accountants in 1977. Although these pronouncements differed in their authoritativeness and emphasized the importance of professional judgment in differing degrees, each identified factors to be considered in making materiality judgments, and each presented quantitative rules of thumb.

The Accountants International Study Group [1974] issued a report on the materiality-related practices of Canada, the United Kingdom, and the United States. The Study Group concluded (paragraph 30), "Quantitative guidelines within broad parameters are usually used in practice and we believe they could be developed by the authoritative bodies in the three nations."

David Selley, Auditing Standards Director of the CICA, discussed several of the major issues and historical developments related to materiality judgments (Selley [1984]). With respect to the materiality policies of 16 of the largest Canadian accounting firms, Selley reported (p. 18):

... 11 of the 16 provide quantitative guidelines to their auditing professional staff for determining what is material and that all 11 of them use as the primary or as one among several presumptive guidelines 5% to 10% of normal pre-tax net income for most profit-oriented clients. Two more firms have guidance material in process.

In summarizing the world wide development of materiality criteria, Selley commented (p. 18), "With the exception of the U.S. profession
(FASB and AICPA), study groups and standard-setting bodies in several important Anglo-Saxon-dominated professions have recommended and/or adopted [rules of thumb] in some form."

Thus, developments both within and outside the United States suggest that the FASB and ASB may be subjected to increased pressure to adopt more specific materiality guidelines. As noted above, the FASB and ASB have expressed their reluctance to establish such guidelines due to the many possible factors which might enter into these judgments. However, with respect to materiality judgments in auditing contexts, one might anticipate that the proponents of explicit criteria may concede the impracticality of issuing materiality guidelines applicable to the evaluation stage, and focus their arguments on the planning stage. They may argue that, for planning purposes, SAS No. 47 acknowledges that the materiality judgment is primarily a function of the quantitative variables and the qualitative variables are substantially ignored at this level. Therefore, in their view, the preliminary judgment would not be required "... to take into account all the considerations that enter an experienced human judgment" (FASB [1980, paragraph 131]). To the extent that more specific materiality guidelines are indeed under continuing consideration by the FASB and ASB, it would be helpful to better understand how and why auditors may differ in their judgments.
Empirical Research Regarding Materiality

Holstrum and Messier [1982] classified research regarding materiality as questionnaire survey, archival, or judgment-capture experiments. The following summary of the literature focuses on that research dealing with materiality judgments of auditors (even though these studies have tended to deal with accounting rather than auditing contexts\textsuperscript{4}). The earliest research into materiality judgments tended to be the questionnaire survey type (e.g., Woolsey [1954a and b]). In addition, some of the more recent, practice-oriented studies have been of the questionnaire survey type (e.g., Dyer [1975] and Pattillo [1976]). These studies are summarized in Table 1.

These questionnaire survey studies of accountants' and auditors' materiality judgments have generally shown consistent results (although the results of such surveys have not been amenable to statistical analysis). Relative amounts have been more important than absolute amounts, and the most important single factor has been the particular item in question as a percentage of net income (or income before taxes).

Woolsey [1954a] surveyed several hundred individuals and different business organizations as to factors affecting the classification of three particular items (gain from the sale of fixed assets, write-off of unamortized bond discount, and loss due to an earthquake casualty loss) as material versus immaterial. Woolsey reported four conclusions: (1) relative amounts were more important
Table 1
Summary of Questionnaire Survey Studies

1. **Woolsey [1954a]**

   **Subjects:** Approximately 120 respondents from public accounting, industry, and user groups (response rate was about 25%).

   **Context of Judgments:** 1. Gain from sale of fixed assets; 2. write-off of bond discount; and 3. casualty loss.

   **Conclusions:** 1. Relative amounts were more important than absolute amounts; 2. item as a percentage of net income was the most important factor; 3. items in the range of 5-15% of net income hinged on other considerations; and 4. thresholds for national firms were about 4 percentage points higher than those for non-national firms.

2. **Woolsey [1954b]**

   **Subjects:** Approximately 120 respondents from public accounting, industry, and user groups (response rate was about 20%).

   **Context of Judgments:** 1. Long-term leases; 2. marketable securities; and 3. contingent liabilities.

   **Conclusions:** 1. The basis for determining the relative percentage was influenced by the context under consideration; and 2. one factor was usually used to determine whether the item was obviously material or immaterial.

3. **Dyer [1975]**

   **Subjects:** 85 CPA-respondents of whom 36 were from national firms and 49 were from non-national firms (response rate was about 30%).

   **Context of Judgments:** Same as Woolsey [1954a].

   **Conclusions:** Same as Woolsey [1954a], except that Dyer did not find any difference between the national and non-national firms.
Table 1 (continued)


Subjects: Approximately 700 respondents from public accounting, industry, and user groups (response rate was about 45%).

Context of Judgments: 1. Extraordinary items; 2. contingencies; 3. accounting changes; 4. policy disclosures; 5. business segments; 6. cumulative judgments; and 7. interim disclosures.

Conclusions: 1. Groups varied, but the auditors tended to be the most conservative; 2. "determinant criteria" (quantitative factors) formed the basis for tentative judgments and "modifying criteria" (qualitative factors) influenced the final judgments; and 3. the item as a percentage of net income was the most influential of 7 determinant criteria considered.
than absolute amounts; (2) the most important single factor was the relationship of the loss to current net income; (3) an item in the range of five to fifteen percent of net income could be classified as either material or immaterial, and the judgment usually depended upon other circumstances; (4) CPAs affiliated with national firms tended to have materiality thresholds about four percentage points higher than their regional and local firm counterparts.

In a related study, Woolsey [1954b] considered judgments involving long-term leases, marketable securities, and contingent liabilities. The results suggested that the basis for determining the relative percentage of the item in question may vary depending upon the nature of the item. For example, the primary factor in the lease case was average net income and current net income was most important in the marketable security case. However, in the contingent liability case, the percentage of the item relative to working capital as well as simply the absolute amount of the item were regarded as the most important factors by an equal number of respondents. Woolsey suggested that the auditors generally had in mind one important factor to determine if an item could be classified as obviously material or immaterial. If this comparison yielded no clear answer, the auditor then considered other factors by which to make the judgment. (This suggests the possibility of interactions among cues.)

Dyer [1975] substantially replicated Woolsey [1954a]. Again, the survey focused on unusual items under three scenarios (the same as
Woolsey's). The subjects were classified as national or non-national. They were asked to identify the single most important factor relative to which the amount in question would be compared for purposes of determining materiality. Then they were asked to classify each situation as material or immaterial for each set of cues. Materiality "thresholds" were inferred based on the most important factor identified and the lowest value of that factor for which the subject decided that the item in question was material.

Dyer concluded that the clearly most important factor in the materiality judgment was the relationship of the item to net income—similar to Woolsey. However, the materiality "thresholds" were substantially lower for all subjects on average and there were no clear-cut differences between national and non-national subjects. Dyer attributed this difference with respect to Woolsey's study to environmental changes (such as the legal environment) over the intervening twenty years.

Pattillo [1976] surveyed nearly 700 subjects representing industry, public accounting, various user groups and academia. A series of 28 cases (each containing six levels of quantitative cues) were mailed to the participants. For each judgment situation (involving unusual and extraordinary items, contingencies, accounting changes, accounting policy disclosures, business segments, cumulative materiality judgments with both similar and dissimilar items, and interim report disclosures) the subjects were asked to identify the
appropriate materiality threshold, applicable to the evaluation of year-end financial statements. About two-thirds of the subjects provided a point estimate, whereas about one-fourth offered an interval estimate; presumably, the remainder responded that a given situation was either always material or always immaterial. One general conclusion was that the judgments among and within groups varied considerably. Yet there was a tendency for the auditors to be more conservative than the other categories of participants.

Pattillo identified two categories of materiality criteria. First, "determinant criteria" are the factors which provide the basis for the tentative judgment. These factors are quantifiable and relate to selected elements of the financial statements. Second, "modifying criteria" represent additional considerations which affect the final materiality judgment. These factors are generally qualitative in nature.

Seven determinant factors were incorporated into each case: (1) the absolute dollar amount of the item in question; (2) the item as a percent of sales; (3) the item as a percent of net income; (4) the item as a percent of total assets; (5) the item as a percent of total liabilities; (6) the item as a percent of stockholders' equity; and (7) the effect of the item on earnings per share. For selected cases, the subjects were asked to assign measures of importance from "1" ("high importance") to "4" ("low importance") to each of the determinant factors. Pattillo reported the following results:
<table>
<thead>
<tr>
<th>Item as a % of net income</th>
<th>1.75</th>
<th>Moderate to high</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute dollar amount</td>
<td>1.92</td>
<td>Moderate to high</td>
</tr>
<tr>
<td>Effect on E.P.S.</td>
<td>2.04</td>
<td>Moderate</td>
</tr>
<tr>
<td>Item as a % of equity</td>
<td>2.70</td>
<td>Low to moderate</td>
</tr>
<tr>
<td>Item as a % of assets</td>
<td>2.88</td>
<td>Low to moderate</td>
</tr>
<tr>
<td>Item as a % of liabilities</td>
<td>3.22</td>
<td>Low to none</td>
</tr>
<tr>
<td>Item as a % of sales</td>
<td>3.32</td>
<td>Low to none</td>
</tr>
</tbody>
</table>

Much of the materiality research in the 1960s and early 1970s tended to be of the "archival" type (e.g., Bernstein [1967]; Neumann [1968]; Copeland and Fredericks [1968]; and Frishkoff [1970]). In addition, Morris, Nichols, and Pattillo [1984] was a recent archival study of auditors' materiality judgments. These studies are summarized in Table 2.

These studies examined actual financial statements in search of a pattern indicative of the materiality criteria applied to those financial statements. Such studies have been unsuccessful in detecting a "dividing line" (based on the item in question as a percentage of net income) between immaterial and material items. Bernstein and Neumann interpreted their results as evidence of a lack of consensus among auditors in the determination of materiality.

Bernstein [1967] examined a sample of 324 income statements containing extraordinary gains or losses. (The accounting treatment of such items was specified at that time in Accounting Research Bulletin No. 43, Chapter 8, and required material extraordinary items to be excluded from income.) Bernstein attempted to infer materiality in relation to net income based upon any apparent "dividing line" between the magnitude of extraordinary items included in income versus those items which were excluded. In other words, an extraordinary
Table 2
Summary of Archival Studies

1. Bernstein [1967]

Issue: Whether a threshold could be identified regarding extraordinary items, such that amounts above that threshold were excluded from income (as "material") and amounts below that threshold were included in income (as "immaterial").

Sample: 324 financial statements containing extraordinary items.

Conclusion: No pattern emerged and, as a result, consensus among auditors was inferred to be low.

2. Neumann [1968]

Issue: Same as Bernstein [1967].

Sample: 300 Fortune 500 companies' financial statements with changes in accounting principles (listed on the NYSE for 12/31/54 to 12/31/64).

Conclusion: Same as Bernstein [1967].

3. Copeland and Fredericks [1968]

Issue: The extent to which the degree of disclosure made with respect to additional stock issuances was related to the relative size (i.e., materiality) of the new issue.

Sample: 200 listing applications of firms on the NYSE for 1964.

Conclusion: Materiality was positively (but not significantly) associated with their measure of the degree of disclosure.

4. Frishkoff [1970]

Issue: The extent to which auditors' qualifications as to consistency (used as a surrogate for materiality) can be predicted by multiple discriminant analysis.

Sample: 190 annual reports for 1963 containing changes in accounting principles (130 were qualified).
Table 2 (continued)

Conclusions: 1. Only 3 of the 17 variables considered were significant at the .10 level (the item as a percent of net income, the amount of net worth, and a dummy variable which identified reclassifications from more consequential changes); and 2. the item as a percent of net income was not a good predictor of qualification decisions.

5. Morris, Nichols, and Pattillo [1984]

Issue: Whether consistency qualifications with respect to mandatory changes in interest capitalization policies (SFAS No. 34) were driven by quantitative financial statement relationships.

Sample: 221 annual reports for 1979/80 or for 1980/81 affected by SFAS No. 34.

Conclusion: No pattern emerged and, as a result, consensus among the auditors was inferred to be low.
item excluded from income was interpreted to be material, whereas such an item included in income was presumed to be immaterial. Bernstein concluded that there was no consensus among auditors as to the determination of materiality, since no pattern emerged from his analysis. He then proceeded into a broad discussion of materiality and advanced his arguments for the establishment of specific criteria to determine materiality. Bernstein argued (p. 89), "No pronouncement can be stronger than the most weakly defined concept used in formulating it."

Neumann [1968] used a similar approach, but instead of focusing on extraordinary items, Neumann examined changes in accounting principles. Again, the objective was to infer materiality based upon a review of actual financial statements. Specifically, Neumann selected a sample of 300 firms from the Fortune 500 companies which were listed on the NYSE for the period 12/31/54 to 12/31/64. Neumann was interested in whether the sampled firms disclosed any changes in accounting principles following two events which had rather pervasive effects. The first event was the Revenue Act of 1954, which permitted broad usage of accelerated depreciation methods for tax purposes. The second event was APB Opinion No.4, which permitted either the deferral method or the flow-through method for reporting investment tax credits. Previously the deferral method had been advanced by the APB as the appropriate method. In response to these two environmental changes, Neumann anticipated a large number of changes in accounting
principles (of a comparable nature) in 1954 and again in 1964. By choosing changes affecting many firms across industry lines, Neumann argued that the "qualitative aspects" of materiality were effectively held constant. Hence, the determination of materiality hinged upon its "quantitative dimensions."

The audit reports were examined for those firms which reported such a change in accounting principle in 1954 and/or 1964. An audit opinion which was qualified as to consistency was presumed to imply a material change in accounting principle, whereas an unqualified opinion was construed to imply immateriality. The magnitude of the effect of the change in accounting principle relative to net income (including the change as reported) was used to determine whether any "dividing line" as to materiality versus immateriality could be observed in the application of the consistency reporting standard. As did Bernstein before him, Neumann concluded that the disclosure decision under consideration did not vary systematically with his measure of materiality, and added (p. 12), "It is my opinion, as the result of this study, that the failure to more specifically define the dimensions of materiality has led to a lack of consensus as to the 'when' and 'how' of the implementation of the consistency standard." Neumann argued that materiality should not be a criterion for the consistency qualification.

Copeland and Fredericks [1968] developed a measure for the extent of disclosure across sets of financial statements and related that
measure to the concept of materiality. In this regard, the authors examined a sample of 200 listing applications of firms on the NYSE for 1964. The object of their interest was any intention to issue additional shares of stock. They defined the materiality variable to be the number of shares listed in the application divided by the total number of shares outstanding prior to the new issuance. In order to quantify the degree of disclosure, the authors prepared a checklist of items of information one might expect to be disclosed, and organized these items according to the reasons for issuing the additional shares. (There were six such reasons considered by the authors.) The numerical measure for the extent of disclosure was calculated separately for each reason for issuing the additional shares in the aggregate for all firms issuing stock for that reason. The measure was the total number of "check marks" (indicating the presence of disclosure regarding particular items of information contained in the authors' checklist) divided by the total number of checkmarks possible for the number of firms issuing additional shares for the specified reason. The result was the percentage of the possible number of disclosures of designated items of information which were in fact disclosed in the filings included with the listing applications.

After the materiality and disclosure variables were quantified, Spearman's rank correlation test was used to measure the extent to which materiality and degree of disclosure were correlated. The result was a correlation coefficient equal to +.595; the authors
concluded that there was a positive association between materiality and disclosure, although the coefficient was not significant at the .05 level.

Frishkoff [1970] focused on consistency qualifications in his study of auditors' materiality judgments, similar to Neumann [1968]. However, whereas Neumann considered only one variable, Frishkoff considered seventeen. In addition, Frishkoff introduced multiple discriminant analysis in the study of materiality.

Frishkoff reviewed approximately 2200 annual reports for 1963. His sample was comprised of 190 reports involving financial statements containing changes in accounting principles. Of these, 130 contained qualified opinions and 60 were unqualified. Again, it was presumed that qualification implied materiality.

Using step-wise discriminant analysis, Frishkoff found that only three independent variables were significant at the .10 level. These were: the absolute value of the effect divided by net income, the amount of net worth, and a dummy variable distinguishing reclassifications from other changes. The last two were significant at about the .065 level, whereas the first was significant at the .025 level.

Based upon the single independent variable, percent of net income, Frishkoff attempted to identify a "dividing line" as to immaterial versus material amounts. He inferred that this distinction occurred at about 25 percent of net income. This was surprisingly
large relative to prior studies (e.g., Bernstein suggested ten to fifteen percent of net income as a line of demarcation). In addition, Frishkoff pointed out that there was virtually no ability to predict accurately the actual sample results based upon this guideline.

Morris, Nichols, and Pattillo [1984] studied auditors' materiality judgments with respect to disclosure of capitalized interest costs under Statement of Financial Accounting Standards (SFAS) No. 34. Since SFAS No. 34 changed the financial reporting treatment of interest costs under certain circumstances, a consistency qualification in the audit report was assumed to imply a material item in the financial statements; an audit report which did not mention the change in treatment of interest was assumed to imply immateriality. The authors argued that they expected consensus to be relatively high in their study for two primary reasons. First, only auditors were considered by focusing on the audit opinions (thereby excluding differences in the materiality judgments of other interested parties, such as users, management, and the courts). Second, only items of a similar nature were considered, i.e., capitalized interest costs.

Morris, et al., made the following observation about the consistency qualifications versus the unqualified audit opinions (p. 553):

Although the mean value of the material subsample is [higher than] the mean value of the immaterial subsample for each judgment factor [i.e., the item as a percent of net income and as a percent of net plant], the distribution of values in each subsample encompasses the entire range of values. The empirical results suggest that there is no general rule such that values less than X percent are always immaterial
and values greater than Y percent are always material. Instead, there appears to be one big area of overlap, regardless of which factor is investigated.

Therefore, the authors concluded that there is an absence of consensus among auditors in these materiality judgments. In addition to suggesting additional research to develop general materiality guidelines, the authors recommended that a company's management disclose its materiality criteria in its financial report, in order to facilitate the comparison of financial statements across companies.

Most of the recent empirical research into auditors' materiality judgments has been of the judgment-capture experiment type (e.g., Boatsman and Robertson [1974]; Moriarity and Barron [1976] and [1979]; Ward [1976]; Newton [1977]; Hofstedt and Hughes [1977]; Firth [1979]; Schultz and Reckers [1981]; Emery, Thakkar, and Moriarity [1982]; Messier [1983]; and Krogstad, Ettenson, and Shanteau [1984]). The participants in the studies were asked to express their materiality judgments based on the case materials presented. These studies are summarized in Table 3.

Boatsman and Robertson [1974] used multiple discriminant analysis to model the materiality judgments of groups of CPAs and securities analysts. Rather than examining actual financial statements in conjunction with the auditors' observed opinions, the authors developed a set of hypothetical cases for which individual judgments were required.
Table 3
Summary of Judgment-capture Experiment Studies

1. **Boatsman and Robertson [1974]**

Subjects: 18 partners from national CPA firms and 15 securities analysts.

Experimental task: Sort 30 cases with different reporting issues into 3 sets: line item disclosure, footnote only disclosure, or no disclosure.

Conclusions: 1. Multiple discriminant models were 63% accurate in predicting subjects' classifications; 2. when disclosure of some kind was viewed as "material" and no disclosure was viewed as "immaterial," the models were 84% accurate; and 3. the item as a percent of net income contributed 73% of the predictive ability.

2. **Morarity and Barron [1976]**

Subjects: 15 audit partners from national CPA firms.

Experimental task: Rank 18 cases in order of increasing materiality, based on the 3 financial statement cues provided.

Conclusion: Most of the subjects' judgments could be effectively modeled by linear models (using conjoint measurement), although the models for 4 subjects were nonlinear.

3. **Ward [1976]**

Subjects: 24 audit managers or partners from national CPA firms.

Experimental task: Sort 20 cards (each identifying a consequence to the auditors if a reporting error occurred) into stacks of 1, 5, 8, 5, & 1 card(s)—from most to least important consideration.

Conclusion: Based on the Q-sort technique, subjects were in reasonable agreement as to the ordering of the 20 consequences.

4. **Newton [1977]**

Sample: 19 partners of CPA firms—including Big Eight, regional, and local firms.
Table 3 (continued)

Experimental task: Identify preferences among different sets of "gambles" (for 3 different contexts); these involved different dollar amounts. Probabilities were elicited with respect to qualification decisions.

Conclusions: 1. Most subjects were consistent as to risk-seeking or avoidance behavior across different contexts (most were risk averse); and 2. subjects tended to use probabilities (at least implicitly) in their decisions to qualify.

5. Hofstede and Hughes [1977]

Subjects: 19 MBA students.

Experimental task: Whether a loss on the write-down of a subsidiary should be classified as an extraordinary item (on a 100-point scale).

Conclusions: 1. The explanatory ability of the ANOVA models decreased when individuals were aggregated; 2. simple models (reflecting only main effects) had about as much explanatory ability as more complicated models (including interaction terms); and 3. individual differences were conspicuous in several respects.

6. Morar and Barron [1979]

Subjects: 8 audit partners from 1 national CPA firm.

Experimental task: Specify the planning stage materiality threshold for each of 30 cases.

Conclusions: 1. The task was unfamiliar to the subjects and the experimental materials lacked "mundane reality;" and 2. the poor consensus was influenced by the diversity of the partners' backgrounds.

7. Firth [1979]

Subjects: 150 people--30 each from 3 Big Eight CPA firms, industry, and user groups.

Experimental task: Identify whether an extraordinary item warranted separate disclosure for each of 30 cases.
Table 3 (continued)

Conclusions: 1. The item as a percent of income (before extraordinary items) had the greatest significance; and 2. no significant differences were found among the groups (including no differences among the 3 firms).

8. Schultz and Reckers [1981]

Subjects: 64 audit partners from 1 national CPA firm.

Experimental task: Identify the minimum probability for which footnote disclosure was required for 2 different contingencies.

Conclusion: A lower probability was required for disclosure of the "high" materiality case than for the "low" materiality case.

9. Emery, Thakkar, and Moriarity [1982]

Subjects: 16 securities analysts and 14 commercial lenders.

Experimental task: Same as Moriarity and Barron [1976].

Conclusions: 1. The structural models were linear for most of the subjects, similar to Moriarity and Barron's conclusion; and 2. in comparison to Moriarity and Barron, the structural models of the financial analysts and the loan officers were similar to those of the auditors.

10. Messier [1983]

Subjects: 22 Big Eight audit partners and 7 non-Big Eight audit partners.

Experimental task: Rate the materiality of a write-down of inventory (on a 7-point scale) and identify the probability that the item would be separately disclosed for each of 32 cases.

Conclusions: 1. Net income was significant for 27 of the partners; 2. based on ANOVA for the materiality judgments, there was a significant difference between the Big Eight and non-Big Eight subjects; and 3. experience was not significant, although the explanatory ability of the ANOVA models was greater for the relatively more experienced subjects.
Table 3 (continued)


Subjects: 10 audit partners, 11 audit seniors (the auditors were from 5 Big Eight firms), and 11 accounting students.

Experimental task: Identify the materiality of a proposed adjusting entry (on a 100-point scale) for each of 32 cases.

Conclusions: 1. Effect on net income was the dominant variable for the auditors; 2. no differences were identified between the seniors and partners; 3. the students' judgments differed from the auditors' judgments; and 4. at least 1 qualitative variable was significant for about 2/3 of the auditors.
The following eight variables were considered in the cases:

1. The nature of the item. A dummy variable designated whether the item was (a) a gain or loss on the sale of noncurrent assets, (b) a change in accounting principle, or (c) an uncertainty.

2. The absolute value of the item divided by the current year net income (excluding the items under consideration).

3. The absolute value of the item divided by total revenue (for positive items) or total expenses (for negative items).

4. The amount of the items divided by total working capital.

5. The percentage change in net income relative to the prior year (excluding the items under consideration).

6. A dummy variable coded "1" if recognition of the item changed net income into a net loss or vice versa and "0" otherwise.

7. The absolute value of the item.

8. Risk defined as common stock price volatility measured on a five-point scale.

In developing their 30 hypothetical cases, the authors selected values for all the variables (except risk, which was based upon Standard & Poor's data) from a sample of 103 actual annual reports. Thus, it was hoped the magnitudes of the variables would be realistic. Since it was assumed that these variables would be intercorrelated in reality, the authors desired values of the variables which reflected a reasonable variance-covariance matrix. A mathematical technique was
employed to generate values for the variables having the mean values observed from the actual reports while also having a variance-covariance matrix consistent with the actual reports. The technique generated data in the form of the multivariate normal distribution.

The subjects were eighteen partners from national CPA firms and fifteen securities analysts. Their task was to sort the 30 cases into three disjoint sets corresponding to the desired reporting treatments (i.e., line-item disclosure in the body of the financial statements, footnote only disclosure, and no disclosure of any kind). In the sense that the subjects were instructed to classify the cases as to how they should be treated, the task was normative.

Based upon the 990 observations (30 cases for each of 33 subjects), three multiple discriminate models were determined, one for each possible disclosure treatment. Models were not prepared for individual decision makers, but rather aggregated all subjects' judgments. The models were then used to predict classifications for each case for each individual. The models accurately predicted 63 percent of the classifications as actually made by the subjects. Further, the authors reported that, if line-item disclosure and footnote disclosure categories were combined as indicating materiality and no disclosure interpreted as immateriality, the model was 84 percent accurate in distinguishing material from immaterial cases.

One variable, the absolute value of the item divided by current net income, contributed 73 percent of the predictive accuracy. The
variable identifying gains and losses on noncurrent assets contributed 24 percent of the predictive accuracy. Finally, the risk variable contributed two percent of predictive accuracy and the other variables combined to contribute the remaining one percent.

A separate analysis was performed to investigate whether there were differences between the judgments of CPAs versus securities analysts, despite the relatively small sample sizes. Several different approaches were used, including nonparametric rank correlation, cross-prediction, cluster analysis, and discriminate analysis. The results suggested that there were no systematic differences in the aggregate judgments of CPAs versus securities analysts.

In view of the apparent importance of the amount of the items as a percentage of net income, Boatsman and Robertson attempted to identify a "dividing line" between immateriality and materiality based solely upon that one variable. Accordingly, they identified four percent of net income as the appropriate distinction and determined the predictive accuracy to be 65 percent (as compared to 84 percent for the full model). However, they pointed out that this simple dividing line tended to err on the side of underdisclosure.

Moriarity and Barron [1976] used a different approach. They observed (p. 320), "Four possible sources of difference may exist among auditors' decision models which would explain the lack of consensus in materiality judgments. These are: (1) the variables
deemed relevant to the decision; (2) the materiality thresholds; (3) the scale values of the variables affecting the decision; and (4) the basic forms of the decision models. Suggesting that other researchers had considered issues (1) and (2), the authors indicated that their investigation of how individuals make materiality judgments would focus on issues (3) and (4).

They used an approach known as conjoint measurement (CJM) to explore the apparent structural form (i.e., linear, multiplicative, or some combination) of the models used by decision makers in the context of a particular materiality judgment. The CJM technique was different from earlier methodologies employed to investigate materiality judgments in that only ordinal characteristics of materiality decisions were needed for CJM. This had the effect of segregating the issue of materiality thresholds from the analysis. The contributing variables were also treated as ordinal and, hence, there was no assumption as to the linearity of the contributions based on the size of the variables.

The authors selected three variables based upon their review of the applicable literature; they noted that, as a practical matter, CJM was limited to three variables, despite the existence of procedures to mitigate that restriction. The three variables they selected were net income, earnings trend, and firm size. A factorial design was used involving three levels of income, three levels of asset size, and two
levels of earnings trend. Hence, eighteen sets of financial information were prepared, corresponding to the $3^2 \times 2$ design.

The subjects were fifteen audit partners of national CPA firms and their task was to rank the eighteen cases in order of increasing materiality. CJM techniques were then applied to determine the structural form of the decision model for each individual partner. Following that, the contribution of the levels of the individual variables to the subjects' separate decision models was assessed.

Of the fifteen subjects, eight were described by the authors as "perfectly additive" and three were "nearly additive." However, the remaining four were described by "configural models" in which there were interactions among levels of the independent variables. Even though a clear majority used models of the linear form, scaling procedures indicated that three distinct perfectly linear models and three distinct nearly linear models were used by the partners. The authors concluded (p. 337) that "... linear homogeneous models (as might be used in regression studies) are not appropriate for modeling the auditor's materiality decision models."

Further, because of the nonlinear models noted, they concluded that linear models are not appropriate in every case. However, they commented (p. 337), "To conclude that the additive model would be appropriate for all auditors would be appealing for three reasons. (1) Such a conclusion would allow more efficient research designs for examining scaling, (2) the compensatory nature of an additive model
appeals to intuition, and (3) this study suggests its predominance in use."

Ward [1976] focused on (p. 138) "... the degree of consensus among professional accountants concerning perceptions about the relative importance of factors associated with issues such as materiality." The subjects were 24 audit partners or managers with national CPA firms. The methodology employed was the Q-sort technique, which, similar to CJM described above, required subjects to specify only rankings. The subjects were provided with information about a hypothetical company's internal controls and a set of financial statements. The subjects were asked to consider the scenario of an error within the valuation of accounts receivable having a corresponding effect on net income (first, overstatement and second, understatement situations were considered).

The author provided 20 index cards expressing various implications to the auditors and to the client firm, given that the error affecting accounts receivable and net income was not detected by the auditors. For example, these included "degree to which audit complied with professional standards," "personal liability," and "loss of future engagements with the client," among others. The task required the subjects to sort the 20 cards into one of five different categories as to the degree of importance of the individual items described on each card. In an attempt to approximate a normal distribution, Ward directed the subjects to sort the cards into stacks
of one, five, eight, five, and one card(s) corresponding to the
category in which the item was classified by the individual subject.
Mean scores were calculated across all subjects and the items arrayed
from most to least important.

Nonparametric statistical procedures were used to compare
subjects' rankings to evaluate whether there was agreement as to the
order of the importance of the items. Ward concluded (p. 149) that
"... the consensus ordering or general perception of the materiality
construct in auditing is fairly well established and shared in common
among professional accountants as a class ...."

In addition, Ward provided graphs of various functions (linear,
bounded linear, step, logistic, modified exponential, and
exponential). The subjects were then asked to select which graph best
depicted the manner by which they judged the materiality of an
expected loss in relationship to the percent of income effect. (The
subjects were permitted to specify a functional form other than the
six provided by the author.)

Ward had hypothesized that if differences were observed, it would
be attributable to differences in perception as to the size of the
error in relationship to the "relative severity of the error" (i.e.,
percent of net income). However, after partitioning the observations
according to perceived functional form, he suggested the results were
inconclusive. Hence, no position was taken with respect to this
hypothesis.
Newton [1977] employed cardinal utility analysis in an exploration of the effect of uncertainty upon CPAs materiality judgments. The purpose was to identify and explain the effect of risk attitudes on materiality decisions. Newton hypothesized (p. 98) that "... the decision an accountant reaches concerning the materiality of an item is related directly to the degree of uncertainty about the issue and the accountant's acceptable [Beta] risk" (i.e., issuing misleading financial statements). As risk aversion increases, materiality thresholds are predicted to become smaller.

Cardinal utility theory was used to assess the risk attitudes of the CPA participants. The subjects were nineteen partners of CPA firms encompassing "Big Eight," regional, and local firms. The subjects were required to express preferences as to sets of "gambles" with specified outcomes and designated probabilities. Based upon a subject's preferences, utilities were inferred.

Three unrelated cases were used. The first was a general case to assess each subject's basic attitudes toward risk outside of an accounting context (called the "standard gamble").

The second case involved marketable securities. Given that a client company owned a portfolio of marketable securities costing $25 million, the auditors were asked to specify the amount of error in the valuation of the securities (relative to the company's $50 million net income) which would cause the auditor to qualify the audit report. Probabilities were introduced by asking the subjects to specify the
minimum probabilities necessary to cause qualification if the portfolio's new valuation were (a) $5 million; (b) $10 million; and (c) $17.5 million.

The third case involved a contingent liability associated with the outcome of litigation. Given that a client firm was expected to lose a lawsuit, but refused to disclose the circumstances in its financial statements, the auditors were asked to specify the minimum amount which would cause the audit report to be qualified (relative to the client's $100 million net income). Then probabilities were introduced by asking the subjects to specify the minimum probabilities necessary to force qualification if the loss were limited to (a) $100,000; (b) $1 million; and (c) $3 million.

The author concluded that the subjects tended to be consistent in their attitudes toward risk in the context of the experiment. Only four subjects were said to have changed attitudes toward risk (from averse to seeking or vice versa) when the context of the experiment changed. In addition, Newton concluded that the subjects appeared to incorporate probabilities into their judgments, and added (p. 104), "This is not to say that such probabilistic considerations were explicit in the behavior of the accountants, though it was apparent that the CPAs thought some consideration should be given to the chances of the event's occurrence." Since the majority of subjects were classified as risk averse, Newton argued that the development of
materiality guidelines should include explicit consideration of the risk factor.

Hofstedt and Hughes [1977] used a research approach known as the "clinical judgment" paradigm to investigate the relationship of materiality and disclosure decisions. Two criteria were used to evaluate a subject's performance. One, "convergence," concerned whether a given decision maker tended to make consistent judgments for different sets of data. The other, "self-insight," dealt with how well the decision maker was able to describe his or her decision processes and the extent to which relevant cues were used in accordance with the described processes. These two criteria were used to evaluate the information-processing characteristics associated with the judgment task.

The subjects were nineteen second-year MBA students, concentrating in accounting, at Cornell University. The task was to decide whether a loss due to the writedown of a subsidiary should be classified as extraordinary based upon its relationship to the experiment's variables.

The experimental design was a $3^3 \times N$. There were three independent variables—the loss was expressed as a percent of operating income, as a percent of the parent company's total investment in all its subsidiaries, and as a percent of the net book value of the subsidiary under consideration. Each independent variable had three levels representing "low," "medium," and "high." In addition, subjects were
asked their "prior probability of disclosing" before the experiment was begun. The dependent variable was the participant's "subjective probability of disclosure" elicited by asking the subject to specify a probability for disclosing the writedown as extraordinary for each combination of the independent variables. It was assumed that the subjective probability of disclosure was an appropriate substitute for materiality. (Their idea was to have a variable that could be measured on a 100-point scale, instead of simply "material" versus "immaterial.")

The experimental results in the aggregate for all subjects were evaluated by analysis of variance (ANOVA), and suggested that all main effects and one interaction (operating income x parent investments) were highly significant. A regression equation was calculated that had an adjusted $R^2 = .39$.

In addition, the results were analyzed individually by subject. Three linear regression models were run for each subject—one including all main effects and all two-way interactions, another including all main effects and the one most important two-way interaction, and a third comprised of only main effects. The individual analysis yielded three conclusions. "First, the overall explanatory power increases sharply when individuals rather than aggregates are used. Second, relatively simple models do about as well as more complex models. Third, there are substantial differences among subjects on several dimensions." It appeared that the
complexity of information processing differed widely. Some subjects had as many as five significant factors (main effects plus interactions) whereas others had as few as one. Further, the cue weights appeared to differ.\textsuperscript{7}

Morarity and Barron \textsuperscript{[1979]} followed up their earlier article. Their approach was a modified version of their earlier approach; the authors expressed their view (p. 114) that "... the primary contribution of the paper is ... an illustration of a methodology having significant implications for accounting research." However, because the experiment lacked "mundane reality," they conceded that the results of the study might not apply to materiality judgments in practice.

The approach used, called "conjoint analysis," had the following characteristics: assumed additivity, a fractional factorial design, and part-worth functions. The model assumed to apply was

\[ M = f_1(x_{j1}) + f_2(x_{j2}) + f_3(x_{j3}) + f_4(x_{j4}) + f_5(x_{j5}) \] (2.1)

where:

- $M$ = the amount judged material
- $f_i(x_{ji})$ = the effect of variable $i$ at level $j$
- and $f_1$, $f_2$, ..., $f_5$ were the part-worth functions applicable to the five variables. It was assumed that five factors affecting materiality could be found. The variables chosen were (1) net income, (2) debt-equity ratios, (3) number of shares, (4) total assets, and (5) earnings trend. Total assets was the most important variable for
one partner; net income was the dominant variable for the other subjects.

The subjects were eight audit partners from a national CPA firm. The materiality judgments were couched in a planning context, and participants were asked to quantify the appropriate "pre-audit" materiality thresholds for each of 30 cases. However, Moriarity and Barron reported that the task was unfamiliar to their eight subjects (p. 130):

Only two of the participants indicated that they had thought seriously about the specific problem of specifying preaudit materiality... Four of the other participants specifically mentioned in our discussion that they were uncomfortable making preaudit materiality decisions. One even refused to complete the task, because he believed the question was not an appropriate one to ask. Another participant said he did not think auditors would ever be required to make these judgments.

The analysis of results indicated that a consensus across subjects as to the part-worth functions and the predictions was not achieved. However, the authors pointed out (p. 122) that "...seven of the eight participating partners produced interpretable part-worths, providing good predictions within the qualitative and quantitative differences across auditors."

Moriarity and Barron attributed this failure to achieve consensus across subjects to the diverse specialties and experience differences among the eight audit partners. They speculated (p. 131) that "...had we had participants with more homogeneous experience, we would likely have achieved more homogeneous results."
Firth [1979] compared the disclosure judgments of auditors, managers, and user groups (i.e., securities analysts and bankers) in the United Kingdom. Again, disclosure was used as a surrogate for materiality. The task required subjects to specify whether the extraordinary item warranted separate disclosure for each of 30 cases (half dealt with losses and half dealt with gains of different amounts). A total of 150 subjects participated in the study--30 from each of three Big Eight CPA firms, 30 from industry, and 30 financial statement users. Firth manipulated the gains and losses as a percentage of the following variables: income before extraordinary items, net assets, total assets, market capitalization, sales turnover, and current assets.

The gain or loss distinction was not significant, but the items as a percentage of income before extraordinary items, net assets, total assets, and market capitalization were significant. The variable with the greatest explanatory power was the item as a percentage of income before extraordinary items. The results indicated that the users had lower materiality thresholds than the other groups (which were not significantly different). Even though the judgments of the three CPA firms involved were not significantly different, Firth concluded (p. 291) that "... some consideration should be given to the idea of producing guidelines and yardsticks for materiality-disclosure decisions."
Schultz and Reckers [1981] used ANOVA to model the disclosure judgments of 64 Peat Marwick audit partners with respect to two contingencies. The dependent variable was the minimum probability for which each subject believed footnote disclosure to be required. (Implicitly, "audit risk" was the dependent variable, since audit risk presumably increases as the probability of disclosure decreases.) In this study, materiality was used as an independent variable with two levels—the "low" condition (i.e., about six percent of net income) and the "high" condition (i.e., about fourteen percent of net income). Other independent variables were the nature of the CPA's role ("advisory" versus "decisive"), and the communication channel between group members (telephone versus face-to-face). Measurements were taken at three points for each subject—(1) prior to being subjected to the group's influence; (2) in conjunction with the group's conclusion (required to be unanimous); and (3) subsequent to being subjected to the group's influence. (A group consisted of four subjects.)

The main effects for materiality and authoritative capacity were significant at .01, and the communication mode was significant at .07. The high level of materiality led to a mean probability of disclosure of 40 percent, versus 46 percent for the low level. The authors reported that many of the subjects appeared to use an "expected value" of the loss as a measure of materiality, in contrast to a "sequential process" implied by FASB Statement No. 5.
The "choice shift" factor (i.e., the comparison of the pre-group, group, and post-group measurements) was not significant. Although the group process did not appear to significantly shift the mean responses, the group setting did appear to reduce the variation in the responses and provide the subjects with a higher level of confidence in their judgments.

Emery, Thakkar, and Moriarity [1982] used conjoint measurement (CJM) to compare the materiality judgments of auditors and financial statement users. According to the authors (p. 5), their paper was intended to answer three research questions:

First, do the materiality judgments of users of accounting information differ from those of auditors? Second, among users, do the materiality judgments of users with an equity point of view (financial analysts) differ from those of users with a debt point of view (commercial bank loan officers)? Third, within each professional group (i.e., auditors, financial analysts, and loan officers), how consistent are the materiality judgments?

Similar to Moriarity and Barron (M&B) in their 1976 paper, the authors applied axiomatic CJM to identify subjects whose judgments could be modeled as additive. Then numerical CJM was applied to determine the coefficient of the applicable cues in the linear models associated with these subjects. The experiment was virtually a replication of M&B's, except for the use of different subject groups.

Again, eighteen sets of financial statements were prepared representing a full factorial experimental design where the three independent variables (income level, earnings trend, and asset size) had three, two, and three different levels, respectively. The
securities analysts were told that the companies represented by the financial statements were under consideration for equity investments, whereas the commercial loan officers were told that the companies were applying for loans.

The task consisted of ranking the financial statements in order of materiality given a situation which implied a possible adjustment (decrease) to income before taxes of $500,000. This situation related to the estimated useful life of specialized equipment. The equipment was being depreciated based upon a ten-year useful life when the best information available suggested that a five-year useful life was more likely.

Examining the individual judgments, it appeared that the results were quite similar to M&B. Of the fifteen auditors in their study, M&B reported that only four could not be modeled as additive. Here, of the sixteen securities analysts, four were nonadditive and, of the fourteen commercial lenders, four were also nonadditive.

In addition, the average group rankings of the cases were nearly the same for the auditors in M&B's study and the financial analysts and loan officers of Emery, Thakkar, and Morarity's study. Consequently, the authors concluded that there were no significant differences in the structural models for materiality judgments among auditors, financial analysts, and loan officers. Further, it appeared that the cues were weighted nearly equally, on average, among the three groups.
However they cautioned (p. 7) that their results should be interpreted carefully:

That is, this study provides evidence that auditors and users agree that, in an ordinal sense, a specific item is more material for a firm with $50,000 in income than for a firm with $75,000 in income (all else equal). But the study does not represent evidence about whether the groups agree, in a nominal sense, that the item is or is not material in either case.

They argue that studies which observe differences among various subjects in materiality judgments tend to be studies which focus on the determination of materiality thresholds (e.g., Woolsey [1973b] and Pattillo [1975]).

Messier [1983] examined the joint materiality and disclosure decision for 29 audit partners who were classified as to their levels of experience and their firm type. Experience was differentiated as greater than 20 years of experience, sixteen-20 years, or up to fifteen years. The number of partners in these categories were six, eight, and fifteen, respectively. "Firm type" simply dichotomized "Big Eight" versus "non-Big Eight." The number of partners in these categories were 22 and seven, respectively.

The experiment involved a multivariate lens model framework and required each subject to make two judgments regarding a writedown of inventory in the amount of $1 million: (1) rate the materiality of the writedown on a seven-point scale; and (2) identify the probability that the item should be separately disclosed in the income statement in accordance with APB Opinion No. 30 (on a 100-point scale).
Each subject was asked to consider 32 separate cases based upon the $4 \times 2^4 \times \frac{1}{2}$ fractional factorial design. This reflected five financial variables: (1) net income, (2) earnings trend, (3) total assets, (4) total inventories, and (5) current ratio. Net income had four levels and the other variables each had two levels.

The overall results indicated that net income was significant for 27 partners, earnings trend was significant for thirteen partners, and the other three variables were insignificant for most of the partners. Net income and earnings trend accounted for 78 percent of the variation in the subjects' decisions. Messier measured judgment consensus, self-insight, and stability of the subjects' judgments over time, and concluded that such measures were relatively high and consistent with a number of prior materiality studies.

A fixed effects ANOVA model was determined for each subject for both the materiality and the disclosure decisions. The experience factor tended to increase the proportion of variation in the data that could be explained. Messier concluded (p. 615), "This suggests that auditors with more experience had a better understanding of the decision process and placed greater reliance on the net income variable."

The analysis of the firm type factor suggested that the net income variable explained a greater proportion of the variation in the data for the Big Eight partners than those of non-Big Eight firms. However, Messier stated (p. 615), "I have no explanation for this."
Although it is difficult to make generalizations due to sample size limitations, this latter result is consistent with the hypothesis that, on average, a year of Big Eight experience is more concentrated than a year of non-Big Eight experience. It may be that "experience" should reflect more than simply a chronological dimension.

Using MANOVA, the author reported that both main effects (i.e., experience and firm type) and the experience–firm type interaction were significant at .01. A separate ANOVA was calculated for both the materiality and disclosure decisions. For the materiality judgment, only firm type was significant at .01. For the disclosure judgment, experience, firm type and the experience–firm type interaction were significant.8

Concerning consensus, Messier reported that the less-than-fifteen-years-of-experience group had less agreement within the group than did the other two groups. Also, the non-Big Eight group had less agreement than the Big Eight group. As to self-insight, the author concluded that experience appeared to have no effect on self-insight and firm type may have had only minor effect. Stability did not appear to be influenced by either experience or firm type.

Messier summarized that experience and firm type had an effect on the consensus of the subjects' judgments (based on MANOVA), but no effect on self-insight or stability. This was interpreted as being attributable to the complexity of the task. Consequently, Messier suggested (p. 618) that "... researchers interested in studying
materiality and disclosure judgments further should address the effects of complexity and frequency of occurrence of the task."

Krogstad, Ettenson, and Shanteau [1984] investigated the effect of certain nonfinancial variables and experience upon auditors' materiality judgments. Carmichael [1972, p. 54] stated, "The major overriding conclusion drawn from the study of the decision-making criteria used by auditors for audit reports is that materiality--as it relates to the type-of-report decision--is a matter primarily of qualitative importance" (emphasis added). According to Krogstad, et al., previous research may have suffered from "artificiality" caused by omitting certain nonfinancial or contextual factors from the experimental designs. Consequently, they included five nonfinancial (contextual) cues in their study. These were (pp. 57 and 59):

1. a diversified firm in an expanding industry vs. a nondiversified firm in a declining industry, 2. a closely held private corporation with creditors being the primary users of the financial statements vs. a publicly held corporation with stockholders being the primary users of the financial statements, 3. a management which is completely cooperative and open in all dealings with your (the audit) firm vs. a management which is less than completely cooperative and open in all dealings with you (the audit) firm, 4. a management that follows conservative accounting policies vs. a management that follows unconservative accounting policies, 5. the presence of strong internal control procedures vs. the presence of inadequate internal control procedures.
In addition, they included three financial (quantitative) variables:

(6) the effect of the proposed adjustment on the current ratio (a reduction of 2.5 to 2.1 vs. a reduction of 2.1 to 1.7), (7) the effect of the proposed adjustment on net income (a decrease of 2.7 percent vs. a decrease of 7.3 percent), and (8) the effect of the proposed adjustment on the earnings-per-share (EPS) trend (no reversal of an upward EPS trend vs. reversal of an otherwise upward EPS trend).

The study considered professional experience by distinguishing the subjects as audit partners ("very experienced"), audit seniors ("moderately experienced"), and accounting students ("inexperienced"). There were eleven partners, ten seniors (the auditors were from five of the Big Eight firms), and eleven students. The authors hoped that including accounting students would be helpful in evaluating the extent to which students may serve as surrogates for professionals in research situations.

The authors prepared 32 hypothetical cases (sixteen separate cases were each repeated). The experimental design was a $3 \times 2^8 \times \frac{1}{16}$ fractional factorial where the main effects were: (1) experience (three levels—a between-subjects variable) and (2) eight cues (two levels each—all within-subjects variables).

The experimental task required each subject to assume the role of the in-charge accountant on a hypothetical audit engagement. A judgment was needed as to the materiality of a proposed adjusting journal entry affecting the "Allowance for Doubtful Accounts." Subjects were asked to identify the materiality of the adjustment on a
scale from "1" ("clearly immaterial") to "100" ("clearly material"). Each case was to be treated independently.

The overall results, based on omega squares, suggested that "effect on net income" was the dominant variable for partners and seniors. However, no single cue was dominant for the students. On average, partners used 3.3 cues, seniors used 2.7 cues, and students used 2.5 cues. Of the cues determined to have statistical significance, the contextual cues accounted for 48, 40, and 61 percent of the cues used by partners, seniors, and students, respectively.

Consensus was evaluated by intra-group correlation coefficients. The partners and seniors appeared no different, whereas the students achieved a lower level of consensus. Judgment consistency was compared by calculating intra-subject correlations on the replications of each case. Again, the results suggested that there were no differences in consistency between partners and seniors, but that there was a difference between the auditors and the students.

The results of this study regarding the financial cues were comparable to earlier research. The "effect on income" was easily the most important financial variable, followed by the "effect on earnings trend" which was, in turn, followed by the "effect on current ratio." The following identifies the average percentage variation explained by each variable:
<table>
<thead>
<tr>
<th>Variable</th>
<th>Partners</th>
<th>Seniors</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Industry trend</td>
<td>.01</td>
<td>.03</td>
<td>.01</td>
</tr>
<tr>
<td>2. Primary users</td>
<td>.03</td>
<td>.01</td>
<td>.03</td>
</tr>
<tr>
<td>3. Management cooperativeness</td>
<td>.01</td>
<td>.01</td>
<td>.06</td>
</tr>
<tr>
<td>4. Accounting policies</td>
<td>.03</td>
<td>.03</td>
<td>.11</td>
</tr>
<tr>
<td>5. Internal control</td>
<td>.02</td>
<td>.02</td>
<td>.12</td>
</tr>
<tr>
<td>6. Effect on current ratio</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>7. Effect on net income</td>
<td>.62</td>
<td>.56</td>
<td>.18</td>
</tr>
<tr>
<td>8. Effect on EPS trend</td>
<td>.07</td>
<td>.05</td>
<td>.06</td>
</tr>
<tr>
<td>Total</td>
<td>.86</td>
<td>.72</td>
<td>.59</td>
</tr>
</tbody>
</table>

Of the 21 auditors included as subjects, fourteen used one or more of the contextual (nonfinancial) cues. The partners, on average, used 1.6 contextual cues versus 1.7 financial cues. Seniors, on average, used 1.1 contextual cues versus 1.6 financial cues. Of the eleven students in the study, ten used at least one contextual cue and, on average, students used 1.5 contextual cues versus 1.0 financial cue.

The variable "primary users" related to the distinction between public corporations (whose primary users were assumed to be stockholders) versus private companies (whose primary users were assumed to be creditors). It was interesting to note that the variable was significant for only four of the ten partners. Of those partners who used it, six to seven percent of the variance was explained solely by this variable. However, this variable was significant for only two of the eleven seniors, and explained four percent of the variance for one senior and one percent for the other. This variable was significant for only one of the eleven students, but explained 29 percent of the variance for that subject. It is not
clear how this variable affects materiality judgments, nor why the subjects differ in the application of this cue.

Focusing on experience, it appeared that there were no significant differences between partners and seniors as to consistency, consensus, and the number of cues used (along with their applicable weights), although the partners appeared to be slightly higher than the seniors on each of these three dimensions). On the other hand, students differed from the auditors in each of these respects. Accordingly, the authors concluded (p. 70), "These results challenge the validity of using students as surrogates for professionals in judgment research."

Contrasting these results with some earlier research which suggested that students could serve reasonably well as surrogates for practitioners, the authors hypothesized that the different conclusions relate to differences in the complexity of the experimental task. That is, when the experimental task is complex or the judgment environment not well defined, it seems likely that experienced subjects may demonstrate skills not yet attained by inexperienced subjects. The authors concluded (p. 71), "... future judgment researchers are cautioned to control carefully for background and professional experience in view of the specific task characteristics being investigated."
Other Relevant Research

The Lens Model

Egon Brunswik developed a framework for representing human judgment known as the "lens" model. The lens model has been summarized by a number of authors, including Ashton [1982b] and Dudycha and Naylor [1966]. Their descriptions are relied on here.

As suggested by Figure 1, the lens model partitions the decision or judgment situation into two parts. The first part (the left-hand side of the model) represents the environment. $Y_e$ denotes the criterion or "distal" variable and represents the actual outcome. $X_i$ denotes cue $i$ representing an item of information which can be used in some respect to predict the criterion variable. Brunswik viewed the cues as being linked probabilistically to the uncertain criterion event. Hence, the decision maker has to base his or her judgment on a subset of the total information set available to the environment. The second part of the model (the right-hand side) represents the individual's judgment system. $Y_s$ denotes the individual's judgment or prediction. The cue validity coefficients are denoted by $r_{ei}$ and represent the relationship between $Y_e$ and each $X_i$. The cue utilization coefficients are denoted by $r_{si}$ and represent the extent to which the individual utilizes each $X_i$ to formulate $Y_s$. 
Source: Dudycha and Naylor [1966, p. 111]

Figure 1

The Lens Model and the Relationships Among the Cues, the Criterion, and Subjects' Responses
Using linear multiple regression, the left-hand side of the lens model can be represented by the following:

\[ \hat{Y}_e = b_{e1}X_1 + b_{e2}X_2 + \ldots + b_{en}X_n \]  

(2.2)

where \( Y_e \) is the predicted criterion variable given the \( X_1, \ldots, X_n \) cues and each \( b_{ei} \) coefficient represents the optimal weighting of cue \( i \). The correlation coefficient, \( R_e \), represents the linear predictability of the criterion event based on the cue set.

The right-hand side of the model can be represented in a similar manner:

\[ \hat{Y}_s = b_{s1}X_1 + b_{s2}X_2 + \ldots + b_{sn}X_n \]  

(2.3)

where \( Y_s \) represents the prediction of the subject's response based on cues \( X_1, \ldots, X_n \) and \( b_{si} \) represents the extensiveness of the usage by the individual on cue \( X_i \). The correlation coefficient, \( R_s \) (called the "consistency index" by Dudycha and Naylor [1966]), represents the linear relationship between the individual's judgments and the given cue set.

As discussed by Ashton [1982b, p. 17] (see Figure 2), there are six correlation coefficients which may be calculated as indices of human judgment. Two important ones, \( R_e \) and \( R_s \), were described above. The other four are denoted as follows: \( r_a, r'_a, r_0, \) and \( G \). The so-called "achievement index" is represented by \( r_a \) and measures the individual's achievement in predicting the criterion variable. Related somewhat to that is \( r'_a \), which reflects the predictability of the criterion variable based on the individual's regression model.
Source: Ashton [1982b, p. 17]

Figure 2
Indices Related to Human Judgments
The "optimal index" is represented by $r_0$ and measures the correspondence between the individual's judgments and those of the optimal regression model's prediction of the criterion variable. The "matching index" is represented by $G$ and reflects the correspondence between the individual's regression model and the optimal regression model of the environment.

The above discussion is presented from a linear multiple regression point of view. Analysis of variance (ANOVA) is also applicable to a lens model formulation. ANOVA is particularly useful in incorporating nonlinear or configural cue usage. "Main effects" represent the linear components of the regression model and "interactions" represent the configural components. A main effect merely indicates that the subject responds systematically to a particular cue while all other cues are held constant. A two-way interaction suggests that the way the individual responds to a given cue is conditional on the level of another cue. Three-way (and higher-order) interactions are more difficult to interpret (Keppel [1982]).

In the context of this study, no criterion variable exists. In many auditing contexts, the environment does not provide a "correct" outcome. (In rare instances, the courts may be viewed as providing such a determination on a case-by-case basis.) However, in most auditing situations there will not be a criterion variable for materiality. Hence, this study focuses on the right-hand side of the
lens model. As a result, this research emphasizes consensus rather than "accuracy."

Ashton [1982a, p.74] notes that audit decision making has been evaluated with respect to six criteria: (1) accuracy; (2) normativeness; (3) consistency with professional pronouncements; (4) consensus; (5) self-insight; and (6) judgment stability. The first three criteria are not applicable to this study. No external criterion exists with respect to auditors' materiality judgments and, hence, optimality can not be assessed. Due to the generality of the auditing standards (i.e., SAS No. 47), it is not possible to assess inconsistencies with the pronouncements in an empirical study such as this. (See Cushing and Loebbecke [1983] for an evaluation of twelve national firms' policies relative to the SASs.)

Therefore, consensus, self-insight, and stability are available for consideration in this study. Ashton [1982, p. 75] notes that these three criteria are usually used when accuracy and normativeness can not be used, and comments:

A rationale for both stability and consensus is that the cost and/or quality of an audit may fluctuate needlessly if decision making is inconsistent over time or across auditors. The rationale for insight involves the importance of an auditor's understanding his or her own decision making if he or she attempts to train other auditors in decision making.

In view of prior research which has reported relatively low consensus on materiality tasks, but relatively high self-insight and stability (e.g., Messier [1983]; and Krogstad, et al. [1984]), this study
focuses on consensus as a criterion. Since this study emphasizes comparisons among groups of auditors, rather than examining the judgments within individual auditors, consensus is of primary importance.

Ashton [1982b, pp. 124-152] reviewed a large number of studies in accounting which have adopted the lens model framework. The contexts included internal control evaluation, materiality judgments, audit report messages, bankruptcy prediction, stock recommendations and price predictions, functional fixation, and managerial decision making, among others. The materiality studies which Ashton described as lens studies (many of the judgment-capture experiments regarding auditors' materiality judgments discussed in detail above) were: Boatsman and Robertson [1974]; Moriarity and Barron [1976, 1979]; Hofstedt and Hughes [1977]; Firth [1979]; and Messier [1979]. In addition, although not included in Ashton's review of materiality-related lens studies, Schultz and Reckers [1981] and Krogstad, et al. [1984] would be appropriately viewed as applications of the lens model. Accordingly, there is ample empirical support for the lens model as a framework within which to explore individuals' judgments, in general, and auditors' materiality judgments, in particular. An advantage of the lens model is its applicability to diverse contexts.
The Role of Attitudes

As indicated above, prior studies of auditors' materiality judgments suggest that auditors differ in their judgments. However, little attention has been paid to why such differences arise. A secondary purpose of this research is to explore whether there are attitudinal differences among auditors associated with systematic differences in their materiality judgments. One might anticipate that auditors have different priorities and different concerns which may affect their views of the appropriate materiality threshold. In other words, there may be certain attitudinal differences among auditors who have "high" materiality thresholds relative to auditors who have "low" materiality thresholds. The concept of "risk attitude" used in this study deals with that notion.

Palmerino, et al. [1984, p. 179], observed, "The study of attitudes has had a long and honored tradition among social psychologists, even to the extent of being defined as the core of social psychology itself." Allport [1968, p. 59] referred to the concept of attitude as "... the most distinctive and indispensable concept in contemporary American social psychology..." Yet, Fishbein and Ajzen [1975] point out that there is considerable disagreement about the appropriate definition of "attitude." Palmerino, et al. [1984, p. 179], characterized an "attitude" as follows: "First, an attitude develops through experience with an object. Second, it predisposes one to act in a predictable manner with respect to an
object. Third, an attitude consists of positive or negative evaluations." (The same three characteristics were identified by Fishbein and Ajzen [1975, p. 6] in their definition of attitude.)

In addition, some issues still remain with respect to the measurement of attitudes. In their 1972 review article, Fishbein and Ajzen reported on over 750 articles published in 1968-1970 related to attitudes. They noted that several hundred different techniques or variations were used to measure attitudes in those articles.

Further, there is some uncertainty about the link between attitudes and behavior. Commenting on a renewed interest in research on attitude-behavior relationships, Ajzen and Fishbein [1977, p. 888] reported, "The emerging position seems to be that attitude is only one of many factors determining behavior." Furthermore, in promoting the "Theory of Mindfulness-Mindlessness," Palmerino, et al. [1984, p. 186], cautioned, "Attitudes, and contexts in general, have greater influence upon behavior when a person is mindless.... Laboratories may induce self-consciousness in our subjects and as such may be thought-provoking and habit-inhibiting." Consequently, attitudes which might be relevant to behavior at a "subconscious" level may be made irrelevant by raising the issues to a "conscious" level. Although attitudinal factors may only explain a portion of the observed behavioral differences, the investigation of differences in subjects' attitudes may, nevertheless, offer insights into differences in their behavior of interest.
Fishbein and Ajzen [1975] presented a "conceptual framework" within which empirical research could be organized. Their framework consisted of four primary constructs: (1) belief; (2) attitude; (3) intention; and (4) behavior. They commented (p. 29) on their theory of attitudes:

(1) An individual holds many beliefs about a given object; i.e., the object may be seen as related to various attributes, such as other objects, characteristics, goals, etc. (2) Associated with each of the attributes is an implicit evaluative response, i.e., an attitude. (3) Through conditioning, the evaluative responses are associated with the attitude object. (4) The conditioned evaluative responses summate, and thus (5) on future occasions the attitude object will elicit this summated evaluative response, i.e., the overall attitude.

They summarized these relationships as follows:

$$A_0 = \sum_{i=1}^{n} b_i e_i$$

(2.4)

where $A_0$ represents the attitude toward some object, $0$; $b_i$ represents belief i about 0 (i.e., the subjective probability that attribute i is related to object 0); $e_i$ represents the evaluation of attribute i; and $n$ represents the number of beliefs.

Since the present study deals with auditors' planning stage materiality judgments, some connection must be made between "attitude" and "behavior." Fishbein and Ajzen [1975, p. 301] report on Fishbein's earlier work in representing the attitude-behavior relationship. According to Fishbein's theory, there are two factors which drive behavioral intentions. The first is attitude (an individual factor) and the second is a social or normative factor.
These two factors are then weighted; the applicable weights may differ across individuals. The theory can be represented as follows:

\[ B^I = (A_B)w_1 + (SN)w_2 \]  \hspace{1cm} (2.5)

where \( B \) represents the behavior of interest; \( I \) represents the intention to perform the behavior \( B \); \( A_B \) represents the attitude toward performing behavior \( B \); \( SN \) represents the subjective norm associated with behavior \( B \); and \( w_1 \) and \( w_2 \) represent the individual's weights associated with the attitude and norm, respectively.

\( A_B \) was represented (2.4) as the summation of the product of the applicable beliefs and evaluations. The subjective norm, \( SN \), deals with the influence of the social environment on behavior, and can be represented as follows:

\[ SN = \sum_{i=1}^{n} b_i m_i \]  \hspace{1cm} (2.6)

where \( b_i \) represents the normative belief that the behavior \( B \) should be performed in view of the reference group or individual \( i \); \( m_i \) represents the motivation to comply with the reference group or individual \( i \); and \( n \) represents the number of referents.

To the extent that the behavior and the intention to perform the behavior may be assumed to be highly related, then factors which influence the intention to perform the behavior will influence the behavior, too. Fishbein and Ajzen [1975, p. 369] consider three factors which may influence the relationship between intention and behavior: (1) the degree to which intention and behavior correspond in their levels of specificity; (2) the stability of intention; and
(3) the individual's volitional control over performing the intended behavior.

Fishbein and Ajzen suggest that the first factor may be the most important of the three. They emphasize that the relationship between the intention and behavior will be strongest when the intention and behavior are examined at the same level of specificity. The stability of intention deals with the possibility that a person's intention may change over time. To the extent that the intention is stable, the predicted behavior is more likely to be observed over time. Finally, the notion of volitional control acknowledges that a person may have an intention to perform a particular behavior (e.g., quit smoking), but be unable to implement that intention.

Fishbein and Ajzen [1975] included a review of a large body of literature related to psychology studies of attitudes, and a number of observations follow. They cited several studies which provide good support for the expectancy-valence approach to model attitudes. In addition, they discussed over a dozen studies which provide good support for the relationship of attitude toward the behavior (A<sub>b</sub>) and the subjective norm (SN). They referenced several other studies which provide support for the assertion that intention and behavior are highly correlated when examined at the same level of specificity. Ajzen and Fishbein [1977] reviewed a large number of studies and, similarly, concluded that attitudes are highly correlated with behavior when they are examined at the same level of specificity.
Thus, the psychology literature provides substantial empirical support for the constructs of the Fishbein attitude model. Recently, van der Plight and Eiser [1984, p. 165] commented on the Fishbein and Ajzen [1975] attitude model: "Since the introduction of this model research has shown a renewed interest in attitudinal phenomena...The theory of reasoned action has been applied to a wide variety of behaviors and a considerable amount of supporting evidence has accumulated in both experimental and naturalistic settings...."

In addition to the psychology literature, empirical support for the relevance of attitudinal differences in explaining behavioral differences can be found in the accounting literature. Kida [1980] is relevant to the issue of risk perception by auditors, and the consequent influence on their predispositions toward the alternative opinion choices. Kida studied auditors' decisions to issue going concern opinions to hypothetical clients relative to their classifications of the companies as troubled versus untroubled firms. He measured their attitudes toward issuing a qualified opinion to a financially troubled firm (a correct judgment) and toward issuing a qualified opinion to a financially untroubled firm (a Type I error); similarly, he measured their attitudes toward issuing an unqualified opinion to a troubled firm (a Type II error) and toward issuing an unqualified opinion to an untroubled firm (a correct judgment). Attitudes were measured first by semantic differential scales (which used bi-polar adjectives), and second by expectancy-valence (which
combined both likelihood and evaluative measures). Kida concluded that the auditors' attitudes toward issuing going-concern qualifications (i.e., their perceptions of consequences) affected their decisions to issue such qualifications in the experiment.

Kida used 27 partners from national CPA firms as subjects. He selected 20 actual "problem" firms and 20 actual "non-problem" firms. Using five ratios for each firm, the subjects were asked first to identify whether each firm was a problem or a non-problem firm (on a six-point scale), and second to identify the type of opinion which the auditor would render (unqualified, going concern qualification, or disclaimer). Each auditor's tendency to issue a going concern opinion was measured by the number of times the going concern qualification was indicated relative to the number of times a problem firm was specified. This tendency was then compared to the auditor's attitudes toward issuing a going concern qualification as measured by the semantic differential technique. The resulting correlation was .419 which was significant at about .016. When attitudes were measured by expectancy-valence, the correlation was .301 which was significant at about .068. The author attributed the weaker association to the nonsalience of some of the belief statements under the latter approach.

This research will focus on the attitude construct denoted $A_o$ (2.4). It represents the "individual factor" (as distinguished from the social or normative factor denoted $SN$ above) which has been linked
to behavior. In this study, behavior will consist of the reported materiality judgments of the participating auditors. As previously indicated, the attitude measure is a summation of the products of certain underlying beliefs ($b_i$'s) and evaluations ($e_i$'s). It is expected that auditors may identify different materiality thresholds because of different perceptions of the consequences (either differences in the perceived likelihoods of particular outcomes or in the perceived importance of those outcomes, or both) associated with auditors' materiality judgments. The expectancy-valence measure of attitude captures both of those dimensions by measuring $b_i$ and $e_i$ for each identified outcome, $i$.

Within the context of this study, the notion of "intention" would not appear to be directly relevant to auditors' materiality judgments. Perhaps one could incorporate the intention of an auditor to establish a "conservative" or a "liberal" materiality threshold. On a somewhat different level, one might consider the intention of an auditor to issue an unqualified opinion (versus a qualified opinion); this intention may be manifested by establishing higher (lower) materiality thresholds. However, it is believed that including intention in the study would not add any significant insights relative to focusing on behavior directly. In this study, the specified materiality thresholds are within the volitional control of each participating auditor. As a result, the participants should be able to
operationalize their intentions and there is little reason to expect a disparity between intention and behavior.

Other Accounting Research

Crosby, Plante, and Tang [1985] developed an analytical model of materiality which, they argue, is optimal in an economic sense. They define the auditor's optimal materiality threshold as that level which minimizes the total expected cost of the opinion choice. In simplest form, they consider the total expected cost of the audit decision to consist of the expected cost of "acceptance" (i.e., issuing an unqualified opinion) plus the expected cost of "rejection" (i.e., issuing a qualified opinion). By explicitly considering that individuals may differ in their perceptions of the likelihoods and/or their evaluations of the consequences associated with "acceptance" and "rejection" (i.e., risk attitude), this dissertation adopts a similar view of materiality.

Cushing and Loebbecke [1983] and Kinney [1986] relate to the issue of differences among national CPA firms. The former developed a model of generally accepted auditing standards (GAAS) based on a detailed review of the SAS's. After developing individual firm models for the twelve participating national firms based on a review of the firms' audit manuals, they compared the individual firm models across firms and to the GAAS model. They concluded (p. 68) that there was greater nonconformity in the preliminary estimate of materiality than
in all the other steps of their GAAS model. Although two firms had developed specific analytical approaches for quantifying the preliminary estimate of materiality and incorporating it into audit planning, they noted that five of the nation's twelve largest CPA firms did not have formal policies regarding preliminary estimates of materiality.

Kinney [1986] classified the Big Eight and several non-Big Eight firms as "structured," "intermediate," or "unstructured" based on their audit technologies. According to Kinney (p. 10), "Structured firms are more likely to argue that auditor consensus is desirable, that is, in a given situation different auditors should conduct similar audits and give the same opinion." Unstructured firms tend to leave more matters to the judgment of the individual auditors and to the practice offices. Morris and Nichols [1986, p.6] reported the specific measures of structure used by Kinney. Five individuals were asked to rate the degree of audit structure for twelve large CPA firms (including each of the Big Eight). Four of these individuals were members of the Auditing Standards Board who rated each of the firms; the other was someone within each firm who rated his or her own firm. The rating was a three-point scale (where "1" represented the least structured and "3" represented the most structured). Thus, the maximum possible score (summed across the five individuals) was fifteen and the minimum was five. The following scores were identified for the Big Eight firms, from most to least structured:
Deloitte, Haskins, & Sells 15 structured
Peat, Marwick, Mitchell 15 structured
Touche Ross 13 structured
Arthur Andersen 10 intermediate
Ernst & Whinney 10 intermediate
Arthur Young 10 intermediate
Coopers & Lybrand 5 unstructured
Price Waterhouse 5 unstructured

Summary

The literature described above was presented in three parts. The relevant pronouncements by the U.S. standard-setting bodies were discussed first, along with several pronouncements by professional bodies outside of the U.S., since the issues related to auditors' materiality judgments are not indigenous to the United States. Second, the prior empirical studies of auditors' materiality judgments were presented. Finally, several other publications relevant to this dissertation were discussed. The summary remarks which follow focus on the prior empirical studies of auditors' materiality judgments as a basis for positioning this research.

Various mathematical and statistical models have been used to represent auditors' materiality judgments. Boatsman and Robertson [1974] used multiple discriminant analysis. Morarity and Barron [1976] and Emery, et al. [1982], used conjoint measurement. Morarity and Barron [1979] used conjoint analysis, which assumed additivity, but not necessarily linearity. Ward used the Q-sort technique, and Newton [1977] used expected utility theory. Several of the recent judgment-capture experiments have used analysis of variance (ANOVA).
These include: Hofstedt and Hughes [1977]; Firth [1979]; Schultz and Reckers [1981]; Messier [1983]; and Krogstad, et al. [1984].

With one exception, all of the prior materiality-related research has focused on auditors' judgments at the evaluation stage. Only Moriarity and Barron [1979] explicitly examined materiality judgments at the planning stage; and that research occurred prior to the issuance of SAS Nos. 22 and 47, which required auditors to consider their "preliminary estimates of materiality" in audit planning. At that time, Moriarity and Barron reported that the task was unfamiliar to their eight subjects.

The dominant variable in all of the materiality studies has been the item in question as a percentage of net income. Other variables have not been consistently significant in the materiality judgments investigated.

Messier [1983] and Krogstad, et al. [1984], looked for experience effects associated with auditors' materiality judgments. The results have not been clear-cut. Using ANOVA, Messier found no mean difference among the experience groups for the materiality judgment; however, an experience effect was reported for the disclosure judgment. Krogstad, et al., reported that consensus was different for students relative to practitioner-subjects, but that consensus was not different for the audit seniors relative to the partners.

Woolsey [1954a], Dyer [1975], and Messier [1983] looked for systematic differences in the materiality judgments of auditors with
national firms relative to auditors with non-national firms (or, similarly, differences between Big Eight and non-Big Eight firms). The evidence has not been consistent. Furthermore, prior studies have not looked for systematic differences in auditors' materiality judgments among the Big Eight firms.

Although SAS No. 47 discusses the interrelationship of audit risk and materiality, and alludes to a notion known as "business risk," few of the prior research efforts have explicitly considered risk. Boatsman and Robertson [1974] included a risk variable which was defined as common stock price volatility on a five-point scale, but it only accounted for two percent of the model's predictive accuracy. Newton [1977] reported that, as expected, auditors' materiality thresholds moved inversely with the auditors' risk aversion, where risk dealt with uncertainty in a probabilistic sense.

In contrast to prior research, this study focuses on auditors' planning stage materiality thresholds at the aggregate financial statement level. A lens model framework will be used to investigate systematic differences associated with level of responsibility (presumably related to experience), firm affiliation, and audit technology. Further, by explicitly considering that auditors may differ in their perceptions of the likelihoods and/or their evaluations of the consequences associated with accepting/rejecting a client's financial representations, this research explores whether auditors may (rationally) have systematic differences in their
materiality thresholds related to attitudinal differences (as suggested by Crosby, et al. [1985]). Chapter III identifies the specific hypotheses to be examined, and discusses the research methodology to achieve the stated research objectives.
CHAPTER II FOOTNOTES

1. The FASB noted two instances in which the APB provided specific quantitative guidelines (see FASB [1975] paragraph 67):

   Note 2 of APB Opinion No. 15, "Earnings Per Share," states: Any reduction of less than 3% in the aggregate need not be considered as dilution in the computation and presentation of earnings per share data as discussed throughout this Opinion.

   Paragraph 17 of APB Opinion No. 18, "The Equity Method of Accounting for Investments in Common Stock," states: An investment (direct or indirect) of 20% or more of the voting stock of an investee should lead to a presumption that ... an investor has the ability to exercise significant influence over an investee.

   Similarly, the FASB has included specific quantitative materiality guidelines with its pronouncements on a piecemeal basis. For example, in Statement of Financial Accounting Standards No. 14, "Financial Reporting for Segments of a Business Enterprise," the FASB identified a reportable segment as a segment which meets one of the "tests" regarding 10 percent or more of (1) revenue, (2) operating profit or loss, or (3) total identifiable assets. (See FASB [1976, paragraph 15].)

2. As the title of SAS No. 47 suggests, audit risk and materiality judgments are obviously interrelated. SAS No. 47 (paragraph 2) defines "audit risk" as "... the risk that the auditor may unknowingly fail to appropriately modify his opinion on financial statements that are materially misstated." This type of risk is referred to in the statistical sampling literature as Type II error, unwarranted acceptance, or beta risk. Guy [1981, p. 133] comments, "Beta is sometimes called the handcuff or auditor's risk. Beta risk relates to audit effectiveness." According to the SAS, audit risk may be assessed in either quantitative or qualitative terms. The SAS points out (paragraph 17) that "... there is an inverse relationship between audit risk and materiality considerations."

   The definition of audit risk excludes the risk of qualifying the opinion with respect to financial statements that are not materially misstated. This concept of risk is referred to in the statistical sampling literature as Type I error, unwarranted rejection, or alpha risk. Guy [1981, p. 133] observes, "Auditors sometimes refer to alpha as the red face or client's risk. Alpha risk relates to audit
efficiency." Presumably, the ASB reasoned that the client's insistence would cause the auditor to investigate further and, therefore, reach the appropriate conclusion.

The ASB described another concept of risk, although not by name, in its footnote 1 of SAS No. 47. The SAS points out that this risk includes litigation and adverse publicity. "This exposure is present even though the auditor has performed his examination in accordance with generally accepted auditing standards and has reported appropriately on those financial statements" (AICPA [1983, note 1]). Brumfield, et al. [1983, p. 60], referred to this concept of risk as "business risk" which they defined as "... the probability that an auditor will suffer a loss or injury to his professional practice."

3. For a more detailed discussion of these pronouncements related to materiality from Canada, the United Kingdom, Australia, and New Zealand, see FASB [1975, paragraphs 90-103] and Selley [1984].

4. Holstrom and Messier [1982, p. 47] referred to a "... dearth of research concerning the use of the materiality concept in auditing ..." and observed that the focus of researchers and standard-setting bodies has been materiality in accounting. As described by Carmichael [1969], the two concepts are very much related, if not synonymous. Materiality in accounting deals with the fairness of the presentation of the financial statements; materiality in auditing deals with the design and execution of the audit approach to provide the auditor with the appropriate documentation and conclusions about the fairness of the financial statements. Selley [1984, pp. 10-11] summarizes his view of the distinction as follows:

... there is widespread agreement that materiality from the auditor's point of view at the final evaluation stage of the audit is the same as it is from the point of view of the preparers of the financial statements (so-called "accounting materiality"). However, SAS 47 has stated that the auditor for cost/benefit reasons may not in fact plan his procedures to detect some errors that would be qualitatively material. If this view prevails (as it is almost bound to do in the U.S.) we are again back to two materialities: "planning materiality" and "evaluation materiality."
5. Frishkoff's basic model was \( Y = f(x_1, x_2, \ldots, x_{17}) \) where:

- \( Y \) was the dependent variable (a dummy variable coded 1 if the opinion was qualified and 0 if it was not).
- \( x_1 \) was the income effect of the change divided by net income for the year.
- \( x_2 \) was the absolute value of the effect of the change divided by net income for the year. (The purpose of \( x_1 \) vs. \( x_2 \) was to segregate the amount of this change from its directions.)
- \( x_3, x_4, x_5, \) and \( x_6 \) represented the absolute value of the effect of the change divided by average net income for the past two, three, four, and five years, respectively.
- \( x_7 \) was the absolute value of the effect on 1962 net income of any change made retroactively to years prior to 1962 and divided by 1962 net income.
- \( x_8 \) was a dummy variable coded 1 for changes which were reclassifications and 0 otherwise.
- \( x_9 \) was the absolute value of the effect of any direct entries to retained earnings in 1963 divided by retained earnings at the end of fiscal year 1963.
- \( x_{10} \) was similar to \( x_9 \) except using 1962.
- \( x_{11} \) was the absolute value of the effect on income expressed on a per-share basis.
- \( x_{12} \) was a measure of the firm size of the CPA firm performing the audit.
$x_{13}$ was the magnitude of stockholders' equity.

$x_{14}$ was the same as $x_2$, except that it was not divided by net income. (The idea here was to determine whether the absolute magnitude or relative magnitude or both were factors considered in materiality judgments.)

$x_{15}$ was a dummy variable coded 1 if the firm reported a net loss and 0 otherwise.

$x_{16}$ was a dummy variable coded 1 if the change turned net income into a loss or vice versa and 0 otherwise.

$x_{17}$ was a dummy variable coded 1 if the CPA firm was a member of the "Big Eight" and 0 otherwise.

6. Moriarity and Barron cite Woolsey [1973b], and Pattillo and Siebel [1974] as evidence that auditors differ with respect to which variables they deem important in their materiality judgments. They cite Woolsey [1954a], Dyer [1975], and Pattillo [1975] as evidence that auditors differ in their materiality thresholds. Each of these studies was a survey and was, therefore, only amenable to limited statistical analysis.

7. In addition, Hofstedt and Hughes [1977] intended to partition their participants according to "risk-taking propensity," based on the Kogan-Wallach (twelve-item) Risk-taking Scale. However, they abandoned this analysis because (p. 384) "... (1) there was little difference between subjects [MBA students] and (2) the dimensionality of the measuring instrument is not clear...."

8. Messier's materiality judgment required his subjects to rank the materiality of a write-down of inventory (for each of 32 cases) on a seven-point scale. The disclosure task required the subjects to identify the probability that the write-down would be separately disclosed in the financial statements (i.e., a 101-point scale from "0" to "100").

It may not be surprising that Messier did not observe any mean difference among the experience groups (based on ANOVA) for the materiality judgments, since the sample consisted entirely of audit partners with varying years in practice. Within an internal control context, Ashton and Kramer [1980], which used students as subjects,
replicated Ashton [1974], which used auditors having one, two, or three or more years of experience. They reported differences in the judgments of the students relative to the auditors, for whom Ashton reported no differences. Therefore, Ashton and Kramer concluded that the important characteristic was having some experience, not necessarily how much.

9. Fishbein and Ajzen [1975, p. 12] describe "beliefs" as information a person has about some object of interest, and "... recommend that [belief] be measured by a procedure which places the subject along a dimension of subjective probability involving an object and some related attribute."

In contrast they argue (p. 13), "... the concept 'attitude' should be used only when there is strong evidence that the measure employed places an individual on a bipolar affective dimension." The term "affect" is also referred to synonymously as "evaluation."

They describe "behavioral intention" (p. 12) as a class of beliefs ("...in which the object is always the person himself and the attribute is always a behavior") and, hence, argue that the strength of the intention to perform a particular behavior should be measured along a subjective probability dimension.

They describe "behavior" as observable (or overt) acts by the subject.

10. Similarly, Cushing and Loebbecke [1983] classified the firms participating in their study as "highly structured," "semi-structured," "partially structured," and "unstructured." (See pp. 91-110 for their discussion and analysis regarding the firms' methodological structures.) The Cushing and Loebbecke classification cannot be used as a basis for identifying differences in the preliminary judgments about materiality related to audit technology, since they did not disclose the firms' identities.
CHAPTER III -- HYPOTHESES AND METHODOLOGY

Introduction

The purpose of this chapter is to identify the research issues of interest, including the specific hypotheses to be investigated. In addition, this chapter describes the general methodology to be used, including the participants, the experimental task, the independent variables, the experimental design, and the planned analyses to achieve the stated research objectives and to test the specific hypotheses.

In general, the research issues of interest include better understanding how auditors may quantify their preliminary judgments about materiality, which variables they attend to, and whether a consensus appears to exist. In particular, since substantial individual differences are expected, the study concentrates on certain factors which may be associated systematically with such differences: level of responsibility, firm affiliation, and audit technology. As a secondary issue of interest, this research also considers the notion of differing risk attitudes to explore a possible reason why auditors may differ in their materiality thresholds.

Specifically, within the context of auditors' planning stage materiality judgments, this study focuses on the following: (1) the effect of level of responsibility on consensus;¹ (2) the effect of
(Big Eight) firm affiliation on consensus; (3) the effect of differing audit technologies on consensus; and (4) the relevance of differing risk attitudes in explaining differences among auditors' materiality thresholds.

This study deals with auditors' preliminary judgments about materiality for three principal reasons. First, scarcely any of the prior research efforts have examined auditors' preliminary judgments about materiality. Of the studies cited in Chapter II, only Morriarity and Barron [1979] explicitly considered materiality judgments in a planning context; that study preceded the issuance of SAS No.'s 22 and 47 which direct auditors to consider their preliminary judgments about materiality in audit planning. Second, the preliminary judgment about materiality precedes the evaluation stage materiality judgment and, therefore, may be a more fundamental level at which to study auditors' materiality judgments.² Yet, insights gained at the planning stage may extend to the evaluation stage. Third, there should be actual auditing ramifications to these judgments at the planning stage. There is an inverse relationship between the auditor's materiality threshold and the amount of audit evidence which is necessary to corroborate the auditor's opinion (which implicitly allows for deviations which are not deemed material). All else equal, it would require more audit evidence to justify an auditor's opinion that a client's financial statements were fairly stated subject to a low materiality threshold (i.e., within a "tighter" range) than it would
if a higher materiality threshold were adopted. Preliminary judgments about materiality which are "too high" may lead to the gathering of insufficient audit evidence and, as a result, audits which are not in accordance with Generally Accepted Auditing Standards (GAAS). The ASB (AICPA [1983, note 1]) cautioned against (presumably, establishing unduly high materiality thresholds and, thereby,) performing insufficiently extensive audit tests as a consequence of an auditor's perception of low exposure: "Even if an auditor assesses this exposure as low, he should not perform less extensive procedures than would otherwise be appropriate under generally accepted auditing standards." Conversely, preliminary judgments about materiality which are "too low" may lead to audit procedures which are "too extensive" and, thereby, audits which are not efficient (i.e., "over-auditing"). Auditors' materiality judgments are relevant to the issue of determining the amount of audit evidence that is obtained—-to obtain a sufficient amount of evidence to support an auditor's opinion without obtaining more evidence than is necessary, since such evidence is costly.

The specific research hypotheses related to this study of auditors' planning stage materiality judgments will be described in the next section. The hypotheses are organized according to the research purposes introduced above. The primary research hypotheses deal with differences in consensus among auditors' planning stage materiality thresholds with respect to level of responsibility, firm
affiliation, and audit technology. The hypotheses of secondary interest deal with differences among auditors in their risk attitudes and the relationship of differences in their risk attitudes to differences in their planning stage materiality thresholds.

Following the identification of the hypotheses, the empirical research strategy to investigate the hypotheses is discussed. A judgment-capture experiment is selected as the preferred research approach. The discussion describes the subjects, the experimental task, the independent variables used, and the experimental design.

After that, the planned analyses to achieve the stated research purposes are identified and discussed. The analyses related to level of responsibility, firm, and audit technology effects are then discussed. These analyses are presented first with respect to location (i.e., regarding central tendency), and second with respect to variation (i.e., regarding dispersion). Then the analyses related to risk attitudes are described. Several additional analyses of interest follow.

The underlying assumptions related to the selected statistical methods are discussed after the planned analyses are identified. The chapter concludes with comments about the administration of the experiment.
Hypotheses

Level of Responsibility

As described in Chapter II, it is not clear what effect experience has on auditors' materiality judgments, in general, and their preliminary judgments about materiality, in particular. Therefore, one issue in this research project is the effect of differing levels of responsibility (presumably related to experience) on the consensus of auditors' preliminary judgments about materiality. Consensus is examined in terms of two separate dimensions based on the recommendation of Gaumnitz, et al. [1982]—(1) location (regarding central tendency) and (2) variation (regarding dispersion). To address the effect of different levels of responsibility on auditors' preliminary judgments about materiality, the following hypotheses (in null form) are offered:

H1.1: Level of responsibility is unrelated to differences in location among auditors' planning stage materiality thresholds.

H1.2: Level of responsibility is unrelated to differences in variation among auditors' planning stage materiality thresholds.

Based on Krogstad, et al. [1984], which found no difference among seniors and partners, the three levels of responsibility are not expected to differ in their planning stage materiality judgments, either as to location or variation. However, Krogstad, et al. focused on an evaluation stage materiality judgment task, so their results may not extend to this planning stage task.
Firm Affiliation

As discussed in Chapter II, Cushing and Loebbecke [1983] and Kinney [1986] distinguished the large CPA firms by their differing audit technologies. In addition, Cushing and Loebbecke pointed out that these firms had conspicuous differences in the way they dealt with their preliminary judgments about materiality. Therefore, even among the Big Eight firms, which are often viewed as homogeneous, there is reason to expect that differences in planning stage materiality judgments may occur at the firm level. The following hypotheses (in null form) address the appropriateness of assuming that the Big Eight firms are homogeneous within the context of planning stage materiality judgments:

H2.1: Firm affiliation is unrelated to differences in location among auditors' planning stage materiality thresholds.

H2.2: Firm affiliation is unrelated to differences in variation among auditors' planning stage materiality thresholds.

Cushing and Loebbecke [1983] reported substantial differences among twelve of the largest CPA firms in the United States as to how they dealt with the preliminary judgment about materiality. On that basis, it is expected that firm differences among the participating Big Eight firms will be detected in this study with respect to location, variation, or both.
Audit Technology

A related issue involves audit technology. Cushing and Loebbecke [1983] and Kinney [1986] examined several characteristics as a basis for classifying firms according to audit structure (e.g., the extensiveness of sampling procedures). The classifications were consistent across the dimensions considered. That consistency, along with Cushing and Loebbecke's observation that there were conspicuous differences in the firms' treatments of planning stage materiality judgments, and the structured firms' emphasis on consensus, as described earlier by Kinney, provide the basis for believing that the notion of audit technology should be considered in this study. To the extent that differences are identified among the participating firms, systematic differences between the relatively structured and the relatively unstructured firms are of interest. Accordingly, the following hypotheses (in null form) will be examined:

H3.1: Audit technology is unrelated to differences in location among auditors' planning stage materiality thresholds.

H3.2: Audit technology is unrelated to differences in variation among auditors' planning stage materiality thresholds.

As discussed above, firm differences are expected. Further, based on Kinney [1986], differences as to location and/or variation, at the firm level, are expected to relate to differing audit technologies. In particular, the differences are expected to be
greater between the structured and unstructured firms than among firms within the same audit technology group.

Risk Attitude

Although the primary issue of interest is the identification of systematic differences among auditors in their planning stage materiality judgments (in terms of level of responsibility, firm affiliation, and audit technology), a secondary issue explores certain reasons why auditors may specify different materiality thresholds. As described in Chapter II, Crosby, et al. [1985] used an economic modeling approach, based on the maximization of subjective expected utility, to analytically model an auditor's optimal materiality threshold. The optimal materiality threshold minimizes the total expected costs comprised of the costs associated with issuing an unqualified opinion plus the costs associated with issuing a qualified opinion. (See Figure 3 for a simple $2 \times 2$ representation.) If we may assume, for simplicity, that the cost of a "correct" opinion choice is zero, then the auditor's focus becomes one of minimizing the combined costs of Type I and Type II errors. $^3$

Some auditors may emphasize so-called Type II errors due to their concern about litigation and unfavorable publicity if it is discovered that an unqualified opinion was rendered despite the existence of a "material" misstatement. They may be expected to be relatively more
<table>
<thead>
<tr>
<th>Action:</th>
<th>No Material Misstatements Exist</th>
<th>Material Misstatements Exist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concludes that there are material misstatements (qualified opinion --</td>
<td>A Type I Error</td>
<td>A Correct Judgment</td>
</tr>
<tr>
<td>&quot;GAAP departure&quot;)</td>
<td>{Cell #1}</td>
<td>{Cell #2}</td>
</tr>
<tr>
<td>Concludes that there are no material misstatements (unqualified opinion)</td>
<td>A Correct Judgment</td>
<td>A Type II Error (&quot;Audit Risk&quot;)</td>
</tr>
<tr>
<td></td>
<td>{Cell #3}</td>
<td>{Cell #4}</td>
</tr>
</tbody>
</table>

Note: The idea represented by the above matrix is that the auditor may have to render an opinion while uncertainties about some of the financial statement items remain. These uncertainties will not be resolved until a later point in time. It is hypothesized that auditors' individual differences with respect to their perceptions of the likelihoods and the consequences associated with these cells may be systematically related to the differences in their preliminary judgments about materiality.

Figure 3

(Implicit) Trade-offs Inherent in Auditors' Materiality Judgments
inclined to issue qualified opinions and therefore specify relatively lower materiality thresholds. Alternatively, other auditors may emphasize so-called Type I errors due to their concern about client retention and practice development. They may be expected to be relatively less inclined to issue qualified opinions and, hence, specify relatively higher materiality thresholds. Thus, one may expect that the auditor's materiality threshold will increase as the (perceived) relative costs of a Type I error increase and that the threshold will decrease as the relative costs of a Type II error increase. "Risk attitude" in this study may be viewed as an intervening variable affecting the preliminary judgment about materiality as depicted in Figure 4.

As described in Chapter II, Kida [1980] demonstrated that auditors' differing perceptions of the consequences associated with issuing going concern qualifications affected their decisions to issue such qualifications in an experimental context. Similar issues exist with respect to auditors' materiality judgments, since trade-offs presumably must be made between possible Type I and Type II errors. The following hypotheses (in null form) deal with the trade-offs between Type I and Type II errors that are believed to be implicit in auditors' materiality judgments:

H4.1: Auditors do not differ significantly in their attitudes toward trading-off Type I and Type II errors.
Figure 4
Variables Used
H4.2: Auditors' attitudes toward trading-off Type I and Type II errors are unrelated to their planning stage materiality thresholds.

The above hypotheses deal with the assertion that some of the variation in auditors' materiality thresholds may be due to differences in their perceptions of the likelihood and/or the consequences associated with Type I versus Type II errors. An alternative characterization is that auditors may differ in their predispositions toward issuing qualified opinions. That is, those auditors who are relatively more sensitive to Type II errors (e.g., litigation by financial statement users) might be expected to be more likely to issue qualified opinions. Those auditors who are relatively more sensitive to Type I errors (e.g., practice development considerations) might be expected to be less inclined to issue qualified opinions. H4.1 is concerned with whether there are systematic differences among the auditors in their risk attitudes regarding the trade-off of Type I and Type II errors. H4.2 is concerned with whether there is a significant association between differences in auditors' risk attitudes and differences in their preliminary judgments about materiality.

The selected research approach will be described next. The choice is justified in view of the attributes desired to achieve the stated research purposes.
Empirical Research Strategy

In order to achieve the study's research purposes, a number of characteristics should be associated with the chosen research strategy. In general, any study must be conducted with a reasonable cost (including a reasonable amount of time committed by any participants). Further, the obtrusiveness of the researcher should be minimized, in order to strengthen the validity of the study. Finally, to enhance generalizability (i.e., external validity), the study should be as realistic as possible.

Specifically, with respect to the stated research objectives, it is essential that the participants address the same scenarios. Since the study makes comparisons across the participants (in terms of level of responsibility, firm, and audit technology effects, and in terms of different risk attitudes), the participants must face identical cases for the comparisons to be meaningful.

In order to model the participants' judgments, several other attributes are desired. First, it is necessary that certain variables of interest be capable of being manipulated as independent variables. Second, and related to the first, is the need for multiple judgments by each participant. Third, the measurements must be objective and capable of relatively precise measurement. These qualities will make it possible to infer causality and enable the research to be replicated.
The stated hypotheses will be explored by conducting a (judgment-capture) behavioral experiment. Although the research approach employed does not constitute a "pure" laboratory experiment, it is similar in a number of respects. Therefore, the characteristics of a laboratory experiment are considered first, and the differences associated with the chosen approach are discussed.

In the laboratory experiment, the researcher creates the context or setting in which the study is conducted; subjects recognize that the setting is not "natural" and that the primary purpose of their involvement is for research. This strategy is characterized by a high degree of control over the independent variables and their levels; confounding factors can be effectively controlled. Stone [1978, p. 118] defines a "laboratory" as "... any setting in which the experimenter has a high degree of control over the stimuli to which experimental subjects are exposed and the conditions associated with the observation of behavior." The researcher emphasizes the fundamental processes or aspects of the behavioral system of interest, rather than emphasizing the duplication of the naturally occurring behavioral system itself.

The major advantages of the laboratory experiment include the strong internal validity which can be achieved. Measurements can be made relatively precisely and causality can be inferred. Each participant can be exposed to the same scenarios so that meaningful comparisons across subjects can be made. The independent variables of
interest can be specified and their levels designated; and confounding factors can be controlled effectively. Multiple observations can be obtained across different levels of the independent variables. As a result, laboratory experiments are capable of replication. In addition, the cost of the research, including time demands on participants, can be modest.

The principal disadvantage of the laboratory experiment is its artificial setting. As a result of potentially weak external validity, it may not be meaningful to extrapolate the results of the experiment into realistic (natural) settings. In addition, Cook and Campbell [1979, pp. 50-80] describe a number of threats to the validity of research. Regarding the laboratory experiment strategy, a number of "artifacts" (including hypothesis-guessing by the participants, evaluation apprehension, and experimenter expectancies\(^4\)) may affect the results of an experiment.

In view of the stated research objectives, an experiment is the most appropriate approach for the present study. It will facilitate the desired comparisons to be made across subjects (in terms of level of responsibility, firms, audit technologies, and different risk attitudes), and permit the modeling of the participants' judgments. The required research costs, in particular the time demands on the participants, are reasonable.\(^5\) An incidental advantage of conducting an experiment is that it may facilitate any desired comparisons with
prior studies which have used an experimental approach to examine auditors' materiality judgments in an evaluation stage context.

It was not practical to use a "pure" laboratory experiment approach. In order to lessen the inconvenience to the participants and thereby induce the firms to participate, it was necessary to relinquish some control over the conditions under which the experiment was conducted. Since the experimental task was estimated to require about two hours for each participant, allowing the participants to complete the experimental materials at their convenience was deemed necessary. Requiring the participants to complete the materials under the researcher's supervision at a uniform campus location or at the individual firms' offices was impractical. In that sense, the research approach combined some features typically associated with a laboratory approach with other features usually associated with a questionnaire-survey approach.

Although an experimental approach may be susceptible to criticism about realism and obtrusiveness, Swieringa and Weick [1982, p. 79] point out that "... artificiality may make it easier to observe clear disconfirmations of theory and examples of novel relationships that are normally concealed in the sheer mass of covarying variables in realistic settings." In order to strengthen external validity, practicing auditors were used as subjects; the experimental task was couched in relatively realistic terms; and the participants were asked to complete the experimental materials in their offices where they
normally work. Permitting the participants to complete the experimental materials in their regular work environments also served to minimize the obtrusiveness of the researcher, since the subjects could complete the materials at their own convenience. Based on the other dimensions considered (i.e., cost, manipulation of independent variables, use of the same scenarios, multiple observations, and relatively objective measurements), an experimental approach is preferred.

**Subjects**

This study involves practicing auditors at the senior, manager, and partner levels. The "contact person" at each firm indicated that the seniors, managers, and partners make planning stage materiality judgments in practice. Based on representations by these individuals that staff-level subjects would not generally be involved with the real-world analogue of the experimental task, staff-level auditors were excluded from this experiment.

As described in Chapter II, there may be systematic differences in the judgments of auditors across Big Eight firms. In addition, there may be differences in the judgments of auditors with "structured" firms relative to auditors with "unstructured" firms. The Columbus, Ohio, offices of seven of the Big Eight firms were asked to participate; Arthur Young was not contacted due to the small size of their Columbus office. The following firms participated:
(1) Arthur Andersen; (2) Coopers & Lybrand; (3) Deloitte, Haskins, & Sells; (4) Peat, Marwick, Mitchell; (5) Price Waterhouse; and (6) Touche Ross. Participants from these firms were auditors in their Columbus, Ohio, offices. Ernst & Whinney indicated that they could not participate because they were unusually busy and working a lot of overtime during the summer.

Thus, the sample consisted of 72 professional auditors--12 from each of the six participating Big Eight firms; there were 24 auditors each at the senior, manager, and partner levels. Half of the subjects were affiliated with firms classified as relatively "structured," and the other half were from firms classified as relatively "unstructured."

Experimental Task

The experimental materials consisted of two booklets. The first booklet contained the instructions to the participants and the background information about the hypothetical audit client. (See Appendix A for the instructions and background information provided to the subjects in booklet #1.) Since the background information was applicable to each specific case in the second booklet, the first booklet was paper-clipped to the inside of the second booklet so that the participants could consult the background information easily as desired. The second booklet consisted of three sections: (1) the 28 combinations (including four repetitions) of the three independent
variables for which subjects were asked to provide their planning stage materiality judgments at both the aggregate income statement and balance sheet levels; (2) the attitude measurement questionnaire (see Appendix B) which dealt with auditors' differing sensitivities to so-called Type I and Type II errors; and (3) a survey or debriefing questionnaire (see Appendix C) which solicited additional personal information about each participant.

As described above, a laboratory experiment of audit judgment may be criticized for its artificiality. However, that concern may be mitigated to some extent by providing relatively realistic information to the participants. The background information included the principal items which SAS No. 22 suggested that auditors should consider when involved in audit planning.9 Further, the case materials were modeled from actual retail companies' annual reports, and subjects were so informed.

Participants were asked to specify the dollar amounts which they believed constituted the appropriate planning stage materiality thresholds for the hypothetical client for each case. Specifically, subjects were asked to make two materiality judgments for each case—one representing their materiality threshold with respect to the income statement; the other representing their materiality threshold with respect to the balance sheet. (Figure 5 provides one page of the document used to elicit subjects' income statement and balance sheet materiality judgments.10) As described in Chapter II, SAS No. 47
FINANCIAL STATEMENT AMOUNTS:
Total assets = $35,388,746
Net sales = $119,202,666
Income (loss) before taxes = $1,429,976

PRELIMINARY JUDGMENT ABOUT MATERIALITY:
Applicable to the Income Statement $___________________
Applicable to the Balance Sheet $___________________

BUSINESS RISK: 1 2 | 3 4 5 | 6 7
(please circle) (very) (very)

Figure 5
Sample of Materiality Judgment Questionnaire
(paragraph 12) indicates that a planning stage materiality judgment should be made at the aggregate financial statement level; the appropriate preliminary judgment about materiality would be the smallest materiality judgment applicable to any one of the financial statements. Consequently, it is appropriate to ask the participants to view the planning stage materiality judgment separately with respect to the income statement and the balance sheet. Subjects were permitted, if they desired, to refer to their firms’ materials and other professional literature to assist them in their judgments.

As described in Chapter II, prior research has consistently reported net income (or income before taxes) to be the dominant variable in studies of materiality judgments at the evaluation stage. In those studies, balance sheet items (e.g., total assets) have tended to be unimportant variables. In addition, Zuber, et al. [1983, p. 43] asserted that planning stage materiality judgments are driven to a much greater extent by income statement rather than balance sheet considerations. Therefore, in this study, the income statement materiality judgments are deemed of primary interest. Nevertheless, comparable analyses will be conducted separately with respect to the income statement and balance sheet judgments, since so little is known about materiality judgments at the planning stage.

Subjects' risk attitudes were measured by their responses to a questionnaire (Appendix B), after their materiality judgments were elicited. The attitude measure used an expectancy-valence approach to
assess participants' attitudes toward the four cells represented by Figure 3:

1. qualifying the opinion on financial statements which are not materially misstated (i.e., a Type I error);
2. qualifying the opinion on financial statements which are materially misstated (a correct judgment);
3. not qualifying the opinion on financial statements which are not materially misstated (a correct judgment);
4. not qualifying the opinion on financial statements which are materially misstated (i.e., "audit risk" as defined in SAS No. 47; a Type II error).

As described in Chapter II, the expectancy-valence model requires measurements of "beliefs" (i.e., assessments of likelihoods) and "evaluations" (i.e., assessments of importance) for outcomes of interest. Outcomes which might be associated with each of the four scenarios above were identified. The principal source for these outcomes was Kida [1980], but additional consideration was given to the items used by Ward [1976] and to discussions with the "contact" people at the participating CPA firms. Participants were asked to express their beliefs that these outcomes would occur given a particular scenario (e.g., qualifying the opinion on financial statements which are not materially misstated—a Type I error). Separately, participants were asked to provide their evaluations of the positive or negative impact of the outcomes associated with the
four scenarios above. As shown in Appendix B, the belief statements were represented by a seven-point scale where "1" represented "practically no chance that this would occur" and "7" represented "this almost certainly would occur." The evaluations were based on a five-point scale where "1" represented "neither bad nor good (indifferent)" and "5" represented "extremely bad" for unfavorable outcomes or "extremely good" for favorable outcomes.\footnote{11}

A "survey" questionnaire (frequently referred to as a "debriefing" questionnaire in the accounting literature) was administered at the end of the experiment to accumulate certain demographic and other background information about each participant. (See Appendix C for the survey questionnaire.)

**Independent Variables**

As indicated above, this study was predicated on the expectation that there would be substantial individual differences among the participating auditors in their planning stage materiality thresholds. In order to gain further insights as to the nature of these differences, individual models for each subject's judgments were desired. The purpose of these individual models was to identify whether the structural form of the models differed among the subjects and whether any such differences in the structural forms of the models related to differences in their materiality thresholds. To make
possible the modeling of subjects' judgments, multiple observations encompassing the desired independent variables were needed.

Three quantitative independent variables related to the client's financial statements were manipulated: (1) income (loss) before taxes; (2) net sales; and (3) total assets. Although empirical support for the selection of independent variables is somewhat lacking, Gafford and Carmichael [1984, p. 112] assert that these three are commonly used in practice as a basis for the preliminary judgment about materiality.

As indicated in Chapter II, prior research into auditors' evaluation stage materiality judgments has consistently shown the item in question as a percentage of net income to be the dominant variable. Based on the significance of net income in those studies, in addition to Morarity and Barron [1979] which focused on planning stage materiality judgments, the income (loss) before taxes variable is likely to be important in the present judgment task. However, in those prior studies, net sales and total assets have been relatively insignificant variables.12

**Experimental Design**

Two levels were selected for net sales; three levels were selected for total assets; and four were selected for income (loss) before taxes. As mentioned in the preceding paragraph, the income (loss) before taxes variable was expected to be the dominant variable
for most auditors in quantifying their preliminary judgments about materiality with respect to the income statement. In addition to two levels of the variable which are relatively high (one may be viewed as "high" and the other as "medium"), one level was chosen at approximately the break-even level (since it was deemed of interest for future purposes to gather data about auditors' judgments as they approach a zero level of income). Finally, a loss situation was included of approximately the same absolute magnitude as the next to highest positive value; it was of interest for future purposes to gather data about whether the auditors' judgments were symmetrical between income and loss situations or whether the judgments were conditional on the positive or negative nature of the income (loss) before taxes. Only two levels of sales were included, since it was not expected that sales would be a significant variable for many of the participants. However, including two levels was sufficient to ascertain whether the level of sales was a significant variable in the auditors' judgments (either as a primary variable or as an interaction with one of the other variables). Three levels of total assets were chosen (representing "high," "medium," and "low") since it was expected to be the dominant variable with respect to the balance sheet materiality judgment. Using three levels made it possible to identify, for future purposes, whether the judgments with respect to the balance sheet were of a linear form. That was of interest since
prior research has not distinguished between materiality judgments with a view toward the balance sheet versus the income statement.

The specific amounts for the cues were determined judgmentally. **Ward's Directory of 51,000 U.S. Corporations** (Ward [1984]) provided a relative indication of "high," "medium," and "low" levels for the cues, as applicable, for retail department stores. After particular values were selected for income (loss) before taxes (four levels), sales (two levels), and total assets (three levels), the contact person at each firm was asked, among other things, whether the different amounts for these variables were reasonable within the context of the retail environment of Columbus, Ohio. Without exception, these auditors agreed that the levels of the variables were reasonable.

A full factorial, repeated measures design was used. (Figure 6 depicts the experimental design and identifies the specific levels of each independent variable.) Four cases were repeated in order to measure (test-retest) judgment consistency. Consequently, each subject was asked to make two materiality judgments (one with respect to the income statement and the other with respect to the balance sheet) for each of 24 different fact situations--representing each combination of net sales, total assets, and income (loss) before taxes--plus the four repeat cases. A full factorial design was chosen instead of a fractional factorial design, since it was of interest to detect the existence of any interactions among the independent
Income (loss) before taxes: 

<table>
<thead>
<tr>
<th>IBT_1</th>
<th>IBT_2</th>
<th>IBT_3</th>
<th>IBT_4</th>
</tr>
</thead>
<tbody>
<tr>
<td>($1,787,446)</td>
<td>$215,712</td>
<td>$1,429,976</td>
<td>$4,366,238</td>
</tr>
</tbody>
</table>

Net sales: 

<table>
<thead>
<tr>
<th>S_1</th>
<th>S_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$119,202,666</td>
<td>$205,790,182</td>
</tr>
</tbody>
</table>

Total assets: 

<table>
<thead>
<tr>
<th>TA_1</th>
<th>TA_2</th>
<th>TA_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$9,260,686</td>
<td>$35,388,746</td>
<td>$90,591,695</td>
</tr>
</tbody>
</table>

Figure 6
Experimental Design
variables. A fractional factorial design has the advantage that it requires fewer observations and therefore would require less time by the participants. However, the major disadvantage is that interactions may not be identifiable (Montgomery [1984]). Since so little is known about auditors' materiality judgments in a planning context, all the interactions were deemed potentially interesting, and a full factorial design was preferred.

**Analyses To Be Performed**

As described above, the purpose of this study is to identify systematic differences in the planning stage materiality judgments among Big Eight auditors in an experimental context. The differences of primary interest relate to level of responsibility, firm affiliation, and audit technology. The analyses focus on consensus as a criterion by which to compare the judgments of the participating auditors. "Consensus" is examined in terms of two separate dimensions, location and variation. A secondary factor of interest deals with auditors' differing risk attitudes. The discussion which follows describes the particular analyses to be performed in order to achieve the stated research objectives. Analyses will be performed separately with respect to the income statement and balance sheet materiality judgment tasks relative to the primary hypotheses (i.e., H1.1, H1.2, H2.1, H2.2, H3.1, and H3.2). The income statement
materiality judgments are of primary interest, and the results of the analyses of the income statement judgments are reported in Chapter IV. Since the balance sheet materiality judgments are of lesser interest, the results of the analyses are reported in Appendix D for the interested reader.

Conceivably, the subjects might differ systematically on many (or perhaps most) of the cases, without necessarily differing consistently on every case. Therefore, a single measure is desired as a basis for making overall inferences with respect to systematic differences among the subjects. This will avoid having to qualify the observations with respect to particular case numbers. This overall measure should capture the relative extremeness of each individual's judgments, as well as provide a reference point from which to measure variability within certain groupings of the participating auditors.

For all the subjects in the aggregate, the responses will be transformed to a distribution with mean zero and variance one, on a case-by-case basis. For a particular subject, an above average response for a given case would be represented by a positive score; an individual with a below average response on that case would be represented by a negative score. Each individual's standardized scores will be summed across the 24 unique cases to provide an overall measure of the extremeness of his or her materiality thresholds. These measures will be referred to as the "standardized materiality scores."
Level of Responsibility Effects

"Experience" will be represented by individuals' levels within the hierarchy of senior, manager, and partner. An alternative construct would be to measure experience in terms of years as an auditor. The former was chosen for three basic reasons. First, a high degree of correlation has been reported between subjects' levels within the hierarchy and their years of audit practice. Prior studies which have considered experience in terms of both constructs have reported virtually identical results (e.g., Nanni [1984]). Both of these constructs may be viewed as surrogates for the number of times an auditor has actually made planning stage materiality judgments similar to the experimental task. It was not deemed appropriate to operationalize experience directly in terms of the number of times each individual has actually made such planning stage materiality judgments, due to the inherent subjectivity of such self-reports. Second, the specified level of responsibility provides a convenient, objective, and meaningful partition of the subjects for purposes of investigating the existence of experience effects. Partitioning the subjects into categories based on their years of experience would be far more subjective. Third, the level of responsibility is believed to have implications with respect to "risk attitude." The differing responsibilities associated with the
respective levels of the CPA firms may affect the priorities and concerns of the individual auditors.

Each subject can be viewed as a member of one of the three levels of responsibility, while simultaneously a member of one of the six participating firms, while simultaneously a member of one of two audit technologies. Consequently, subjects were nested within level of responsibility, within firm, within audit technology. In the context of this study, systematic differences among subjects' materiality thresholds might arise from two distinct sources—level of responsibility and firm affiliation (along with audit technology, which is related to firm affiliation). The effects associated with these different factors must be separated in order for the conclusions to be valid. The planned analyses reflect the need to separate the differences associated with level of responsibility from the differences associated with firm affiliation. The statistical tests will identify whether there are differences among the three levels of responsibility, first as to location and second as to variation.

Location

The first analyses to be conducted are concerned with differences among the three levels of responsibility with respect to location (i.e., central tendency). A one-way ANOVA involving levels of responsibility is not appropriate because such an analysis ignores the influence of firm affiliation on subjects' responses while considering
only the effect of level of responsibility. In order to disentangle the effects associated with the two distinct factors, one possibility would be to construct a two-way ANOVA, reflecting both levels of responsibility and firm affiliation. However, such an analysis would implicitly test whether the effects associated with the levels of responsibility were constant across the six firms. Conceivably, there may be a differential effect across the firms.

The first analyses with respect to location will be conducted firm-by-firm to ascertain whether any of the three levels of responsibility differ. Using the notation of Figure 7, the following comparisons will be made: $A_1$, $B_1$, and $C_1$; $A_2$, $B_2$, and $C_2$; ...; and, finally, $A_6$, $B_6$, and $C_6$.

Differences in location with respect to the three levels of responsibility will be tested, firm-by-firm, by one-way ANOVAs of the standardized materiality scores. If level of responsibility is found to be significant for one or more of the firms, Tukey's all pairwise multiple comparison test will be used to identify which of the levels of responsibility differ significantly from one another.

A p-value with respect to level of responsibility will be obtained for each firm. In order to test H1.1, regarding differences in location among the three levels of responsibility, the p-values obtained in the one-way ANOVAs for each firm will be combined into an overall test statistic using Fisher's method for combining separate
\[
\begin{array}{cccc|cccc}
\text{Structured Firms} & F_1 & F_2 & F_3 & \text{Sub-T} & F_{1-3} & \text{Unstructured Firms} & F_4 & F_5 & F_6 & \text{Sub-T} & F_{4-6} & \text{Total} & F_{1-6} \\
\hline
\text{PARTNER} & A_1 & A_2 & A_3 & A_s & A_4 & A_5 & A_6 & A_u & A \\
\text{MANAGER} & B_1 & B_2 & B_3 & B_s & B_4 & B_5 & B_6 & B_u & B \\
\text{SENIOR} & C_1 & C_2 & C_3 & C_s & C_4 & C_5 & C_6 & C_u & C \\
\hline
\text{TOTAL} & T_1 & T_2 & T_3 & T_s & T_4 & T_5 & T_6 & T_u & T
\end{array}
\]

where:
- Each \( A_1, A_2, A_3, \ldots, C_6 \) represents 4 subjects;
- each \( A_s, \ldots, C_s, A_u, \ldots, C_u \) represents 12 subjects;
- each \( A, \ldots, C \) represents 24 subjects;
- each \( T_1, \ldots, T_6 \) represents 12 subjects;
- both \( T_s \) and \( T_u \) represent 36 subjects;
- \( T \) represents 72 subjects.

**Figure 7**

Notation Regarding Comparisons To Be Made
tests. Hedges and Olkin [1985, p. 37] describe Fisher's method for combined tests (sometimes referred to as "the inverse chi-square method") as "... the most widely used combination procedure..." Based on the six independent one-way ANOVAs yielding the p-values for each of the participating firms, the test procedure is as follows:

\[ \text{reject H1.1 if } P = -2 \sum_{i=1}^{6} \ln(p_i) \geq C, \text{ where } C \text{ is the critical value based on the chi-square distribution with 12 degrees of freedom (i.e., } 2k \text{ where } k=6 \text{ firms).} \]

The inability to reject H1.1 based on Fisher's method for combining results would be interpreted as finding no difference in means among levels of responsibility within any of the six participating firms. This (null hypothesis) is more general than testing whether there is a constant difference associated with levels of responsibility across the six firms.

A supplemental comparison examines differences among the three levels of responsibility within each of the two audit technology groups. The purpose of this additional comparison is to gain possible insights regarding any differential effect related to audit technology. Intuitively, one might expect experience to have a greater impact on consensus within less structured firms, since the structured firms, to a greater extent, have adopted detailed guidelines to direct the judgments of their auditors at all levels of responsibility. In terms of Figure 7, the following comparisons will be made: \( A_S, B_S, \text{ and } C_S; \) and \( A_U, B_U, \text{ and } C_U. \) With respect to location, the planned tests consist of two separate one-way ANOVAs
dealing with level of responsibility (first among the structured firms, and second among the unstructured firms).

**Variation**

Similar comparisons will be conducted to identify any differences among the levels of responsibility with respect to variation. Overall tests for homogeneity of variance cannot be appropriately conducted in the aggregate among the levels of responsibility, since that would ignore the potential firm effect as a source of variation. Accordingly, the tests for homogeneity of variance will be applied on a firm-by-firm basis.

The sums of squares, related to variance, will be calculated for each level of responsibility, separately for each firm, based on the standardized materiality scores. Cochran's test for homogeneity of variance (Kirk [1982, p. 78]) will be applied, firm-by-firm, to the sums of squares associated with the seniors, managers, and partners. The p-values will be calculated with respect to level of responsibility for each firm. As described above, Fisher's method for combining tests will be used to combine the p-values for the six firms into an overall test of H1.2 (regarding level of responsibility effects with respect to variation).

A supplemental comparison examines differences among the three levels of responsibility separately within each of the two audit technologies. With respect to differences in variation, Cochran's
test for homogeneity of variance will be applied to the sums of squares calculated separately for the seniors, managers, and partners within each audit technology. Table 4 summarizes the principal analyses to be performed with respect to level of responsibility.

**Firm Effects**

As suggested in the preceding section, there are two distinctly different potential sources of systematic effects in this study. Level of responsibility effects were discussed above; firm effects are discussed in this section. (Audit technology effects are related to firm affiliation and are discussed in the next section.) Based on Krogstad, et al. [1984], which found no difference among seniors and partners in an evaluation stage context, no differences are expected among the seniors, managers, and partners in their planning stage materiality judgments. The planned analyses with respect to firm (and audit technology) effects reflect that expectation; if level of responsibility effects are significant, then the analyses of firm (and audit technology) effects will have to consider those experience-related effects simultaneously.\(^{18}\)

Similar to the level of responsibility effects, two hypotheses were stated with respect to firm effects, regarding differences in auditors' planning stage materiality judgments associated with Big
Table 4
Analyses of Level of Responsibility Effects

<table>
<thead>
<tr>
<th>Comparisons To Be Made (Notation per Figure 7)</th>
<th>Data To Be Used</th>
<th>Primary Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Location:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Level of responsibility, by firm affiliation</td>
<td>Materiality scores standardized across all 72 subjects, partitioned by firm</td>
<td>One-way ANOVA for each firm</td>
</tr>
<tr>
<td>$A_1, B_1, C_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_2, B_2, C_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\vdots$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_6, B_6, C_6$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Overall (H1.1)</td>
<td>Same as above</td>
<td>P-values for firms combined by Fisher's method</td>
</tr>
<tr>
<td>Same as above</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Level of responsibility, by audit technology</td>
<td>Same as above, but partitioned by audit technology</td>
<td>One-way ANOVA for each audit technology</td>
</tr>
<tr>
<td>$A_s, B_s, C_s$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_u, B_u, C_u$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II. Variation:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Level of responsibility, by firm affiliation</td>
<td>Sums of squares (SS's) for each level of responsibility for each firm</td>
<td>Cochran's test for homogeneity of variance</td>
</tr>
<tr>
<td>Same as I(A) above</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Overall (H1.2)</td>
<td>Same as II(A) above</td>
<td>P-values for firms combined by Fisher's method</td>
</tr>
<tr>
<td>Same as I(A) above</td>
<td>Same as II(A) above</td>
<td></td>
</tr>
<tr>
<td>C. Level of responsibility, by audit technology</td>
<td>SS's associated with each level of responsibility for each audit technology</td>
<td>Cochran's test for homogeneity of variance</td>
</tr>
<tr>
<td>Same as I(C) above</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Eight firm affiliation. H2.1 deals with differences at the firm level with respect to location; H2.2 deals with differences at the firm level with respect to variation. In terms of Figure 7, the overall tests regarding firm effects (both with respect to location and variability) will be based on $T_1$, $T_2$, $T_3$, $T_4$, $T_5$, and $T_6$.

A supplemental comparison deals with firm differences at each level of responsibility to provide further insights regarding firm effects. Comparisons will be made across the participating firms, by level of responsibility, with respect to location and variation. In terms of Figure 7, the following comparisons will be made: $A_1$, $A_2$, ..., $A_6$; $B_1$, ..., $B_6$; and $C_1$, ..., $C_6$. Should differences be found at the firm level, this type of comparison will indicate how pervasive those differences are across the hierarchy of the firms.

Another supplemental comparison tests for firm differences within each audit technology group (i.e., firm differences within the structured grouping and, separately, firm differences within the unstructured grouping). In terms of Figure 7, the following comparisons will be made: $T_1$, $T_2$, and $T_3$; and $T_4$, $T_5$, and $T_6$. This type of comparison will provide insights regarding the relevance of audit technology in explaining any observed firm differences.
Location

H2.1 (regarding differences at the firm level with respect to location) will be tested by running a one-way ANOVA of the standardized materiality scores. If firm affiliation is found to be significant, Tukey's pairwise comparison test will be used to identify which firms differ from one another.¹⁹

A supplemental comparison explores whether firm effects are present at each (or any) of the levels of responsibility. A one-way ANOVA will be run separately for each level of responsibility. In each instance that a significant firm difference is detected, Tukey's pairwise comparison test will be used to identify which firms differ.

Similarly, another supplemental comparison examines firm differences, within each audit technology group. (There are three firms in each of the two audit technology groups.) A one-way ANOVA will be run separately for each of the two audit technology groups. Again, for any significant result, Tukey's pairwise comparison test will be used to identify which of the firms differ.²⁰

Variation

To test H2.2 (regarding firm differences with respect to variation), Cochran's test for homogeneity of variance will be applied to the sums of squares associated with each firm.²¹ This test will indicate whether the variances among the firms are the same or whether at least one of the firms is significantly more variable than at least
one of the others. Cochran's test for homogeneity of variance will be applied sequentially among the firms, deleting each firm that is found to be the most variable. This will provide an indication of the extensiveness of any difference among the firms with respect to variation.

A supplemental comparison investigates differences among the firms, separately, at each level of responsibility. Cochran's test for homogeneity of variance will be applied to the variances calculated, by firm and by level of responsibility. This test will indicate whether firm affiliation is significantly related to differences in variability for any or all of the levels of responsibility.

Finally, another supplemental comparison investigates whether there are differences in variability among the firms within each of the two audit technology groups. Cochran's test for homogeneity of variance will be applied to the sums of squares for the structured firms and, separately, for the unstructured firms. Table 5 summarizes the primary analyses related to firm effects.

Audit Technology Effects

If firm differences are detected with respect to location and/or variation, audit technology will be examined to explore how well firm
Table 5
Analyses of Firm Effects

Comparisons To Be Made  | Data To Be Used  | Primary Method
(Notation per Figure 7)  |  |  

I. Location:

A. Overall (H2.1)
   \[ T_1, T_2, \ldots, T_6 \]
   Materiality scores standardized across all 72 subjects
   One-way ANOVA across all 6 firms

B. Firm affiliation, by level of responsibility
   \[ A_1, A_2, \ldots, A_6 \]
   \[ B_1, B_2, \ldots, B_6 \]
   \[ C_1, C_2, \ldots, C_6 \]
   Same as above, but partitioned by level of responsibility
   One-way ANOVA for each level

C. Firm affiliation, by audit technology
   \[ T_1, T_2, T_3 \]
   \[ T_4, T_5, T_6 \]
   Same as above, but partitioned by audit technology
   One-way ANOVA for each audit technology

II. Variation:

A. Overall (H2.2)
   Same as I(A) above
   Sums of squares (SS's) associated with each firm
   Cochran's test for homogeneity of variance

B. Firm affiliation, by level of responsibility
   Same as I(B) above
   SS's associated with each level of responsibility within each firm
   Cochran's test for homogeneity of variance

C. Firm affiliation, by audit technology
   Same as I(C) above
   SS's associated with each firm, partitioned by audit technology
   Cochran's test for homogeneity of variance
effects correspond with differences in audit technology. Audit technology effects will be tested similarly to the firm effects described above. The notion of audit technology is closely related to that of firm affiliation; whereas firm affiliation categorizes each subject into one of six firms, audit technology partitions the firm affiliations into two groups—the structured and unstructured firms. As discussed above with respect to firm effects, the planned analyses with respect to audit technology reflect the expectation that no differences will be found with respect to level of responsibility. In terms of Figure 7, the technology effects with respect to location and variation will be tested by comparisons of $T_s$ and $T_u$.

In order to obtain insights with respect to technology and experience-related effects, one further supplemental comparison will be made. The two different audit technologies will be examined, level-by-level, with respect to location and variation. The purpose of this comparison is to explore the pervasiveness of any audit technology effects across the hierarchy of the firms. In terms of Figure 7, the following comparisons will be made: $A_s$ and $A_u$; $B_s$ and $B_u$; and $C_s$ and $C_u$.

**Location**

To test H3.1 (regarding audit technology effects with respect to location), a one-way ANOVA will be conducted which distinguishes the subjects affiliated with structured versus unstructured firms.
Firms #1-3 constitute the structured firms, whereas firms #4-6 constitute the unstructured firms.

Each level of responsibility will be examined separately to identify whether the structured and unstructured firms differ. Therefore, three separate one-way ANOVAs dealing with audit technology will be conducted. This supplemental comparison is intended to identify whether the different audit technologies have an effect on any or all of the levels of responsibility.

**Variation**

To test H3.2 (regarding audit technology effects with respect to variation), Cochran's test for homogeneity of variance will be applied to the sums of squares associated with the structured and unstructured firms. A supplemental analysis explores each level of responsibility separately for differences in variability associated with audit technology. The sums of squares for each level of responsibility will be calculated within each audit technology group. Cochran's test for homogeneity of variance will be applied to those sums of squares, on a level-by-level basis. This analysis will provide insights as to the consistency of any overall audit technology effect across the three levels of responsibility. Table 6 summarizes the primary analyses related to audit technology effects.
### Table 6

Analyses of Audit Technology

<table>
<thead>
<tr>
<th>Comparisons To Be Made (Notation per Figure 7)</th>
<th>Data To Be Used</th>
<th>Primary Method</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. Location:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Overall (H3.1)</td>
<td>Materiality scores standardized across all 72 subjects</td>
<td>One-way ANOVA with both audit technologies</td>
</tr>
<tr>
<td>$T_s, T_u$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Audit technology, by level of responsibility</td>
<td>Same as above, but partitioned by level of responsibility</td>
<td>One-way ANOVA with each level</td>
</tr>
<tr>
<td>$A_s, A_u$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B_s, B_u$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C_s, C_u$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>II. Variation:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Overall (H3.2)</td>
<td>Sums of squares (SS's) associated with each audit technology</td>
<td>Cochran's test for homogeneity of variance</td>
</tr>
<tr>
<td>Same as I(A) above</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Audit technology, by level of responsibility</td>
<td>SS's associated with each level of responsibility within each audit technology group</td>
<td>Cochran's test for homogeneity of variance</td>
</tr>
<tr>
<td>Same as I(B) above</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Risk Attitude

The notion of risk attitude is concerned with auditors' relative sensitivities to Type I errors (i.e., issuing a qualified opinion on financial statements which do not contain material misstatements) versus Type II errors (i.e., issuing an unqualified opinion on financial statements which do contain material misstatements). As hypothesized above, a difference among auditors in their attitudes about the implicit tradeoff of Type I and Type II errors (H4.1) has implications with respect to their relative materiality thresholds (H4.2). Subjects' attitudes toward Type I and Type II errors will be measured by means of a questionnaire which adopts an expectancy-va(lence approach. As described above and as represented by Figure 3, the attitude measures are organized according to four cells--representing each combination of a simple action set (issue a qualified opinion; issue an unqualified opinion) and a simple state set (the financial statements do not contain any material misstatements; the financial statements do contain material misstatements).

Cells #1 and #4 of Figure 3 represent Type I and Type II errors, respectively. Each cell specifies seven outcomes that might be associated with the particular cell. (Table 7 identifies the seven items specified for each cell.) The questionnaire elicits subjects' views of the likelihood ("belief") associated with each outcome for
Table 7
Items Encompassed by Risk Attitude Measure
(By Cell of Figure 3)

I. **Type I Error (Cell #1)**
   * a. result in losing the firm as a client
   * b. result in litigation by client against my accounting firm
   * c. damage my accounting firm's reputation in the business community
   * d. result in deteriorated relations with the client
   * e. damage my own standing within my accounting firm
   * f. demonstrate auditing negligence
   * g. damage the reputation of the client in the business community

II. **Type II Error (Cell #4)**
   * a. result in litigation by affected parties against my accounting firm
   * b. demonstrate auditing negligence
   * c. result in losing the firm as a client
   * d. damage my accounting firm's reputation in the business community
   * e. damage my own standing within my accounting firm
   * f. result in deteriorated relations with the client
   * g. damage the reputation of the client in the business community

III. "Correct" Qualification (Cell #2)
   * a. ... increase the credibility of the audit report to the public
   * b. enhance my accounting firm's reputation in the business community
   * c. result in losing the firm as a client
   * d. enhance my own standing within my accounting firm
   * e. damage the reputation of the client in the business community
   * f. result in deteriorated relations with the client
   * g. avoid litigation against my accounting firm

IV. "Correct" Ungualification (Cell #3)
   * a. fulfill my accounting firm's responsibility to the public
   * b. result in stronger relations with the client
   * c. enhance my accounting firm's reputation in the business community
   * d. avoid the undue concern ...
   * e. enhance my own standing within my accounting firm
   * f. avoid litigation against my accounting firm
   * g. avoid losing the firm as a client

Note: See Appendix A for the Attitude Measurement Questionnaire
   * Item taken directly from Kida [1980]
each cell along with the importance ("evaluation") of the outcome. The product of the belief and evaluation, termed the "expectancy-valence," will be calculated for each outcome within each cell. The expectancy-valences for the outcomes will be summed within the cells for each participant to obtain a measure of each individual's concern for Type I (i.e., cell #1) and Type II (i.e., cell #4) errors in an absolute sense.

The relevant notion here is the relative sensitivity of an individual to Type I and Type II errors. In other words, it is the individual's tradeoff of Type I versus Type II errors that is of interest. Thus, a single measure of each subject's risk attitude will be calculated as the expectancy-valence measure associated with Type II errors (i.e., cell #4) divided by the measure associated with Type I errors (i.e., cell #1). Individuals who are relatively more concerned with, for example, litigation by financial statement users (associated with Type II errors) would have larger risk attitude measures than individuals who are relatively more concerned with, for example, practice development (associated with Type I errors).

To construct a reasonably powerful test, subjects will be partitioned into the upper, middle, and lower one-thirds, based on the extremeness of their standardized materiality scores. To test H4.1, the subjects comprising the upper one-third will be compared to the subjects comprising the lower one-third for differences in their risk attitudes; a one-way ANOVA will be run on the risk attitude measures.
In order to gain insights related to differing risk attitudes, additional analyses will be conducted at a lower level of aggregation. Similar one-way ANOVAs will be conducted, comparing the subjects in the upper and lower one-thirds for differences in their expectancy-valence measures for cells #1 (related to Type I errors) and #4 (related to Type II errors) from Figure 3, and the individual items reflected by the attitude measurement questionnaire (in terms of the individual belief statements, the evaluation statements, and the related expectancy-valences) from Table 7.26 (There is support for this upper one-third versus lower one-third type of comparison in the accounting literature. See, for example, Kida [1980] and Hopwood and Schaefer [1985].)

In H4.2, auditors' planning stage materiality thresholds are hypothesized to relate to their attitudes toward trading-off Type I and Type II errors. (As depicted by Figure 4, the notion of risk attitude in this study may be viewed as an intervening variable which influences an auditor's materiality judgments across the whole set of cues. Specifically, individuals who are relatively more sensitive to Type II errors (i.e., have larger measures of risk attitude as described above) are expected to have relatively lower materiality thresholds; and individuals who are relatively more sensitive to Type I errors (i.e., have smaller measures of risk attitude as described above) are expected to have relatively higher materiality thresholds. To test H4.2, the risk attitude measures will be correlated with the
standardized materiality scores for the subjects in the aggregate, using the parametric Pearson product-moment correlation. As suggested by the above discussion, a negative correlation is expected.

Additional Analyses

Finally, five additional types of analyses will be conducted. First, individual ANOVA models will be constructed for each subject's judgments. The purpose is to identify any apparent differences among the relevant groupings of subjects (in terms of level of responsibility, firm affiliation, and/or audit technology) with respect to the structural forms of their models. Second, business risk will be examined as a possible confounding factor. Third, test-retest consistency will be examined based on the four repeated cases. Fourth, order effects will be investigated for differences as to location and/or variation between the two orders of case presentation. Fifth, several (ad hoc) comparisons will be made to identify whether the income statement materiality thresholds are less than the balance sheet materiality thresholds, on average as expected (in terms of level of responsibility, firm affiliation, and audit technology).

Individual ANOVA Models

Individual ANOVA models will be calculated to model each participant's planning stage materiality judgments, separately for the
income statement and the balance sheet, for the purpose of assessing the significance of the independent variables. Due to the few financial statement cues provided (e.g., two are apparently directly applicable to the income statement materiality judgment and only one is apparently directly applicable to the balance sheet materiality judgment), the explanatory power of the models for each individual (measured by $R^2$) should be quite high. These ANOVA models will serve as a basis for classifying and comparing the materiality judgments of the subjects as to the structural form of their models (e.g., whether income or sales or both are significant variables in the income statement materiality judgment).

**Business Risk Assessments**

As indicated by Figure 5, subjects will be asked to identify the business risk (also known as "exposure") on a seven-point scale (where "1" identifies "very low" risk and "7" identifies "very high" risk) which they attribute to the financial statement variables for each case. Business risk was regarded as a potential confounding factor. If subjects differ in their perceptions of the auditor's business risk, their materiality judgments may be affected. Higher exposure to the auditor is believed to result in more conservative (i.e., lower) materiality thresholds.

Since this dissertation focuses primarily on systematic differences among auditors in their planning stage materiality
judgments, the examination of business risk is conducted with respect to level of responsibility, firm affiliation, and audit technology. To ascertain whether subjects differ in their business risk assessments, separate one-way ANOVAs will be run on a case-by-case basis with respect to level of responsibility, firm affiliation, and audit technology.

**Test-retest Consistency**

Test-retest consistency will be examined by repeating four of the cases at the end of the judgment task; the judgments on these repeated cases will be compared with the original judgments for the same cases. If subjects are found to be inconsistent between the original and repeated cases, then their original responses may be suspect. Such results would suggest that subjects were either unwilling/unable to provide meaningful responses to the experimental materials or that they were inattentive to the experimental task.

The analyses with respect to test-retest consistency will be performed on a case-by-case basis for the four repeated cases. First, the responses to the repeated cases will be correlated with those on the original cases to ascertain the degree to which the responses move in the same direction. This correlation will merely identify whether relatively large responses on the original cases are paired with relatively large responses on the repeated cases. Second, for each of the repeated cases, a paired t-test will be calculated to ascertain
whether the responses on the repeated case are significantly different, in the aggregate, from those on the original cases. Finally, the differences between the original and repeated cases will be calculated for each subject; and separate one-way ANOVAs will be run to ascertain whether these differences are systematically related to level of responsibility, firm affiliation, and audit technology.

**Order Effects**

Order effects will be tested to investigate whether the order in which the cases were presented affected subjects' responses, in terms of location and/or variation. To the extent an order effect is detected, the responses may be viewed as artifacts of the experiment.

Two different orders were presented. The 24 unique combinations of the independent variables were initially presented in a randomized order (constituting order #1). This initial order was then reversed (constituting order #2). (Four repeat cases were included for both orders; two of the repeat cases were from the first half of the cases and the other two were from the second half of the cases. The same cases were repeated for both orders.)

Since each participating firm provided twelve subjects, the booklets were numbered consecutively, by firm, in order to provide some degree of control over follow-up, etc., but maintain the individuals' anonymity. The even-numbered booklets were the initial order and the odd-numbered booklets were the alternate order. The two
orders were systematically varied over level of responsibility, firm affiliation, and audit technology. Half of the subjects (at each level within each firm) received order #1 and the others received order #2.

A one-way ANOVA will be used to identify any difference as to location between the two orders of presentation, based on the standardized scores of the subjects receiving each order. Cochran's test for homogeneity of variance will be used to identify any difference as to variation between the two orders of presentation, based on the sums of squares associated with the subjects receiving each order.27

**Income Statement Versus Balance Sheet Judgments**

It was stated above that the income statement materiality judgments are of primary interest in this study. The balance sheet materiality judgments were included because SAS No. 47 indicates that such judgments may be made at the aggregate balance sheet level. Consequently, the balance sheet judgments were included for exploratory purposes. The SAS indicates that, as a practical matter, a single "preliminary judgment about materiality" should be determined. It is identified as the smallest of the amounts material to any one of the financial statements. Prior research has consistently reported the dominance of net income (or income before taxes) in auditors' evaluation stage materiality judgments. In
addition, Zuber, et al. [1983] suggest that, in practice, income statement considerations generally drive the "preliminary judgment about materiality." Accordingly, it is expected that the income statement judgments will be less than the balance sheet judgments for most subjects, across most cases.

For each level of responsibility, the average income statement and balance sheet materiality judgments will be calculated using subjects' original responses (i.e., their "raw" or untransformed materiality judgments). The average income statement and balance sheet judgments will be compared for each level of responsibility across the 24 different cases.

For each firm, the average income statement and balance sheet materiality judgments will be similarly calculated. Again, the average income statement and balance sheet judgments will be compared for each firm across the 24 different cases.

Similarly for each audit technology group, the average income statement and balance sheet materiality judgments will be calculated. Again, the average income statement and balance sheet judgments will be compared for each audit technology group across the 24 different cases.

Although these comparisons are ad hoc, they will provide some retrospective evidence as to the appropriateness of focusing on the income statement materiality judgments in the context of studying auditors' "preliminary judgments about materiality" at the aggregate
financial statement level. The income statement judgments are expected to be less than the balance sheet judgments (hence, they will drive the auditors' "preliminary judgments about materiality") for most of the auditors, for most of the cases. Since this dissertation emphasizes systematic differences in terms of level of responsibility, firm affiliation, and audit technology, these ad hoc comparisons are made at those levels.

**Underlying Statistical Assumptions**

This section identifies the principal assumptions underlying these statistical methods to be used in the data analysis and discusses the reasons that those assumptions are not believed to constitute problems for the validity of the planned analyses. As described above, the judgments of the participants will be examined for systematic differences with respect to location and variation. The primary statistical method to be used to identify differences as to location is analysis of variance (ANOVA), a parametric method. Tukey's all pairwise multiple comparison test will be used to identify which levels of responsibility and/or which firms differ, when the related ANOVA is significant.

There are three principal assumptions underlying analysis of variance: (1) normally distributed populations; (2) homogeneous error variances associated with the populations; and (3) independent error
components, both within and between comparison groups (Keppel [1982, pp. 85-87]). Related to the first assumption, the members of the comparison groups are further assumed to be randomly selected from the relevant populations. Keppel (p. 86) points out that, when the sample sizes are equal and reasonably large, only the most severe violations of the normality assumption are problematic (and then only with respect to F-statistics that fall near the critical values). Since all of the cell sizes for all comparisons will be equal by design, this assumption should not be consequential. In addition, as noted, when the parametric tests are significant with respect to the primary hypotheses (i.e., H1.1, H1.2, H2.1, H2.2, H3.1, and H3.2), nonparametric Wilcoxon rank sum tests will be used. These nonparametric tests will provide evidence on the importance of the normality assumption. Similar results with a nonparametric method would provide strong corroboration and establish that the normality assumption is not critical to the parametric tests, since the nonparametric method does not depend on distributional assumptions.

The second assumption involves the equality of the variances of the different populations under study. Again, Keppel points out that the effects of heterogeneity are substantially mitigated by equal cell sizes. While noting that tests of the homogeneity assumption are generally omitted due to the robustness of ANOVA, Keppel (p. 87) suggests that caution should be used when there is extreme heterogeneity (e.g., a ten-fold difference between the largest and
smallest variances), especially when the observed F-value falls near 
the critical F-value. Once again, this assumption should not be 
consequential, since the cell sizes will be equal by design for all 
comparisons. In addition, as noted, when the parametric tests are 
significant with respect to the primary hypotheses (i.e., H1.1, H1.2, 
H2.1, H2.2, H3.1, and H3.2), nonparametric Wilcoxon rank sum tests 
will be used. These nonparametric tests will provide evidence on the 
importance of the homogeneity of variance assumption.

The third assumption involves the independence of the error 
components of the ANOVA model. The error terms are assumed to be 
independent, both within and between comparison groups. Keppel (p.87) 
paraphrases this assumption: "...this is just another way of saying 
that systematic biases must not be present in the assignment of 
subjects to conditions." Independence within comparison groups 
requires that each observation be unrelated to all other observations. 
Independence between comparison groups requires that there be no 
confounding effects which would distort the inferences with respect to 
the relationship between the independent and dependent variables. 
Random selection of subjects is frequently used to achieve 
independence.

The standardized scores for each subject should be independent 
measures of each individual's relative extremeness in the materiality 
judgments examined. One way in which the independence of the data 
could be jeopardized would be by subjects working on the experiment
collectively. However, they were specifically instructed to work alone. Further, the independence of the observations could be adversely affected by confounding influences. It is particularly important to separate the effects carefully with respect to level of responsibility, firm affiliation, and audit technology.

When the ANOVA identifies a significant mean difference (with respect to levels of responsibility and/or firm affiliation), the specific nature of the difference will be explored using Tukey's multiple comparison test. The Tukey test assumes that the observations represent random samples from normally distributed populations, homogeneity of variances, and approximately equal sample sizes (Kirk [1982, p. 116]). These are similar to the assumptions identified above with respect to analysis of variance.

The primary statistical method to be used to identify differences as to variation is Cochran's test for homogeneity of variance, a parametric method. The major assumption associated with Cochran's test for homogeneity of variance is that the populations of interest are normally distributed. According to Kirk (p. 78), Cochran's test is somewhat sensitive to departures from normality.

Therefore, another parametric test will be used, in addition to Cochran's test, to identify between-firm differences in variation. F-tests, reflecting all possible pairwise comparisons among the firms based on ratios of the firms' sums of squares, will be used to identify differences in variation at the firm level. The assumptions
associated with the F-test are as follows: observations are random samples from normally distributed populations; the numerator and denominator of the F ratio are estimates of the same population variance; and the numerator and denominator of the F ratio are independent (Kirk [1982, pp. 55-57]). Kirk (p. 77) points out that the F-test is quite robust with respect to violations of the normality assumption; in addition, it is also robust with respect to violations of the heterogeneity assumption when the cell sizes are equal. (By design, the cell sizes for all comparisons are equal.) Further, nonparametric Wilcoxon rank sum tests will be used to cross-check the results when significance regarding the primary hypotheses is established by the parametric methods.

Administration of the Experiment

During late autumn of 1985, a "contact person" (a partner) at seven of the Big Eight firms' Columbus offices was asked for assistance in conducting the experiment. Arthur Young was not asked to participate due to the small size of its Columbus office. Ernst & Whinney declined to participate because they were too busy at the time. The other six Big Eight firms agreed to participate and help in any way that they could.

Prior to conducting a pilot study, the contact person at each of the firms expressing a willingness to participate in the study was
asked to review the experimental materials. The purpose of that review was to identify any troublesome terminology and to verify that the experimental task was reasonable for those firms. Some changes were made to the background materials as a result of that review, but they were minor and primarily editorial in nature. Without exception, the reviewers felt that the participants from their firms had sufficient information to provide meaningful responses.

A pilot study was conducted during the winter of 1986 for the purpose of verifying that the experimental materials were appropriately constructed. Fourteen subjects participated, and all were from the Columbus office of Coopers & Lybrand. That office was selected because it was the largest office of the Big Eight firms in Columbus, and therefore the pilot study would not interfere with procuring an adequate number of participants for the actual experiment. The participants were explicitly asked not to discuss the experiment with their colleagues, since some of their associates would be asked to participate in the next phase of the study.

As a result of the pilot study a number of changes were made. For example, the pilot study consisted of 36 cases—representing a \(3^2 \times 4\) experimental design with three levels each for total assets and sales, and four levels for income (loss) before taxes. Since the main effect for sales was only significant for three of the participants, one level was dropped in order to reduce the number of cases to a more manageable number. In addition, the pilot study contained two
between-subjects variables (one distinguishing public versus private companies; and the other representing high client-paid fees versus low client-paid fees). The nature of the materiality judgments under study was apparently so "noisy" that the between-subjects variables were likely to contribute additional noise without offering much explanatory power. Thus, the focus of the study became the differences among auditors related to differences in their materiality judgments, rather than focusing on situational factors and their impact on auditors' materiality judgments. The further study of certain situational factors will be among the topics for future research.

In May of 1986 the experimental materials were delivered to the contact person at each firm to oversee distribution to the participants. (In only one case were the materials provided directly to the participants by the researcher. That involved Price Waterhouse, which had already arranged a staff meeting for the entire Columbus office and it was convenient for the firm for the researcher to distribute the materials directly to the chosen participants.) The completed booklets were returned to a designated person within each firm; the researcher then picked up these materials periodically as they became available.

Although deadlines were agreed upon during June, a number of individuals did not complete the experiment within the planned time-frame and some follow-up was necessary. The booklets were kept track
of by number, by firm. The booklets were numbered from one to twelve and a control listing was given to each firm to account for the materials. Thus, follow-up involved identifying the booklet numbers which were still outstanding, since the participants were anonymous to the researcher. Almost all of the booklets were returned and the data loaded by the end of July, although three partners of one firm did not return the completed booklets until the end of September, 1986.
CHAPTER III FOOTNOTES

1. As referenced in Chapter II, Ashton [1982a, p.74] notes that audit decision making has been evaluated with respect to six criteria: (1) accuracy; (2) normativeness; (3) consistency with professional pronouncements; (4) consensus; (5) self-insight; and (6) judgment stability. The first three criteria are not applicable to this study. There is no available criterion variable by which to assess optimality; and, due to the generality of the auditing standards (i.e., SAS No. 47), inconsistencies with the pronouncements are not applicable to an empirical study such as this. (See Cushing and Loebbecke [1983] for an evaluation of 12 national firms' policies relative to the SAS's.)

Ashton notes that consensus, self-insight, and stability are usually used when accuracy and normativeness cannot be used. Since this study focuses on comparisons among groups of auditors, rather than within individual auditors, consensus is of primary importance. Prior research (e.g., Messier [1983]; and Krogstad, et al. [1984]) has reported relatively high self-insight and stability on materiality tasks.

2. In addition to preceding the final materiality judgment, one could argue that quantifying the preliminary judgment about materiality is a more complicated task. That is, it may be more difficult to specify a dollar amount as a basis for audit planning (i.e., quantifying the preliminary judgment about materiality) than, for example, to make the dichotomous judgment that a specific item under consideration warrants separate disclosure, or that an account warrants adjustment.

On the other hand, one might argue that the evaluation stage judgment is more complicated. SAS No. 47 indicates that the evaluation stage judgment requires consideration of both quantitative and qualitative factors; the planning stage judgment primarily involves just the quantitative factors. Thus, one could argue that the judgment at the evaluation stage is more complicated in the sense of involving a larger set of variables which might enter into the judgment. As suggested by Woolsey [1973], there may be a "border zone" such that below some threshold there is a presumption of "immateriality;" above some other threshold there is a presumption of "materiality." Intuitively, the endpoints of this border zone would apparently hinge on quantitative factors. Within the border zone, the determination of materiality may hinge on the qualitative factors. Only within this "grey area" would the evaluation stage materiality
judgment seem to depend upon a larger cue set, and, in that sense, be a more complicated judgment.

3. The emphasis on cost minimization is consistent with a competitive environment for audit services. Anecdotal evidence (i.e., comments by practicing auditors) suggests that the market for audit services is indeed highly competitive. Simon [1985], Dopuch and Simunic [1980], and Simunic [1980] provide empirical evidence of a competitive market for audit services.

For a further discussion of Type I and Type II errors, see footnote #2 in Chapter II.

4. "Hypothesis-guessing" may lead to biased responses by the subjects if the research questions of interest are transparent. For example, a subject who identifies the focus of the research may wish to be a "good subject" and provide responses perceived to be what the researcher "wants." To mitigate this concern, it is important that the research issues not be transparent.

"Evaluation apprehension" refers to biased responses as a consequence of subjects' concerns about performance assessment, etc. As a result, subjects may alter their "real" responses in order to be evaluated more favorably. To mitigate this concern, one may endeavor to persuade the participants that their responses will be anonymous and that they will not be evaluated based on their responses.

"Experimenter expectancies" refers to bias in the data attributable to the perceptions of the researcher. Such bias may be communicated either in personal contact with the participants or in the experimental materials. To mitigate this concern, one may obtain independent review of the experimental materials by others not having similar expectations.

5. On average, across the 72 participants, approximately 2.25 hours were required to complete the experiment and the related questionnaires. The standard deviation was approximately .83 hour. The range was 1-4 hours.

6. One possibility was to contact Arthur Young for the purpose of gaining the participation of auditors from one or more other offices. However, to do so would interject the possibility of a geographical influence, although that would be expected to be a rather minor issue. The six participating firms, involving 72 practitioners, was deemed to be an adequate sample size, and a desirable symmetry between the three very structured and three less structured firms was achieved without Arthur Young's participation.
7. The partition of the firms into two groups--i.e., highly structured ("structured") and relatively less structured ("unstructured")--relied on the ordering identified by Kinney [1986]. The set of highly structured firms was comprised of Deloitte, Haskins, & Sells; Peat, Marwick, Mitchell; and Touche Ross.

For purposes of this study, Arthur Andersen was combined with Coopers & Lybrand and Price Waterhouse to comprise the set of relatively less structured firms. Kinney classified Arthur Andersen as an "intermediate" firm. Review of the firms' audit manuals with respect to planning stage materiality judgments suggested that these three firms were similar in their treatment of planning stage materiality judgments, at the time the experiment was conducted.

Consequently, the classification of Arthur Andersen as "unstructured" was slightly subjective and reflected a somewhat arbitrary cut-off along the continuum of unstructured to structured audit technology. In general, there can be no guarantee that, within either of the two audit technology groups, the firms reflect identical degrees of structure.

8. The information contained in Appendix A specifies the role as "partner." The materials provided to each subject specified the appropriate role in terms of "senior," "manager," and "partner." With the exception of identifying the role conditional on each subject's level of responsibility, the information provided to each subject was identical.

9. SAS No. 22 (AICPA [1978, paragraph 3]) identifies a number of factors that the auditor should consider in audit planning:
   a. the entity's business and industry environments (including competitive conditions);
   b. the entity's accounting policies;
   c. whether reliance on internal controls is anticipated;
   d. preliminary judgments about materiality levels ..." (i.e., the dependent variables of interest in this study);
   e. statement items likely to require adjustment;
   f. related party transactions;
   g. the nature of the reports expected to be issued.

   These factors, among others, were explicitly addressed in the background information regarding the hypothetical client (Appendix A).

10. As indicated by Figure 5 and as explained in Appendix A (page 1), subjects were also asked to identify the business risk implied by the given financial statement amounts. (Business risk is defined on page 1 of Appendix A, and is often referred to as "exposure" in practice.) Risk was assessed on a seven-point scale from very low to very high.
It is expected that, all else constant, an auditor will establish lower materiality thresholds in instances of perceived higher risk. Therefore, exposure was viewed as a potential confounding factor. To the extent that the subjects differed systematically (particularly by level of responsibility or by firm) as to their perceptions of these risks, their materiality judgments could be systematically affected.

To assess whether such a confounding may have occurred, one-way ANOVA's were conducted, case-by-case, to identify whether there were systematic differences, first, among the three levels of responsibility and, second, among the six participating firms.

In addition, as described in Appendix A (page 1), for selected cases (specifically, half of the cases), subjects were asked to make audit hour estimates. The audit hour responses are strictly for future purposes and are only tangential to the dissertation.

11. These scales were similar to those reported by Kida [1980]--he indicated that he measured both the belief and evaluation statements on seven-point scales. However, regardless of whether an outcome was favorable or unfavorable, the seven-point evaluation scale represented three degrees of "bad" (1, 2, and 3), one response that represented a neutral view (4), and three degrees of "good" (5, 6, and 7). Consequently, as a practical matter, Kida's evaluation scale was really a four-point scale, at best, from 1-4 for unfavorable outcomes and 4-7 for favorable outcomes.

12. Both Peat, Marwick, Mitchell and Touche Ross have developed firm-wide guidelines in a tabular format to assist their auditors in quantifying the preliminary judgments about materiality (e.g., Peat Marwick calls this construct "guage"). These tabular approaches both consider sales and total assets as the relevant variables in quantifying the preliminary judgment about materiality. Robert K. Elliott, a partner in the national office of Peat Marwick, disclosed his firm's tabular approach in a recent Journal of Accountancy (Godick [1983]).

13. In this study, the coefficient of correlation between the levels of responsibility and years as an auditor was .84.

14. To ascertain whether the seniors, managers, and partners differ similarly across the six participating firms, a two-way ANOVA (encompassing the main effects for level of responsibility and firms, along with the related level-firm interaction) will be run. The two-way ANOVA is a less general approach than the analysis of level of responsibility effects, on a firm-by-firm-basis. The two-way ANOVA assumes that the main effect associated with level of responsibility is uniformly observed across the firms.
15. According to Keppel [1982, p. 165], Tukey's multiple comparison test is generally preferred to a number of available alternative tests for the purpose of making all pairwise comparisons among means. Its advantages include relative simplicity, favorable power characteristics, and appropriate control of alpha in view of the number of comparisons made.

16. As indicated by Kirk [1982, pp. 57-58], the total variance for np scores in an experiment is given by:

\[ \sum_{i=1}^{n} \sum_{j=1}^{p} (Y_{ij} - \bar{Y})^2 / np \]

The numerator above is referred to as "the sum of squares." When the cell sizes (i.e., p) for n comparison groups are equal, the ratio of their variances is equal to the ratio of their sums of squares.

17. The available tables [e.g., Kirk [1982, p. 829]] provide the critical values for Cochran's test for homogeneity of variance at the .05 and .01 levels. However, these tables are not amenable to identifying particular p-values. In order to use Fisher's method for combining tests, the specific p-values must be derived. Cochran [1941] demonstrates how such p-values can be calculated. For three variances (i.e., for the seniors, managers, and partners), the following formula calculates the p-value for one of the groups:

\[ p = 3 \left[ 1 - 6.5625 \left( \frac{2 z^{3/2}}{3} - \frac{4 z^{5/2}}{5} + \frac{2 z^{7/2}}{7} \right) \right] \]

where \( z = \max \frac{1}{\sum_{j=1}^{3} s_j^2} \sum_{i \leq j < 3} s_i^2 \)

This formula is valid for \( z \geq .5 \); for \( z < .5 \), using \( z = .5 \) gives a lower bound for the p-value.

This approach will be used to infer the specific p-value from Cochran's test for homogeneity of variance for each level of responsibility. Then, Fisher's method will be used to combine the separate p-values into a single test of significance.

18. If significant effects for level of responsibility are found, then firm differences will be evaluated, separately, for each level of responsibility. Fisher's method for combined experiments would be used to combine the p-values associated with each level of responsibility into an overall test of significance regarding firm affiliation. These procedures would be similar to those previously described with respect to level of responsibility in terms of location and variation.
19. In the event that the ANOVA detects a significant firm difference, the nonparametric Wilcoxon rank sum test (Hollander and Wolfe [1973, pp. 27-28] will be used to cross-check the parametric results. Corroboration would be viewed as strong evidence that the results of the ANOVA do not depend upon the distributional assumptions of ANOVA. (These assumptions are discussed near the end of Chapter III.)

Specifically, all pairwise comparisons will be made between the six firms (hence, there are 15 comparisons) using the Wilcoxon rank sum test. In view of the number of pairwise comparisons to be made, a conservative alpha level is required (e.g., .05/15).

20. An additional test for firm differences as to location will be conducted, since the ANOVA's related to firm effects could be affected by substantial heterogeneity of variance. (The assumptions underlying ANOVA are discussed near the end of the chapter.) Therefore, all possible (fifteen) pairwise comparisons will be made among the six firms using two-sample t-tests (Miller [1986, p. 55]). (In view of the number of such comparisons, the level of significance will be established conservatively—e.g., approximately .05/15.)

Of the resulting fifteen pairwise comparisons, three are comparisons between structured firms, three are comparisons between unstructured firms, and nine are comparisons between one structured and one unstructured firm. Consequently, these t-tests also relate to the analysis of firm differences within audit technology groups and to the analysis of audit technology effects (i.e., regarding differences between firms having different audit technologies).

21. If firm differences in variability are shown using the parametric Cochran's test for homogeneity of variance, a nonparametric Wilcoxon rank sum test will be used as a cross-check. Corroboration using a nonparametric approach would demonstrate that the assumptions underlying the parametric test (described near the end of Chapter III) are not driving the results.

Specifically, the variance within each level of responsibility for each firm will be calculated (thus, there will be 3 observations for each firm). These observations will be used to make all pairwise comparisons among the six firms using the Wilcoxon rank sum procedure.

22. An additional test for firm differences as to variation will be conducted, since Cochran's test for homogeneity of variance could be affected by substantial non-normality. (The assumptions underlying ANOVA are discussed near the end of the chapter.) Therefore, all possible (fifteen) pairwise comparisons will be made among the six firms using ratios based on the firms' sums of squares. The ratio of
two firms' sums of squares (or variances, as explained in footnote #16 above) constitutes an F-test. This represents a specific condition of the more general Hartley's test for homogeneity of variance, since, in this case, only two variances are involved in the comparison (Kirk [1982, pp. 54 and 78]). (In view of the number of such comparisons, the level of significance will be established conservatively--i.e., approximately .05/15.)

Of the resulting fifteen pairwise comparisons, three are comparisons between structured firms, three are comparisons between unstructured firms, and nine are comparisons between one structured and one unstructured firm. Consequently, these F-tests also relate to the analysis of firm differences within audit technology groups and to the analysis of audit technology effects (i.e., regarding differences between firms having different audit technologies).

23. If significant experience-related effects are found, then audit technology differences will be evaluated level-by-level for each level of responsibility. Fisher's method for combined experiments would be used to combine the p-values associated with each level of responsibility into an overall test of significance regarding audit technology. These procedures would be similar to those previously described with respect to level of responsibility regarding location and variation.

24. In the event that the ANOVA detects a significant audit technology difference, the nonparametric Wilcoxon rank sum test (Hollander and Wolfe [1973, pp. 27-28]) will be used to cross-check the parametric results. (There would be 36 subjects in each technology group and, hence, the large sample approximation would be used.) Corroboration would be viewed as strong evidence that the results of the ANOVA do not depend upon the distributonal assumptions of ANOVA.

25. If audit technology differences in variability are shown using the parametric Cochran's test for homogeneity of variance, a nonparametric Wilcoxon rank sum test will be used as a cross-check. Corroboration using a nonparametric approach would demonstrate that the assumptions underlying the parametric test are not driving the results.

Specifically, the variance within each level of responsibility for each firm will be calculated (thus, there will be 9 observations for each audit technology). The Wilcoxon rank sum test will then be used to identify whether there is a difference in variability between the two technologies.

26. Additional one-way ANOVA's will be run with respect to subjects' risk attitudes (including the individual belief and evaluation
statements, and the expectancy-vaences for the individual items and for the cell totals) based on levels of responsibility, firm affiliations, and audit technologies.

27. In addition, a test for "drift" will be conducted to identify whether subjects tended to become stricter (or more lenient) as they proceeded through the 24 original cases. A ratio will be calculated, on a case-by-case-basis, where the numerator is the average materiality judgment (based on the "raw" or untransformed responses) for the subjects receiving order #1 and the denominator is the average for the subjects receiving order #2. Since order #2 was the reverse of order #1, a tendency to become stricter (more lenient) would be shown by these ratios decreasing (increasing) as the subjects completed the cases. Kendall's nonparametric test for independence (Hollander and Wolfe [1973, pp. 185-6]) will be used to identify whether there was such a "drift" as subjects progressed through the experiment.

28. The nonparametric test to be used to cross-check the parametric tests is the Wilcoxon rank sum test. Although the Wilcoxon rank sum test makes no assumptions about the underlying distribution of the data, the error terms are assumed to be unobservable random variables which are mutually independent and belong to the same continuous population (Hollander and Wolfe [1973, p. 67]).
CHAPTER IV -- RESULTS OF ANALYSES

Introduction

As described in Chapter III, each participant was asked to make two separate materiality judgments (one with respect to the income statement and the other with respect to the balance sheet) for each of the 24 unique combinations of the three independent variables--total assets, net sales, and income (loss) before taxes. Separate analyses were conducted on the income statement materiality judgments and on the balance sheet materiality judgments.

As discussed in Chapter II, prior research (substantially directed at the evaluation stage) has consistently reported net income (or income before taxes) to be the dominant consideration in the judgments studied. Further, in practice the auditor's planning stage materiality judgment is believed to be driven primarily by income statement considerations (e.g., see Zuber, et al. [1983, p. 43]). Accordingly, the income statement materiality judgment is deemed to be the dependent variable of principal interest in this study, and this chapter reports the analyses only with respect to these income statement judgments. However, since SAS No. 47 (paragraph 12) implies that the planning stage materiality judgment may be made appropriately from a balance sheet perspective, similar analyses were performed on the balance sheet materiality judgments for exploratory purposes. The
results of the analyses of the balance sheet judgments are reported in Appendix D.

A transformation was made to subjects' "raw" materiality judgments for two reasons. First, across the 24 cases, there appeared to be a conspicuous heterogeneity of variance, since some cases naturally yielded relatively large responses and other cases led to relatively low responses. Second, a single (independent) indicator of the extremeness of each individual's materiality thresholds was desired so that the discussion of the results would not have to be conditioned with respect to particular cases. To achieve these two purposes, the materiality judgments were standardized separately for the income statement and balance sheet judgments. On a case-by-case basis, the participants' responses were transformed to a distribution with mean zero and variance one. Consequently, an individual with an above-average judgment on a particular case would have a positive score, and an individual with a below-average judgment on that case would have a negative score. Each subject's standardized scores were summed across the 24 cases to provide an overall measure of the extremeness of the subject's materiality thresholds. Then, the standardized materiality scores were analyzed.

The research objectives of the dissertation were described in the previous chapter. This chapter reports the results of the analyses dealing with systematic differences in the auditors' planning stage materiality judgments with respect to (1) level of responsibility,
(2) firm affiliation, and (3) audit technology (i.e., "structure"), related to firm affiliation. These comparisons emphasize consensus. This investigation of consensus involves two dimensions--location (regarding central tendency) and variation (regarding dispersion).

Effects associated with level of responsibility were examined among the seniors, managers, and partners. No significant differences were identified either with respect to location or variation.

Differences as to location and variation were examined among the six participating Big Eight CPA firms. Significant firm differences were identified among the participants both with respect to location and variation.

Further analyses were conducted to identify any differences as to location and variation related to the firms' differing audit technologies. The firms were partitioned into two groups--the structured firms and the unstructured firms. Significant differences were identified between the structured and unstructured firms both with respect to location and variation. The participants affiliated with structured firms had higher materiality thresholds, but greater variability, relative to the participants from the unstructured firms.

In addition, this chapter includes discussions of the results of the analyses with respect to risk attitudes. Risk attitude was viewed as a potential intervening variable affecting auditors' materiality judgments. Auditors who were relatively more sensitive to Type I errors (e.g., practice development considerations) were expected to
establish relatively higher materiality thresholds; auditors who were relatively more sensitive to Type II errors (e.g., litigation by financial statement users) were expected to establish relatively lower materiality thresholds. However, the participants did not differ systematically in their risk attitude measures, nor were those risk attitude measures associated with systematic differences in the subjects' materiality thresholds.

Five additional types of analyses were conducted. Presented first are the remarks about individual models of the participants' materiality judgments. Conspicuous differences were noted in the structural forms of subjects' individual ANOVA models, related to cue utilization.

Second, business risk (i.e., exposure) was examined as a potential confounding factor, on a case-by-case basis. Subjects' assessments of business risk were not systematically related to differences in their materiality judgments.

Third, this chapter reports the analyses with respect to test-retest consistency and fourth, with respect to order effects. The analysis of test-retest consistency indicated that subjects were consistent in their responses to the original and repeated cases. No differences were found with respect to location or variation associated with order of presentation.

Fifth, some observations are offered, in hindsight, about the appropriateness of focusing the discussion on the income statement
judgments, rather than the balance sheet judgments. These ex post comparisons underscored the appropriateness of that emphasis.

This chapter concludes with comments about the statistical assumptions underlying the principal analyses. Reasons are given as to why those assumptions are not viewed as problematic.

Level of Responsibility Effects

As discussed in Chapters II and III, it is unclear whether auditors' preliminary judgments about materiality differ systematically among auditors of differing levels of responsibility (presumably related to experience). Hypotheses were stated with respect to overall level of responsibility effects, separately in terms of location and variation. H1.1 deals with differences as to location; and H1.2 deals with differences as to variation.

These analyses distinguish the responses of the seniors, managers, and partners. The demographic information that was obtained provided the following profile of the subjects' years of experience:

<table>
<thead>
<tr>
<th></th>
<th>Range</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seniors</td>
<td>2.0-6 years</td>
<td>3.9</td>
<td>1.3</td>
</tr>
<tr>
<td>Managers</td>
<td>4.5-14 years</td>
<td>7.9</td>
<td>2.1</td>
</tr>
<tr>
<td>Partners</td>
<td>9.5-40 years</td>
<td>17.6</td>
<td>6.3</td>
</tr>
</tbody>
</table>

Similar profiles were constructed separately for each firm and are reported in Table 8. Years of experience overlap with level of
Table 8

Years of Experience, By Level of Responsibility and Firm

<table>
<thead>
<tr>
<th>Firm #1</th>
<th>Range (years)</th>
<th>Average (years)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior</td>
<td>4-6</td>
<td>5.3</td>
<td>.83</td>
</tr>
<tr>
<td>Manager</td>
<td>7-11</td>
<td>8.3</td>
<td>1.64</td>
</tr>
<tr>
<td>Partner</td>
<td>14-40</td>
<td>23.1</td>
<td>10.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Firm #2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior</td>
<td>2-4</td>
<td>2.9</td>
<td>.89</td>
</tr>
<tr>
<td>Manager</td>
<td>6-10</td>
<td>7.8</td>
<td>1.79</td>
</tr>
<tr>
<td>Partner</td>
<td>9.5-19</td>
<td>14.4</td>
<td>3.38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Firm #3</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior</td>
<td>1.5-6</td>
<td>3.9</td>
<td>1.60</td>
</tr>
<tr>
<td>Manager</td>
<td>6-7</td>
<td>6.5</td>
<td>.50</td>
</tr>
<tr>
<td>Partner</td>
<td>13.5-21</td>
<td>16.6</td>
<td>2.72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Firm #4</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior</td>
<td>3-4</td>
<td>3.5</td>
<td>.50</td>
</tr>
<tr>
<td>Manager</td>
<td>8.5-14</td>
<td>10.3</td>
<td>2.19</td>
</tr>
<tr>
<td>Partner</td>
<td>13-27</td>
<td>18.4</td>
<td>5.35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Firm #5</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior</td>
<td>3-7</td>
<td>4.3</td>
<td>1.64</td>
</tr>
<tr>
<td>Manager</td>
<td>4.5-8</td>
<td>6.4</td>
<td>1.29</td>
</tr>
<tr>
<td>Partner</td>
<td>10-17</td>
<td>13.6</td>
<td>2.63</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Firm #6</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior</td>
<td>3-4</td>
<td>3.8</td>
<td>.43</td>
</tr>
<tr>
<td>Manager</td>
<td>7-11</td>
<td>8.0</td>
<td>1.73</td>
</tr>
<tr>
<td>Partner</td>
<td>14-24</td>
<td>19.5</td>
<td>3.57</td>
</tr>
</tbody>
</table>
responsibility in some instances, but are obviously correlated with level of responsibility. (As noted in Chapter III, the overall correlation coefficient between years of experience and level of responsibility was .84.)

As previously indicated, this dissertation investigates two distinct potential sources of systematic differences among the subjects (experience-related effects and firm-related effects.) Level of responsibility and firm effects were separated by analyzing level of responsibility effects within each firm; the results for each firm were then combined into an overall test of significance with respect to level of responsibility. In terms of Figure 7, the following comparisons were made: A₁, B₁, and C₁; ...; and A₆, B₆, and C₆.³

In order to further explore any possible differences associated with level of responsibility, an additional type of comparison was made, and the results are also reported below. Each of the three levels of responsibility was compared within the two audit technologies (structured and unstructured). In terms of Figure 7, the following comparisons were made: Aₛ, Bₛ, and Cₛ; and Aᵤ, Bᵤ, and Cᵤ. As indicated, formal hypotheses were not stated with respect to this supplemental comparison, since the purpose was to glean possible insights regarding effects associated with level of responsibility.

As described above, in order to obtain a single, independent measure of the extremeness of each subject's materiality judgments and to mitigate the substantial heterogeneity of variance with respect to
the individual cases, the responses for the seniors, managers, and partners were standardized on a case-by-case basis. For each case, the materiality judgments were transformed to a distribution with mean zero and variance one, across the 72 subjects. To get a measure of the extremeness of the individuals' judgments, each subject's standardized scores were summed across the 24 cases. Figure 8 plots these standardized materiality scores by level of responsibility (where levels 2, 3, and 4 identify the seniors, managers, and partners, respectively). These analyses are discussed first with respect to location and second with respect to variation.

Location

H1.1 deals with differences as to location among the seniors, managers, and partners. (Differences in location with respect to their standardized materiality scores would be interpreted as differences in their mean materiality thresholds.) As noted above, Figure 8 plots the standardized materiality scores by level of responsibility. Examination of that plot does not indicate any obvious differences in the means among the seniors, managers, and partners.

Six separate one-way ANOVAs were conducted to identify whether there were any detectable differences as to location among the seniors, managers, and partners of each participating firm. As summarized in Table 9, level of responsibility was significant
Figure 8

Standardized Scores for the Income Statement Materiality Judgments
Plotted by Level of Responsibility
Table 9

Standardized Materiality Scores and Averages
By Level of Responsibility and By Firm Affiliation

<table>
<thead>
<tr>
<th>Firm #1</th>
<th>Seniors</th>
<th>Managers</th>
<th>Partners</th>
<th>Average</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>17.37</td>
<td>32.60</td>
<td>24.39</td>
<td>24.79</td>
<td>.77</td>
</tr>
<tr>
<td>3</td>
<td>11.75</td>
<td>3.93</td>
<td>-5.30</td>
<td>3.46</td>
<td>.16</td>
</tr>
<tr>
<td>4</td>
<td>-13.94</td>
<td>-10.77</td>
<td>-14.15</td>
<td>-12.95</td>
<td>.62</td>
</tr>
<tr>
<td>5</td>
<td>-12.75</td>
<td>-8.88</td>
<td>-10.98</td>
<td>-10.87</td>
<td>.39</td>
</tr>
<tr>
<td>6</td>
<td>-15.68</td>
<td>-10.05</td>
<td>-4.92</td>
<td>-10.22</td>
<td>.06</td>
</tr>
<tr>
<td>Average</td>
<td>2.86</td>
<td>.71</td>
<td>-3.57</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p-value based on one-way ANOVA for level of responsibility calculated on a firm-by-firm basis.
(p ≤ .05) only for firm #2. In that instance, Tukey's multiple comparison test identified the seniors as having higher standardized scores (reflecting higher materiality thresholds) relative to the managers and partners.5

In order to make an overall inference regarding H1.1, the six separate p-values (obtained from the one-way ANOVAs with respect to level of responsibility on a firm-by-firm basis) were combined using Fisher's method for combining tests. An overall test statistic of approximately 21.86 resulted. This was not significant at the .01 level (for which the critical value was 24.05), but was significant at the .05 level (for which the critical value was 21.03). The results of the ANOVA with respect to firm #2 apparently caused the overall test statistic to be significant at the .05 level. The observed effect was limited to one firm, and therefore was not pervasive.6 The evidence was not sufficiently compelling to reject H1.1 at the .01 level. Accordingly, no real differences in location were identified among the seniors, managers, and partners.

For exploratory purposes, a supplemental type of comparison was made. Effects associated with level of responsibility were examined within each audit technology. That is, the standardized scores of the seniors, managers, and partners were compared, separately, within each technology group.

Two separate one-way ANOVAs explored each audit technology group for differences as to location. Level of responsibility was not
significant for either the structured or the unstructured firms. The p-value was approximately .20 for the structured firms and .08 for the unstructured firms. The p-value for the unstructured firms might be viewed as borderline significant. Intuitively, one might expect experience-related factors to be more influential within the unstructured firms (relative to the structured firms, which have developed more explicit guidelines to provide more direction to their auditors at all levels of responsibility). For the structured firms, the means were approximately 19.8, 11.3, and 2.9 for the seniors, managers, and partners, respectively. For the unstructured firms, the means were approximately -14.1, -9.9, and -10.0 for the seniors, managers, and partners, respectively.

Variation

H1.2 deals with differences as to variation among the seniors, managers, and partners. Based on Figure 8, which plots the standardized materiality scores by level of responsibility, no conspicuous differences in variation about the apparent means among the seniors, managers, and partners are visible.

Similar to the analyses reported above with respect to location, effects associated with level of responsibility were examined within each firm. The sums of squares were calculated for the seniors, managers, and partners, separately for each firm; and Cochran's test for homogeneity of variance was applied to those sums of squares, on a
firm-by-firm basis. As explained in Chapter III, the specific p-value was inferred for each firm with respect to this test for homogeneity of variance. Then, using Fisher's method for combining tests, the six p-values (one for each firm) were combined into an overall test statistic for the purpose of addressing H1.2.

No differences in variation were found among the seniors, managers, and partners for any of the six firms at the .05 level. When the six p-values were combined using Fisher's method, an overall test statistic of approximately 10.7 resulted. This was not significant at any reasonable level. (The critical values associated with the .01 and .05 levels were 24.05 and 21.03, respectively.) These results were consistent and unambiguous; there were no real differences as to variation among the seniors, managers, and partners. Hence, one cannot reject H1.2.

A supplemental type of comparison with respect to variation was made. Effects associated with level of responsibility were examined within each audit technology group. The sums of squares for the seniors, managers, and partners were calculated separately for each audit technology group; and Cochran's test for homogeneity of variance was separately applied. No significant difference was found for either audit technology group at the .05 level.
**Location**

H2.1 deals with differences as to location among the participating Big Eight firms. As noted above, Figure 9 plots the standardized materiality scores by firm affiliation. A casual examination suggests that there are distinct differences among the firms' apparent means (e.g., firm #1 appears to be high and firm #6 appears to be low, relative to the other firms).

In order to test H2.1, a one-way ANOVA compared the standardized scores for systematic differences in the means associated with firm affiliation. The means and standard deviations (sd) of the standardized materiality scores, by firm, were as follows:

<table>
<thead>
<tr>
<th>Firm</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>24.78</td>
<td>26.46</td>
</tr>
<tr>
<td>#2</td>
<td>5.79</td>
<td>21.28</td>
</tr>
<tr>
<td>#3</td>
<td>3.46</td>
<td>11.99</td>
</tr>
<tr>
<td>#4</td>
<td>-10.21</td>
<td>4.88</td>
</tr>
<tr>
<td>#5</td>
<td>-10.87</td>
<td>3.63</td>
</tr>
<tr>
<td>#6</td>
<td>-12.95</td>
<td>6.50</td>
</tr>
</tbody>
</table>

As documented in Table 11, firm affiliation was highly significant (p ≤ .0001). Specifically, Tukey's multiple comparison test identified firm #1 as significantly different than each of the other firms (#2-6); no other differences were significant at the .05 level. On this basis, one can reject H2.1; there were real differences at the firm level with respect to the central tendencies of the materiality judgments.

Interestingly, the means associated with the structured firms (firms #1-3) are all positive (hence, above average) and the means associated with the unstructured firms are all negative (hence, below
Summary of Level of Responsibility Effects

As described by the analyses above, one cannot reject the null hypothesis for either H1.1 (regarding location) or H1.2 (regarding variation). Overall, there were no real differences, either with respect to the means of the standardized materiality scores or with respect to the sums of squares associated with the standardized materiality scores, among the seniors, managers, and partners. These results are summarized in Table 10.

The one-way ANOVAs with respect to level of responsibility, calculated separately for each firm, identified only one significant difference. For firm #2, the seniors had higher materiality thresholds, on average, than the managers and partners (p ≤ .01). However, that was an isolated instance. Using Fisher's method for combining tests, the overall test statistic was significant at the .05 level (and was driven by the significant result with respect to firm #2), but it was not significant at the .01 level. Further, no effect associated with level of responsibility was identified as to location for either audit technology group, although the unstructured group could be viewed as having a borderline level of significance (p = .09).

Similar to the tests for differences as to location, the tests for differences as to variation were applied firm-by-firm. The seniors, managers, and partners were not different with respect to variation for any of the firms at the .05 level. The specific p-values inferred from Cochran's test for homogeneity of variance were
Table 10
Summary of Level of Responsibility Effects

<table>
<thead>
<tr>
<th>Comparisons Made</th>
<th>Re: Location</th>
<th>Re: Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Level of responsibility by firm affiliation</td>
<td>Seniors from firm #2 differed from others in that firm (p ≤ .01). No other differences were significant for any of the firms at the .05 level.</td>
<td>No significant differences at the .05 level.</td>
</tr>
<tr>
<td>B. Overall effect associated with level of responsibility</td>
<td>No significant differences at the .01 level. (A significant difference at the .05 level was driven by the seniors in firm #2.)</td>
<td>No significant differences at the .05 level.</td>
</tr>
<tr>
<td>C. Level of responsibility by audit technology</td>
<td>No significant differences for either audit technology (p=.20 and p=.08 for the structured and unstructured firms, respectively).</td>
<td>No significant differences at the .05 level.</td>
</tr>
</tbody>
</table>

Note: This table summarizes the analyses identified in Table 4. Table 4 summarizes the primary method of analysis for differences as to location and variation, identifies the data on which the above conclusions are based, and ties the analyses conducted to the notation introduced in Figure 7.
combined into an overall test statistic using Fisher's method. The overall test statistic was not significant at any reasonable level. Further, the seniors, managers, and partners were not significantly different with respect to variation for either of the two audit technology groups at the .05 level.

Based on these results, one may appropriately regard the seniors, managers, and partners as homogeneous, and proceed with the overall comparisons at the firm and audit technology levels. Those analyses follow.

Firm Effects

As described in Chapters II and III above, systematic differences among the participants' materiality judgments at the firm level were of interest. Prior research suggests that there may be differences between Big Eight and non-Big Eight firms or between national and non-national firms. Big Eight firms have been viewed as largely homogeneous due to the similarities in their client bases, national/international employee pool, sophisticated training, etc. Prior research has not explored whether there may be systematic differences in auditors' materiality thresholds associated with Big Eight firm affiliation. To address that issue, analyses similar to those described above with respect to level of responsibility were performed. Two hypotheses were explicitly stated with respect to firm
differences. H2.1 deals with firm differences as to location; and H2.2 deals with firm differences as to variation.

Using the notation presented in Figure 7, the overall firm effects were examined by comparing \( T_1, T_2, \ldots, \) and \( T_6 \). In order to further explore the overall firm differences, two additional types of comparisons were made, and the results are also reported below. First, each of the levels of responsibility was examined separately across the six participating firms for differences associated with each level. In terms of Figure 7, the following additional comparisons were made: \( A_1, A_2, \ldots, \) and \( A_6 \); \( B_1, B_2, \ldots, \) and \( B_6 \); and \( C_1, C_2, \ldots, \) and \( C_6 \). Second, each audit technology was examined separately for firm differences within the audit technologies. In terms of Figure 7, the following additional comparisons are made: \( T_1, T_2, \) and \( T_3 \); and \( T_4, T_5, \) and \( T_6 \). Formal hypotheses were not stated with respect to these supplemental comparisons, since the purpose was to provide further insights regarding overall firm differences.

Figure 9 plots these standardized materiality scores by firm affiliation (where firms #1-3 are the structured firms and firms #4-6 are the unstructured firms). Differences among the six participating Big Eight firms are discussed first with respect to location, and second with respect to variation.
Figure 9

Standardized Scores for the Income Statement Materiality Judgments Plotted by Firm Affiliation
Table 11
ANOVA Table For Firm Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
<th>Mean Square</th>
<th>F-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5</td>
<td>12,599.5</td>
<td>2,519.9</td>
<td>10.07</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>66</td>
<td>16,512.4</td>
<td>250.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>29,111.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
average). The mean standardized score for each structured firm exceeds the mean standardized score for each unstructured firm. This issue will be discussed in more depth in a subsequent section with respect to "Audit Technology Effects."

In order to explore further the overall firm differences identified above, two additional types of comparisons were made. First, each of the levels of responsibility was examined separately across the six participating firms for firm differences associated with each level. Second, each audit technology group was examined separately for firm differences within the audit technologies.

Three separate one-way ANOVAs were conducted to identify whether there were any significant differences as to location among the firms at each level of responsibility. The purpose of this analysis was to explore the pervasiveness of the firm differences, and to ascertain whether the overall results described above were driven by any particular level of responsibility. For the seniors, firm effects were highly significant ($p \leq .0001$); Tukey's multiple comparison test identified the seniors from firms #1 and 2 to be different than the seniors from firms #4-6. For the managers, firm effects were also highly significant ($p \leq .001$); Tukey's multiple comparison test identified the managers from firm #1 to be different than the managers from firms #2-6. For the partners, firm affiliation was borderline significant ($p \approx .06$); the partners from firm #1 had the highest average standardized scores, while the partners from firm #4 had the lowest.
Consequently, the existence of firm differences in location was fairly consistently observed across all three levels of responsibility. The overall firm differences reported above do not appear to be driven by the judgments of a single level.

Two separate one-way ANOVAs looked for firm differences as to location within each audit technology group. Among the three structured firms (firms #1-3), firm #1 was weakly significant (p=.043) relative to firm #3. (Since firm #1 had the highest mean and firm #3 had the lowest mean with respect to the structured firms, the p-value obtained in the one-way ANOVA can be attributed to the comparison of those two firms.) Among the three unstructured firms (firms #4-6), firm differences were not significant (p=.44). This result, in conjunction with the observation made earlier that the mean for each structured firm exceeded the mean for each unstructured firm, suggests that firm differences as to location were more conspicuous between firms having different audit technologies.

**Variation**

H2.2 deals with differences as to variation across the six participating Big Eight CPA firms. Based on Figure 9, which plots the standardized materiality scores by firm affiliation, differences in variation about the firms' means are visible. For example, the distribution for firm #1 is widely dispersed, whereas the distribution
for firm #4 is tightly compacted. Thus, firm differences with respect to variation are conspicuous.

In order to test H2.2, Cochran's test for homogeneity of variance was applied to the sums of squares, calculated separately for each firm, with respect to their standardized materiality scores. The following sums of squares were used to test H2.2:

<table>
<thead>
<tr>
<th>Firm</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>8,399</td>
</tr>
<tr>
<td>#2</td>
<td>5,436</td>
</tr>
<tr>
<td>#3</td>
<td>1,725</td>
</tr>
<tr>
<td>#4</td>
<td>286</td>
</tr>
<tr>
<td>#5</td>
<td>158</td>
</tr>
<tr>
<td>#6</td>
<td>507</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>16,511</strong></td>
</tr>
</tbody>
</table>

Based on a calculated measure of approximately .509 (i.e., 8,399/16,511), Cochran's test for homogeneity indicates that one must reject H2.2; the firms do not have equal variances \((p \leq .01)\). The next most variable firm (#2) was tested in the same way, after deleting the sums of squares attributable to firm #1. Firm #2 contributed 67 percent of the remaining sums of squares; again, the remaining firms (#2-6) did not have equal variances \((p \leq .01)\). Firm #3, the next most variable firm, was tested after further deleting the sums of squares related to firm #2. Firm #3 contributed 64.5 percent of the remaining sums of squares; again, the remaining firms (#3-6) did not have equal variances \((p \leq .01)\). The variances of the other firms (#4, 5, and 6) were not found to be different at the .05 level.
This indicates that each of the structured firms was more variable than at least one of the unstructured firms.\textsuperscript{11}

Similar to the tests for location differences reported above, two types of supplemental comparisons with respect to variation were made for exploratory purposes. First, tests were conducted to identify firm differences in variation at each level of responsibility. Second, tests were conducted to identify whether there were any firm differences as to variation within each of the audit technology groups.

To test for differences in variation across firms, by level of responsibility, the variances of the standardized scores were calculated, by firm and by level of responsibility. (These variances are summarized in Table 12.) Cochran's test for homogeneity of variance was applied to these variances across firms, on a level-by-level basis.

For the seniors, significant differences in variation were detected among the firms. Cochran's test was applied sequentially, deleting each significant firm. Firm #1 was the most variable, followed by firms #2, 4, 3, 6, and 5 in order of decreasing variability. Firms #1, 2, and 4 were significant at the .01 level, and firms #3 and 6 were significant at the .05 level.

For the managers, significant differences in variation were also found among the firms. Firm #1 was the most variable, followed by
Table 12

Average Variance By Firm and Level of Responsibility Based on Standardized Materiality Scores

<table>
<thead>
<tr>
<th>Firm #1</th>
<th>Seniors</th>
<th>Managers</th>
<th>Partners</th>
<th>Total</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>304.66</td>
<td>231.78</td>
<td>21.73</td>
<td>558.17</td>
<td>186.06</td>
</tr>
<tr>
<td>3</td>
<td>21.59</td>
<td>172.66</td>
<td>186.54</td>
<td>380.79</td>
<td>126.93</td>
</tr>
<tr>
<td>4</td>
<td>53.31</td>
<td>28.49</td>
<td>3.89</td>
<td>85.69</td>
<td>28.56</td>
</tr>
<tr>
<td>5</td>
<td>1.99</td>
<td>17.56</td>
<td>23.17</td>
<td>42.72</td>
<td>14.24</td>
</tr>
<tr>
<td>6</td>
<td>11.39</td>
<td>21.99</td>
<td>58.35</td>
<td>91.73</td>
<td>30.58</td>
</tr>
<tr>
<td>Average</td>
<td>160.61</td>
<td>162.85</td>
<td>310.52</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
firms #2 and 3. Each was significant at the .01 level; no other differences among the managers were significant at the .05 level.

For the partners, significant differences in variation were also identified among the firms. Firm #1 was the most variable, followed by firms #3 and 6. Firms #1 and 3 were significant at the .01 level, and firm #6 was significant at the .05 level. These were the only differences among the partners that were significant at the .05 level.

The results of these analyses of firm differences, by level of responsibility, are consistent with the overall firm effects reported above. These analyses identified fairly pervasive firm effects across each of the levels. The overall firm effects discussed above do not appear to be driven by any particular level of responsibility.

Finally, two other tests looked for firm differences as to variation within each audit technology group. Cochran's test for homogeneity of variance was applied to the sums of squares for the structured firms (firms #1-3 above) and, separately, to the sums of squares for the unstructured firms (firms #4-6 above). No differences in variation were detected for the three firms in the structured group or for the three firms in the unstructured group at the .05 level. This supplemental analysis suggests that the differences in variation were much greater for firms having different audit technologies than for firms having the same audit technology.
Summary of Firm Effects

Differences at the firm level were identified both with respect to location and variation. Accordingly, both H2.1 (regarding differences as to location) and H2.2 (regarding differences as to variation) can be rejected. These results are summarized in Table 13.

With respect to location, differences among the means of the firms' standardized materiality scores were highly significant (p \leq .0001). A one-way ANOVA regarding firm affiliation was used to detect significant differences among the firms. Although only firm #1 was demonstrated to be significantly different than the other firms, interestingly, the means for each of the structured firms (#1-3) exceeded the means for each of the unstructured firms (#4-6).

A supplemental analysis explored whether there were firm differences as to location among each of the three levels of responsibility. Firm differences were identified separately for the seniors, managers, and partners. The results suggested fairly pervasive firm effects. The overall firm differences did not appear to be driven by any particular level of responsibility.

Another supplemental analysis examined each audit technology group separately for differences as to location among the firms. Within the structured audit technology group, firm #1 was weakly significant relative to firm #3 (p=0.043). Within the unstructured firms, firm affiliation was not significant (p=0.44). Thus,
Table 13
Summary of Firm Effects

<table>
<thead>
<tr>
<th>Comparisons Made</th>
<th>Re: Location</th>
<th>Re: Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Overall firm effect</td>
<td>Firm affiliation was highly significant (p ≤ .0001). Firm #1 was high relative to all other firms.</td>
<td>Firm differences were found at the .01 level. (Firms #1, 2, and 3 were high relative to firms #4-6.)</td>
</tr>
<tr>
<td>B. Firm affiliation, by level of responsibility</td>
<td>Firm affiliation was highly significant for the seniors (p ≤ .0001) and managers (p ≤ .001); the partners were borderline (p = .06).</td>
<td>Firm differences were found at the .01 level among the seniors, the managers, and the partners.</td>
</tr>
<tr>
<td>C. Firm affiliation, by audit technology</td>
<td>Firms #1 and 3 were weakly significant for the structured firms (p = .043); there were no differences among the unstructured firms (p = .44).</td>
<td>No difference was found at the .05 level for either audit technology.</td>
</tr>
</tbody>
</table>

Note: This table summarizes the analyses identified in Table 5. Table 5 summarizes the primary method of analysis for differences as to location and variation, identifies the data on which the above conclusions are based, and ties the analyses conducted to the notation introduced in Figure 7.
differences in location at the firm level were more conspicuous for firms having different audit technologies than for firms within the same audit technology group.

With respect to variation, distinct firm differences were also identified. Cochran's test for homogeneity of variance, based on the sums of squares associated with each firm's standardized materiality scores, established that the variances of the firms were not all equal. (Sequential analyses were made by deleting the most variable firm when Cochran's test was significant. Firm #1 was significant relative to at least one other firm; after deleting firm #1, firm #2 was significant relative to at least one of the remaining firms; and, after further deleting firm #2, firm #3 was significant relative to at least one of the remaining firms.) Interestingly, the sums of squares for each of the structured firms exceeded those for each of the unstructured firms.

A supplemental analysis explored whether there were firm differences as to variation among each of the three levels of responsibility. Firm differences were identified separately for the seniors, managers, and partners. These results indicated fairly pervasive firm effects with respect to variation. Again, the overall firm differences were not driven by the judgments of subjects from any particular level of responsibility.

Another supplemental analysis examined each audit technology group separately for differences as to variation among the firms.
Firm differences were not significant for either of the two audit technologies. As was pointed out with respect to location, differences in variation at the firm level were more conspicuous for firms having different audit technologies than for firms within the same audit technology group.

These results indicate that there were real differences in the judgments of the participants at the firm level. In addition, the supplemental analyses conducted across firms within each audit technology suggested that differing audit technologies may have been a factor in the differences among the firms, both as to location and variation. The next section looks specifically at that issue.

Audit Technology Effects

In order to explore further the issue of systematic differences among the Big Eight firms, the participating firms were classified as structured or unstructured, based on the notion developed by Cushing and Loebbecke [1983] and Kinney [1986]. As discussed in Chapters II and III, this concept of audit technology has implications with respect to auditors and their differing materiality judgments. In particular, Cushing and Loebbecke commented on the divergent policies and practices with respect to planning stage materiality judgments across the firms that they studied. Kinney (in addition to Cushing and Loebbecke) pointed out that the structured firms provide more
detailed guidelines to achieve greater uniformity in judgmental areas; the unstructured firms allow individual auditors relatively more autonomy in judgmental areas.

The following discussion is related to the previous section regarding firm affiliation. Now, instead of partitioning the 72 subjects into six categories based on firm affiliation, they are partitioned into two categories—the structured and unstructured firms. Two hypotheses were explicitly stated with respect to audit technology differences. H3.1 deals with differences between the structured and unstructured firms with respect to location; and H3.2 deals with differences between the structured and unstructured firms with respect to variation. Analyses similar to those described above with respect to firm affiliation were performed.

Using the notation presented in Figure 7, audit technology effects were examined by comparing $T_s$ and $T_u$. In order to further explore the pervasiveness of any overall audit technology effects, an additional type of comparison was made. Audit technology effects were examined, on a level-by-level basis, to ascertain whether the structured and unstructured firms differed at each of the levels for the seniors, managers, and partners. In terms of Figure 7, this supplemental comparison involved $A_s$ v. $A_u$, $B_s$ v. $B_u$, and $C_s$ v. $C_u$.

As was indicated in the preceding section regarding firm effects, Figure 9 plots the standardized scores by firm affiliation (where firms #1-3 are the structured firms and firms #4-6 are the
unstructured firms). That plot is useful in identifying both
differences among the six firms and differences among the two audit
technology groups. Differences among the structured and unstructured
firms are discussed first with respect to location, and second with
respect to variation.

**Location**

H3.1 deals with differences as to location between the structured
and unstructured audit technology groups. As mentioned, Figure 9
plots the standardized materiality scores by firm affiliation; there
are visually apparent differences in the means for firms #1-3 (the
structured firms) relative to those for firms #4-6 (the unstructured
firms). Each of the structured firms has an apparently higher mean
than each of the unstructured firms.

To test H3.1, a one-way ANOVA compared the standardized scores
for systematic differences in the means for each audit technology
group. Based on the 36 subjects from each audit technology group, the
means of the standardized scores were 11.345 (-11.345) for the
structured (unstructured) firms. As documented in Table 14, the ANOVA
found audit technology to be highly significant (p ≤ .0001). On this
basis, one can reject H3.1; there were real differences between the
structured and unstructured firms with respect to the central
tendencies of the materiality judgments.
Table 14
ANOVA Table For Audit Technology Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
<th>Mean Square</th>
<th>F-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>9,266.4</td>
<td>9,266.4</td>
<td>32.69</td>
<td>$\leq .0001$</td>
</tr>
<tr>
<td>Residual</td>
<td>70</td>
<td>19,845.5</td>
<td>283.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>29,111.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
That audit technology was highly significant is not surprising, in view of the results regarding firm effects reported in the preceding section. With respect to firm differences as to location discussed above, only firm #1 was shown to be significantly different than the other firms, based on a one-way ANOVA. However, the fact that the mean for each of the three structured firms exceeded the mean for each of the three unstructured firms strongly suggests that the overall significance of audit technology is not driven by the earlier reported significance of firm #1.

In order to explore the pervasiveness of the overall audit technology effect, an additional type of comparison was made. The purpose of this supplemental analysis was to ascertain whether the overall audit technology effect was consistently observed across the three levels of responsibility. Each level was examined separately to identify audit technology differences among the seniors, the managers, and the partners. Three separate one-way ANOVAs tested for audit technology differences between the structured and unstructured firms, on a level-by-level basis.

For the seniors, audit technology was highly significant (p < .0001). The seniors affiliated with the structured (unstructured) firms had a mean standardized score of 19.8 (-14.1). For the managers, audit technology was also highly significant (p < .005). The managers affiliated with the structured (unstructured) firms had a mean standardized score of 11.3 (-9.9). For the partners, audit
technology was not significant ($p = .12$). Nevertheless, the partners affiliated with the structured firms also had a higher mean standardized score than their counterparts with the unstructured firms (2.9 and -10.0, respectively). These results indicate fairly pervasive and consistent audit technology differences as to location across levels of responsibility.

**Variation**

H3.2 deals with differences as to variation between the structured and unstructured audit technology groups. Based on Figure 9, which plots the standardized materiality scores by firm affiliation, there are clear differences in the variability about the means for firms #1-3 (the structured firms) relative to those for firms #4-6 (the unstructured firms). Each of the structured firms has greater variability about its apparent mean relative to each of the unstructured firms.

To test H3.2, Cochran's test for homogeneity of variance was applied to the sums of squares associated with the standardized materiality scores for the structured and unstructured firms. The following sums of squares were obtained:

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured firms</td>
<td>18,845</td>
</tr>
<tr>
<td>Unstructured firms</td>
<td>1,000</td>
</tr>
<tr>
<td>Total</td>
<td>19,845</td>
</tr>
</tbody>
</table>

Cochran's test for homogeneity of variance verifies the obvious. The participants affiliated with structured firms were significantly
more variable than the participants from unstructured firms—95
percent of the total sums of squares was attributable to the
structured firms (p ≤ .01). Accordingly, one can reject H3.2,
regarding equal variances between the two audit technology groups. 14

In view of the analysis of firm effects reported in the preceding
section, it is not surprising that the structured firms (#1-3) were
found to be more variable than the unstructured firms (#4-6). With
respect to firm differences in variation discussed above, Cochran's
test was applied sequentially to the firms' sums of squares, deleting
the most variable firm each time the firms were shown to have unequal
variances. When all six firms were considered, firm #1 was the most
variable and the firms were shown to have unequal variances. After
deleting firm #1, firm #2 was the most variable and, again, the
remaining firms were shown to have unequal variances. After further
deleting firm #2, firm #3 was the most variable and, again, the
remaining firms were still shown to have unequal variances. Yet, a
supplemental analysis regarding variation was unable to demonstrate
any statistically significant differences among the three structured
firms and, separately, among the three unstructured firms.

In order to explore the pervasiveness of the overall audit
technology effect with respect to variation, an additional type of
comparison was made. The purpose of this supplemental analysis was to
ascertain, whether the greater variability of the structured firms was
consistently observed across the three levels of responsibility. Each
level was examined separately to identify audit technology differences among the seniors, the managers, and the partners.

The sums of squares were calculated separately for the twelve subjects at each level of responsibility within each audit technology group. (These are summarized in Table 15.) Cochran's test for homogeneity of variance was used to test the equality of variance between the structured and unstructured firms, on a level-by-level basis. The seniors, the managers, and the partners affiliated with structured firms were, at each level, more variable than their counterparts affiliated with unstructured firms (p ≤ .01). The overall greater variability of the structured firms was consistently observed at each level of responsibility.

Summary of Audit Technology Effects

Audit technology effects were identified with respect to both location and variation. Accordingly, both H3.1 (regarding differences as to location) and H3.2 (regarding differences as to variation) were rejected. These results are summarized in Table 16.

With respect to location differences, audit technology was highly significant (p ≤ .0001), based on the 36 subjects representing each of the two audit technology groups. Participants from the structured firms had higher materiality thresholds relative to the participants from the unstructured firms. Indeed, the mean associated with each of the structured firms was higher than the mean associated with each of
Table 15
Sums of Squares of the Standardized Scores
By Level of Responsibility and By Technology

<table>
<thead>
<tr>
<th></th>
<th>Structured Firms</th>
<th>Unstructured Firms</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seniors</td>
<td>3424</td>
<td>217</td>
<td>3641</td>
</tr>
<tr>
<td>Managers</td>
<td>5529</td>
<td>211</td>
<td>5740</td>
</tr>
<tr>
<td>Partners</td>
<td>8164</td>
<td>432</td>
<td>8596</td>
</tr>
<tr>
<td>Total</td>
<td>17,117</td>
<td>860</td>
<td>17,977</td>
</tr>
</tbody>
</table>
Table 16
Summary of Audit Technology Effects

<table>
<thead>
<tr>
<th>Comparisons Made</th>
<th>Re: Location</th>
<th>Re: Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Overall audit technology effect</td>
<td>Audit technology was highly significant ( (p \leq 0.0001) ). The structured firms had higher thresholds than the unstructured firms.</td>
<td>The structured firms had greater variability than the unstructured firms ( (p \leq 0.01) ).</td>
</tr>
</tbody>
</table>

| B. Audit technology, by level of responsibility | Audit technology was highly significant for the seniors \( (p \leq 0.0001) \) and for the managers \( (p \leq 0.005) \); the partners were not significant \( (p = 0.12) \). At each level, the structured firms had higher thresholds. | At each level, the structured firms were more variable \( (p \leq 0.01) \). |

Note: This table summarizes the analyses identified in Table 6. Table 6 summarizes the primary method of analysis for differences as to location and variation, identifies the data on which the above conclusions are based, and ties the analyses conducted to the notation introduced in Figure 7.
the unstructured firms. Further, the characteristic of higher materiality thresholds for participants from structured firms was consistently observed across each level of responsibility. (Although audit technology was not significant for the partners, it was highly significant for the seniors and the managers.)

With respect to variation, the participants from structured firms were found to be significantly more variable in their judgments than the participants from unstructured firms ($p \leq 0.01$). The sums of squares associated with each of the structured firms was higher than that for each of the unstructured firms. The greater variability of the structured firms was a pervasive phenomenon, as the participants from structured firms were more variable than the subjects from unstructured firms at each of the three levels of responsibility.

**Risk Attitude**

A secondary research issue identified in Chapters II and III related to why individual auditors might (rationally) differ in their materiality thresholds. One reason offered was that auditors might differ in their views of the appropriate tradeoffs between the so-called Type I and Type II errors. In that regard, two specific hypotheses were stated. $H_4.1$ dealt with the existence of differences among the auditors with respect to the risk attitude notion; related to that, viewing risk attitude as an intervening variable affecting
the materiality judgment, H4.2 dealt with the association between the auditors' risk attitudes and the extremeness of their materiality thresholds. It was expected that auditors who were relatively more concerned with Type I errors (e.g., practice development) would tend to have higher materiality thresholds, whereas auditors who were relatively more concerned with Type II errors (e.g., litigation by financial statement users) would tend to have lower thresholds.

As explained in Chapter III, risk attitude was based on an expectancy-va lance measure with respect to the four cells identified in Figure 3. (The individual items associated with the four cells were identified in Table 7.) Subjects were asked to specify their beliefs (likelihoods on a seven-point scale) and their evaluations (assessments of importance on a five-point scale). The product of the belief and evaluation was calculated for each item (termed the item's expectancy-va lance), for each subject. The items' expectancy-va lences were summed within each cell for each subject.

Risk attitude was defined to be the subject's tradeoff of Type I and Type II errors. It was measured by the expectancy-va lance for cell #4 (regarding Type II errors) divided by the expectancy-va lance for cell #1 (regarding Type I errors) for each subject.

To construct a reasonably powerful test to identify any differences among the participants as to risk attitude, and to measure the strength of the association between risk attitude and their materiality judgments, subjects were partitioned into one of three
groups. Based on a ranking of their standardized materiality scores from lowest to highest, the 72 subjects were partitioned into the lower one-third, middle one-third, and upper one-third. Then, the subjects in the upper and lower thirds were compared for systematic differences.

A one-way ANOVA was conducted on these overall risk attitude measures to identify any differences in the means between subjects in the upper and lower one-thirds. The risk attitudes were not found to differ between the two groups (p = .28). Contrary to expectations, the risk attitude measure was higher for the subjects with the highest materiality judgments. It was expected that subjects with a relatively greater sensitivity to Type II errors (reflected by the numerator in the risk attitude measure) would tend to establish relatively lower materiality thresholds.

In order to examine the risk attitude measures more closely, one-way ANOVAs were run on the expectancy-valence measures for each of the four cells represented by Figure 3. The p-values associated with cells #1 (Type I error), 2, 3, and 4 (Type II error) were .0507, .1456, .7877, and .0824, respectively. Comparing the means of cells #1 and 4, the subjects with the lowest materiality judgments had the higher mean for cell #1 and the lower mean for cell #4. These were opposite to the hypothesized relationship. (As indicated, it was expected that the low materiality threshold group would emphasize cell #4 regarding Type II errors; and it was expected that the high
materiality threshold group would emphasize cell #1 regarding Type I errors.) On the basis of the ANOVAs with respect to the overall risk attitude measures and with respect to the risk attitude measures calculated separately for Type I and Type II errors, H4.1 (that there are no systematic differences among the auditors with respect to the risk attitude measures) cannot be rejected.\(^16\)

Using the Pearson product-moment correlation, subjects' risk attitude measures were correlated with their standardized materiality scores. The overall correlation coefficient between the risk attitude measures and the standardized materiality scores was .21203, and was not significant at the .05 level \((p = .074)\). It was surprising that this coefficient was positive, since a negative association was hypothesized. (The \(p\)-value of .074 was comparable in magnitude to the overall significance reported by Kida \(1980\) between his expectancy-value measure of attitude and his subjects' qualifying behavior—\(p = .068\). However, as indicated, the observed direction was opposite to that hypothesized here.)\(^17\)

These results suggest that the null hypothesis for H4.2 cannot be rejected (which is not surprising since the null hypothesis for H4.1 could not be rejected either). The measure of risk attitude adopted in this study was not found to vary systematically among the subjects, nor was it systematically related to participants' materiality thresholds. The overall notion of risk attitude, as operationalized
here, was not helpful in explaining why subjects differed in their materiality thresholds.

Further investigation was conducted to identify whether there were attitudinal differences, at a lower level of aggregation, between the subjects with the lowest and the highest materiality thresholds. Attitudinal differences were explored on an item-by-item basis, in terms of the belief, evaluation, and the product of belief and evaluation, termed the item's expectancy-valence (EV), within each cell (see Figure 3) represented by the attitude measurement questionnaire (Appendix B). Table 17 identifies the attitudinal differences, on an item-by-item basis, between the subjects with the lowest and highest thresholds.

As reported above, the subjects with the lowest thresholds tended to attach greater emphasis (based on the EV's) to Type I errors relative to subjects with the highest thresholds. In particular, the low threshold group had higher EV's regarding the following items (at the .10 level): litigation by the client, damage to one's own standing within the firm, and enhancing the accounting firm's reputation. Subjects with the highest thresholds tended to attach greater emphasis to Type II errors relative to subjects with the lowest thresholds. In particular, the high threshold group had a higher EV with respect to damaging the client's reputation. As indicated, it was expected that the high threshold group would attach
Table 17

Attitudinal Differences (p ≤ .10) Among Subjects With the Lowest Versus Highest Thresholds

<table>
<thead>
<tr>
<th>Cell of Fig.3.1</th>
<th>Item</th>
<th>Belief Eval'n</th>
<th>Exp.Val.</th>
<th>Highest v.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lose the client</td>
<td>.0555</td>
<td></td>
<td>L&gt;H</td>
</tr>
<tr>
<td>1</td>
<td>Litigation by client</td>
<td>.0937</td>
<td></td>
<td>L&gt;H</td>
</tr>
<tr>
<td>1</td>
<td>Damage own standing</td>
<td>.0192</td>
<td></td>
<td>L&gt;H</td>
</tr>
<tr>
<td>2</td>
<td>Enhance firm's reputation</td>
<td>.0203</td>
<td></td>
<td>L&gt;H</td>
</tr>
<tr>
<td>2</td>
<td>Damage client's reputation</td>
<td>.0797</td>
<td></td>
<td>L&gt;H</td>
</tr>
<tr>
<td>4</td>
<td>Deteriorated client relations</td>
<td>.0873</td>
<td></td>
<td>L&lt;H</td>
</tr>
<tr>
<td>4</td>
<td>Damage client's reputation</td>
<td>.0212</td>
<td>.0494</td>
<td>L&lt;H</td>
</tr>
</tbody>
</table>
more emphasis to Type I errors and that the low threshold group would attach more emphasis to Type II errors.

**Additional Analyses**

This section reports five additional types of analyses. First, a comparison is made of the ANOVA models calculated for each subject across the 24 cases to which the subjects responded. The purpose is to identify any apparent differences, with respect to level of responsibility and/or firm affiliation, in the structural forms of the ANOVA models. Differences in subjects' materiality thresholds may be influenced by differences in the financial statement cues to which they attended. Second, subjects' responses with respect to the degree of business risk (i.e., exposure) associated with the financial statement cues were examined on a case-by-case basis. The purpose of this analysis was to ascertain whether the differences identified above were confounded by different perceptions of risk implied by the financial statement variables. Third, the four repeated cases for each subject were compared to their original judgments. This test-retest consistency was then examined for significant differences among the three levels of responsibility, among the six participating firms, and between the two audit technologies. Fourth, the data were examined to ascertain whether the two orders in which the cases were presented had any apparent effect on subjects' responses. Fifth and
finally, several comparisons were made to identify whether the income statement materiality thresholds were less than the balance sheet materiality thresholds, as expected on average (for each level of responsibility, for each firm, and for both audit technologies). These comparisons were made, on a case-by-case basis, to explore ex post the appropriateness of viewing the income statement materiality judgments as the dependent variable of primary interest.

**Individual ANOVA Models**

In order to gain insights into the nature of individual differences among the participants, ANOVA models were calculated for each subject across the 24 cases. These models were based on subjects' untransformed (or "raw") judgments; and they were intended to be descriptive, not to represent the actual "processes" which subjects followed in arriving at their judgments.

In the aggregate, for all 72 subjects, income (loss) before taxes ("income") was significant for 49 subjects. The total asset variable was the second most frequently significant variable, and was significant for 24 subjects. The sales variable was significant for 15 subjects. Only one interaction was often significant; income x assets was significant for 19 subjects. The seniors tended to rely more on sales relative to the managers and partners, who apparently relied on income to a greater extent. Table 18 summarizes the
Table 18
All Significant Cues (p ≤ .05)
By Level of Responsibility, By Firm Affiliation

<table>
<thead>
<tr>
<th>By Level of Responsibility:</th>
<th>Seniors</th>
<th>Managers</th>
<th>Partners</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income (loss) before taxes</td>
<td>11</td>
<td>17</td>
<td>21</td>
<td>49</td>
</tr>
<tr>
<td>Net sales</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Total assets</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>Income x sales</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Income x assets</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>Sales x assets</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>By Firm Affiliation:</th>
<th>Structured</th>
<th>Unstructured</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Firm #1</td>
<td>#2</td>
</tr>
<tr>
<td>Income (loss) before taxes</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Net sales</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>Total assets</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Income x sales</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Income x assets</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Sales x assets</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
significant cues \((p \leq 0.05)\), by level of responsibility and, separately, by firm affiliation.

Table 19 summarizes the significant cues for the subjects for whom only one variable was significant. Income was the only significant variable for 29 subjects, followed by sales for eleven subjects, and total assets for seven subjects.

Firm differences were identifiable. Total assets was significant for 24 subjects. Nearly half (eleven) were from firm #1, a structured firm. In addition, all seven of the subjects for whom total assets was the only significant variable were affiliated with firm #1. Further, only one subject from firm #1 for whom just one variable was significant relied on a variable other than total assets (i.e., income).

The standardized materiality scores for the seven subjects from firm #1 for whom total assets was the only significant variable were examined further. Of the seven highest standardized scores, six were subjects from firm #1, and all six had only one significant variable--total assets. These subjects comprised half of the twelve subjects from firm #1. Apparently, the conspicuously high materiality thresholds associated with subjects from firm #1 were driven, or at least influenced, by their relatively strong balance sheet orientation within the context of this experiment.

Although somewhat ad hoc, these observations underscore the existence of individual differences, even in the structural form of
Table 19

Only One Significant Cue (p < .05)
By Level of Responsibility, By Firm Affiliation

<table>
<thead>
<tr>
<th>By Level of Responsibility:</th>
<th>Seniors</th>
<th>Managers</th>
<th>Partners</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income (loss) before taxes</td>
<td>5</td>
<td>10</td>
<td>14</td>
<td>29</td>
</tr>
<tr>
<td>Net sales</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Total assets</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>By Firm Affiliation:</th>
<th>Structured</th>
<th></th>
<th>Unstructured</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Firm #1</td>
<td>#2</td>
<td>#3</td>
<td>#4</td>
</tr>
<tr>
<td>Income (loss) before taxes</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Net sales</td>
<td>-</td>
<td>4</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>Total assets</td>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
the models of subjects' judgments. This is somewhat surprising in view of the fairly limited experimental manipulations—only three financial statement cues were available and only two of these were directly applicable to the income statement. Particularly with respect to firm #1, differences in subjects' thresholds apparently related to differences in subjects' structural models.

**Business Risk Assessments**

To the extent that subjects perceived the financial statement cues as implying different degrees of business risk (i.e., exposure) to the auditors, their materiality thresholds may be affected. All else equal, higher exposure to the auditor is believed to result in a lower materiality threshold. To ascertain whether there was such a confounding effect present in the experiment, subjects were asked to rate the business risk associated with each case on a seven-point scale (where "1" represented "very low" and "7" represented "very high"). The responses were analyzed by one-way ANOVAs, on a case-by-case basis, to evaluate whether there were systematic differences which might affect the above-reported results with respect to level of responsibility, firm affiliation, or audit technology.

With respect to level of responsibility, only two cases were significant at the .05 level. The seniors expressed a higher level of risk associated with case #5 than did the managers ($p = .048$). The
partners expressed a higher level of risk associated with case #19 than did the managers ($p = .013$).

With respect to firm affiliation, four cases were significant ($p \leq .05$). Firm #2 expressed greater risk than did firm #4 for case #1 ($p = .023$). Firm #2 expressed greater risk than did Firm #6 for case #23 ($p = .037$). Firm #5 expressed a higher level of risk than did firm #1 for case #9 ($p = .049$). Finally, firm #5 expressed a higher level of risk than did firm #6 for case #18 ($p = .019$).

Technology was not significant for any of the cases. Overall, no pattern which related to the identified differences in subjects' materiality judgments emerged with respect to these business risk assessments, in terms of responsibility, firms, or audit technology.

**Test-retest Consistency**

As a further investigation of the meaningfulness of subjects' responses, four cases were repeated for each individual following the 24 original cases. Each subject dealt with the same repeated cases which were arbitrarily selected (two were selected from the first half of the cases presented and two were selected from the second half of the original cases). If the subjects' responses were different on the repeated cases, then their original responses may be suspect. Differences would imply either inattention to the task or that the subjects were otherwise unable/unwilling to make meaningful responses based on the experimental materials.
The responses for the repeated cases were compared to the original responses in three ways. First, the overall Pearson product-moment correlation coefficient was calculated to identify whether the responses on the four repetitions moved in the same direction as those to the original cases. Second, paired t-tests were calculated, on a case-by-case basis, to identify whether, on average, the responses to the four repeated cases were comparable to the responses to the original cases. Last, separate one-way ANOVAs were run on the differences between the responses to the original and repeated cases to identify any systematic differences among the levels of responsibility, among the firms, and between the two audit technologies.

Across all 72 subjects, the correlation coefficient of the repeated cases relative to the original cases was approximately .963. Not surprisingly, larger responses on the original cases appear to be paired with larger responses on the repeated cases.

For the 72 subjects in the aggregate, the hypothesis of no difference between the repeated and the original cases could not be rejected for any of the four judgments. The resulting approximate t-values (p-values) for cases #3, 9, 18, and 22 were: -.36 (.72), -.96 (.34), -.67 (.50), and .73 (.47), respectively.

Separate one-way ANOVAs, with respect to level of responsibility, firm affiliation, and audit technology, were run on the differences between the original and repeated cases, on a case-by-case basis.
Level of responsibility, firm affiliation, and audit technology were not significant at the .10 level for any of the repeated cases. Hence, there were no systematic differences among the subjects in the test-retest consistency of the experimental task which would call into question the validity of their responses.22

Order Effects

"Order effects" were examined to identify whether the order in which the cases were presented affected subjects' responses. To the extent that the order of presentation is associated with systematic differences among subjects, the results reported above may be viewed somewhat as artifacts of the experiment.

The cases were presented to the participants in two different randomized orders. The 24 unique combinations of the independent variables were initially ordered randomly (constituting order #1). This initial ordering was then reversed (constituting order #2). Since each of the participating firms contributed twelve subjects, the experimental materials were consecutively numbered to provide control for purposes of follow-up and identification, while maintaining individuals' anonymity. The even-numbered booklets were order #1, and the odd-numbered booklets were order #2. Consequently, the orders were systematically varied across levels of responsibility, firms, and audit technologies.
With respect to location, a one-way ANOVA of the standardized scores did not find any difference in the mean judgments between the two orders (p=.71). With respect to variation, Cochran's test for homogeneity of variance (using the sums of squares associated with each order) did not detect a significant difference at the .05 level. Apparently, the judgments were not influenced by the order in which the cases were presented.23

**Income Statement versus Balance Sheet Judgments**

SAS No. 47 (paragraph 12) describes the preliminary judgment about materiality, at the financial statement level, in terms of the smallest amount material to any one of the financial statements. Prior research has consistently reported net income (or income before taxes) to be the dominant variable in numerous studies of auditors' evaluation stage materiality judgments. Balance sheet items (such as total assets) have tended to be unimportant variables in those studies. Further, others have asserted that the planning stage judgments are generally driven more by income statement considerations rather than balance sheet considerations (e.g., Zuber, et al. [1983, p. 43]). Consequently, it was expected that the income statement materiality judgments would generally be less than the balance sheet materiality judgments (i.e., the preliminary judgment about materiality would result from income statement considerations) for most subjects for most cases. That expectation was examined in a
somewhat ad hoc fashion, separately with respect to level of responsibility, firm affiliation, and audit technology.

Using the participants' untransformed ("raw") materiality judgments, the average income statement and balance sheet materiality judgments were calculated for each level of responsibility (for the seniors, managers, and partners) on a case-by-case basis. Across the 24 cases, there were no instances where the average income statement materiality judgments exceeded the average balance sheet materiality judgments for the seniors, managers, or partners.

Similarly, the average income statement and balance sheet materiality judgments were calculated for each firm on a case-by-case basis. Across the 24 cases, there were seven cases for which, on average at the firm level, the income statement materiality judgment exceeded the balance sheet materiality judgment; two firms were involved in these seven instances and both were structured firms.

Finally, the average income statement and balance sheet materiality judgments were calculated for the structured firms and for the unstructured firms, on a case-by-case basis. Across the 24 cases, there was only one case for which, on average at the audit technology level, the income statement materiality judgment exceeded the balance sheet materiality judgment; and that instance involved the structured firms. In other words, on average for the audit technology groups, the income statement materiality judgments were less than the balance
sheet materiality judgments for 23 of the cases for the structured firms and for all 24 of the cases for the unstructured firms.

These observations provide further (albeit ex post) justification that, for purposes of examining level of responsibility, firm, and audit technology effects, the income statement materiality judgments were appropriately viewed as the dependent variable of principal interest. In the preponderance of cases, the income statement materiality judgments, on average (across the three levels of responsibility, across the six firms, and between the two audit technology groups), would indeed form the basis for the auditors' "preliminary judgments about materiality."

Comments on Statistical Assumptions

The analyses reported above dealt with systematic differences among the auditors with respect to location (i.e., central tendency) and variation (i.e., dispersion) in their materiality judgments. Analysis of variance was the primary statistical method used to identify differences as to location. Cochran's test for homogeneity of variance was the primary method used to identify differences as to variation. Both are parametric procedures involving certain distributional assumptions about the data.

As discussed in Chapter III, ANOVA assumes that the data are normally distributed with independent error components having
homogeneous variances. Cochran's test for homogeneity of variance assumes that the data are normally distributed.

Keppel [1982, p. 87] and Kirk [1982, p. 75] point out that ANOVA is robust with respect to its assumptions. The F-test (used to establish significance in ANOVA) is not sensitive to the assumptions of normality and homogeneity, especially when the sample sizes among the comparison groups are equal. (By design, the sample sizes of all comparison groups in this study were equal.) In view of the robustness of ANOVA, it is unlikely that the results of the analyses were influenced by departures from the methods' underlying assumptions.

However, as a cross-check, nonparametric Wilcoxon rank sum tests were used to further examine the primary significant findings with respect to location. The existence of firm and audit technology differences were corroborated by the nonparametric tests (reported in footnotes #5 and 11, respectively).

The Wilcoxon rank sum test assumes that the observations are independent. In that regard, subjects were instructed to work on the experiment alone. No evidence was found to suggest that subjects failed to work independently. For example, the plots of the data do not indicate collusion among the subjects; and review of subjects' original responses did not indicate any tendency for pairs or groups of subjects to provide identical responses.
The primary significant findings with respect to variation were corroborated in two ways. Differences at the firm and audit technology levels were also identified (as reported in footnote #9) using the conservative Bonferroni simultaneous confidence intervals, involving all possible comparisons (i.e., ratios) of the firms' sums of squares. Since these ratios constituted F-tests, the above remarks with respect to the robustness of the F-test apply with respect to the tests of variation, as well. Again, distinct firm and audit technology effects were observed. Finally, nonparametric Wilcoxon rank sum tests were used to further examine the differences in variation at the firm and audit technology levels (reported in footnotes #8 and 12, respectively). Similar results obtained.

Based on the robustness of the methods used (i.e., the F-tests used in the ANOVAs with respect to location and used to corroborate the results with respect to variation) and based on the substantiation provided by the nonparametric tests, no further analyses related to the underlying statistical assumptions are warranted. The results were consistently observed, and apparently do not hinge on the distributional assumptions underlying the parametric methods that were used in the data analysis.
CHAPTER IV FOOTNOTES

1. The tests for differences as to location dealt with the mean standardized materiality scores, across levels of responsibility, firms, and audit technology groups. Although these scores were averaged across groups of subjects, the standardized materiality scores were obtained by summing across the 24 cases for each subject. For example, a subject whose materiality judgment was one standard deviation above the overall mean for each case would have a standardized materiality score equal to 24. Another subject whose materiality judgment was one-half standard deviation above the overall mean for each case would have a standardized materiality score equal to 12. These two subjects would not be 12 standard deviations different in their materiality judgments, but, rather, they would differ by half a standard deviation on each of 24 cases.

2. In addition, the absolute values of the standardized values were summed for each subject, and analyses were performed on those amounts. Because those results were virtually the same as the results which allowed positive and negative items to offset for each individual, no further discussion will be made of the analyses based on the absolute values. Plots of the standardized scores, by case for each subject, suggested that individuals tended to be fairly consistently high or low across cases.

3. Reference to Figure 7 indicates that the analysis of differences associated with level of responsibility on a firm-by-firm basis, is based on a relatively small number of observations—there are four subjects at each level for each firm. Accordingly, this analysis is particularly susceptible to the presence of relatively extreme observations. Although the data were examined subjectively for the purpose of identifying conspicuously high or low materiality judgments, no formal tests were conducted to identify "outliers." However, nonparametric Wilcoxon rank sum tests were performed when the parametric procedures were significant with respect to the primary research hypotheses (regarding level of responsibility, firm affiliation, and audit technology). Corroboration by nonparametric tests would provide evidence that the parametric procedures are not driven by the extremeness of a few observations.

4. Within the basic hierarchy of the Big Eight CPA firms, one might identify four distinct levels of responsibilities--staff, senior, manager, and partner. Staff subjects were not included in this study, since the contact persons indicated that generally staff would not be involved in the real world analogue of the experimental
task. The seniors, managers, and partners were identified by the codes "2," "3," and "4," respectively, related to the basic hierarchy of the firms.

A few subjects were visibly more extreme than others in their materiality judgments, as demonstrated by Figure 8. However, the seniors, managers, and partners overall were not apparently affected differentially due to these individuals. (Further, it should be noted that there are 24 subjects at each level of responsibility, and individual differences that are not too extreme will be mitigated when averaged across the other subjects within each level of responsibility.) None of the observations are conspicuously extreme among the seniors, managers, and partners. Accordingly, the results of the analyses are probably not greatly influenced by "outliers."

5. Similar results were found using (nonparametric) Wilcoxon rank sum tests. Within firm #2, the judgments of the seniors were significantly higher than those of both the managers and partners (p = .014 in each instance). These results suggest that the (parametric) ANOVA results are not heavily influenced by the presence of a few relatively extreme observations.

Table 8 indicates that the seniors from firm #2 had fewer years of experience, on average, than those from each of the other firms. However, based on pooled t-tests, there were no significant differences in years of experience for the seniors across the six firms at the .01 level. (Since there were five comparisons involving firm #2, the .01 level for each comparison corresponds to an overall .05 significance level.) Conceivably, judgments of the seniors from firm #2 may be influenced by a relative lack of experience in making planning stage materiality judgments. To the extent that the experimental task involved subjects' "rules of thumb" with respect to materiality judgments at the aggregate financial statement level, these subjects' heuristics may be less well defined than other subjects having more practice with the real world analogue of the experimental task.

6. A two-way ANOVA encompassing the main effects for level of responsibility and firm affiliation (as well as the level-firm interaction) was conducted. The F-values (p-values) for level of responsibility, firm affiliation, and the level-firm interaction were as follows: 1.22 (.3041), 11.92 (.0001), and 2.17 (.0340).

At first glance, the significant interaction might suggest that firm effects cannot be discussed meaningfully without considering specific levels of responsibility. However, the significant interaction resulted entirely from the previously reported significant effect (regarding the seniors) in firm #2.
7. The responses at the firm level are not identified by name due to a commitment to one firm to maintain their anonymity. The specific nature of this commitment was that the individual firms' "performance" in this experiment would not be discussed. Nevertheless, it is proper to point out that there was a high degree of correlation between the firms' mean standardized materiality scores and the measures of structure calculated by Kinney [1986], as reported by Morris and Nichols [1986, p.6]. Specifically, the Spearman (nonparametric) test statistic was calculated (Hollander and Wolfe [1973, p. 191]) between the rankings of the observed means for the income statement judgments for each firm and Kinney's measures of structure. The coefficient was very high, approximately .943.

8. Nonparametric Wilcoxon rank sum tests were applied to the twelve standardized materiality scores for each firm. All pairwise comparisons were conducted among the six firms—hence, fifteen comparisons were made. The large sample approximation of the Wilcoxon rank sum test's distribution (with m=n=12) was used (Hollander and Wolfe [1973, p. 68]). Even using a conservative .001 significance level per comparison (i.e., .001 x 15=.015), four of the fifteen comparisons were significant. Firm #1 was different than firms #4, 5, and 6; similarly, firm #3 was different than firm #4.

These nonparametric Wilcoxon rank sum tests substantiate the parametric tests that there were systematic differences across firms in their materiality judgments. There were differences in the median thresholds as well as differences in the mean thresholds, associated with firm affiliation. An important benefit of this nonparametric tests is that no distributional assumptions are involved. Consequently, this provides strong evidence that the results of the parametric tests do not hinge on the assumptions underlying ANOVA.

9. In another examination of mean differences among the firms, all possible between-firm comparisons were made using two sample t-tests (Miller [1986, p. 55]). (Since fifteen comparisons resulted, a conservative level of significance was selected based on Bonferonni simultaneous confidence intervals=.0025 per comparison times 15 comparisons equals .0375 overall level of significance.) None of the comparisons between two structured firms or between two unstructured firms were significant; however, six of the nine comparisons between a structured and an unstructured firm were significant (and another could be viewed as borderline significant). Apparently, differences among the firms in their mean materiality judgments were related to the firms' audit technologies. This issue will be discussed in more depth in a subsequent section entitled, "Audit Technology Effects."

10. Table 12 reports the variances of the standardized materiality scores, calculated by firm and by level of responsibility.
Nonparametric Wilcoxon rank sum tests were applied to these variances. Specifically, all pairwise comparisons were made across the six firms using Wilcoxon rank sums (there were three observations per firm, representing the variances of the seniors, managers, and partners for each firm). Firm #1 was significantly different than each of the other firms (p ≤ .05); none of the other firms were significant at the .05 level. This provides further evidence of differences at the firm level in variation, and corroborates the parametric results reported in the text. An important aspect of this nonparametric result is that no distributional assumptions are required.

11. Further evidence of firm differences as to variation resulted from all pairwise comparisons using the ratios of the firms' sums of squares. A total of fifteen comparisons resulted; three were comparisons between structured firms, three were comparisons between unstructured firms, and nine were comparisons involving one structured and one unstructured firm. Significance was based on the conservative Bonferroni simultaneous confidence intervals involving the .05 level.

None of the comparisons among the structured firms and none of the comparisons among the unstructured firms were significant at the .05 level. However, seven of the nine pairwise comparisons between a structured and an unstructured firm were significant (and another could be viewed as borderline). This provides strong evidence that firm differences in variation were closely related to their audit technologies. This issue will be discussed in more depth in a subsequent section, entitled "Audit Technology Effects."

12. Chapters II and III discussed the Cushing and Loebbecke [1983] and Kinney [1986] papers with respect to audit technology. As noted in Chapter III, Kinney classified Deloitte, Haskins, & Sells; Peat, Marwick, Mitchell; and Touche Ross as the "structured" firms, and they are so classified here. Further, Kinney classified Coopers & Lybrand and Price Waterhouse as "unstructured" firms, and they are so classified here. In addition, Kinney classified Arthur Andersen, Arthur Young, and Ernst & Whinney as "intermediate" firms. As explained in Chapter III, the latter two firms were not involved in this study, and Arthur Andersen was included among the unstructured firms. (Based on a review of the firms' audit manuals and discussions with each "contact person," Arthur Andersen's policies with respect to planning stage materiality judgments appeared to be similar to those of Coopers & Lybrand and Price Waterhouse, at the time the experiment was conducted.)

This dissertation relies on Kinney's classification to compare the most highly structured firms with others viewed as less structured. The classification of Arthur Andersen among the "unstructured" firms is somewhat subjective, and the cut-off between
the "structured" and "unstructured" firms along the continuum related to measures of structure is somewhat arbitrary. A "sensitivity analysis," as such, was not conducted; the results related to audit technology might be affected by including Arthur Andersen among the "structured" firms. Further, within a particular audit technology classification, there can be no guarantee that the firms have identical degrees of structure.

13. The large sample approximation of the Wilcoxon rank sum test (Hollander and Wolfe [1973, p. 68]) was applied to the 72 standardized materiality scores. Audit technology was highly significant (p ≤ .0001). The structured firms had both higher mean and median standardized materiality scores.

The nonparametric Wilcoxon rank sum test substantiates the parametric ANOVA test that there was a systematic difference between the structured and unstructured firms in these materiality judgments. An important benefit of this nonparametric test is that no distributional assumptions are involved. Consequently, this nonparametric result provides strong evidence that the results of the parametric tests do not hinge on the assumptions underlying ANOVA.

14. Table 12 reports the variances of the standardized materiality scores, calculated by firm (where firms #1-3 are the structured firms and firms #4-6 are the unstructured firms) and by level of responsibility. Nonparametric Wilcoxon rank sum tests were used to assess whether the structured and unstructured firms differed in variability, based on the nine variances (i.e., three levels of responsibility for each of three firms) within each technology group. The large sample approximation of the Wilcoxon rank sum test (Hollander and Wolfe [1973, p. 68]) was used. Again, the structured firms were shown to be significantly more variable (p=.004).

This provides further evidence of differences between the two audit technologies in variation, and corroborates the parametric results reported in the text. An important aspect of the nonparametric result is that no distributional assumptions are required.

15. The analyses regarding risk attitudes are discussed only in conjunction with the auditors' income statement materiality judgments, since those materiality judgments are of primary interest.

16. The tests reported in Chapter IV with respect to differences among the participants in their risk attitude measures are based on the comparison of subjects in the upper and lower one-third threshold groups. Separate one-way ANOVA's were also conducted across all 72 subjects with respect to level of responsibility, firm affiliation,
and audit technology; no systematic differences in the risk attitude measures were identified at the .05 level.

17. Similar correlations were calculated by level of responsibility and by firm affiliation. By level of responsibility, the correlation coefficients (p-values) for the seniors, managers, and partners were, respectively: -.08440 (.6950), .40335 (.0506), and .28073 (.1839). By firm, the coefficients (p-values) for firm #1-6 were, respectively: .21125 (.5099), .08309 (.7974), .47506 (.1186), .19222 (.5495), -.08212 (.7997), .17512 (.5862). None of the coefficients, by level of responsibility or by firm, were significant at the .05 level. In addition, nearly all of the coefficients were opposite the direction hypothesized.

18. This section discusses the individual ANOVA models with respect to the income statement materiality judgments. Individual ANOVA models were also constructed with respect to the balance sheet materiality judgments. Those are discussed in Appendix D.

19. The firm's guidelines emphasize owners' equity, a balance sheet concept, in the determination of materiality at the financial statement level. Since the case materials provided the breakdown of the balance sheet in percentage terms, owners' equity could be ascertained from the total assets variable.

20. The possibility must be acknowledged that the extremeness of these standardized scores may be an artifact of the experiment. To the extent that firm #1, a structured firm, bases the planning stage materiality judgment on owners' equity (a balance sheet amount), it may be particularly abstract for subjects of firm #1 to express separate judgments with respect to the income statement and balance sheet. In addition, this firm may be affected more than the others by the orthogonal experimental design. In the experiment, total assets, income, and sales were uncorrelated as a consequence of the full factorial design. However, the "contact person" (a partner) stated that his colleagues would be able to provide meaningful responses based on his review of the experimental materials. In addition, none of the subjects from firm #1 expressed difficulty with the experimental task in the survey questionnaire.

21. This section discusses the analyses with respect to the repeated cases for the income statement materiality judgments. Similar analyses were performed with respect to the repeated cases for the balance sheet materiality judgments. Those are discussed in Appendix D.

22. The survey questionnaire (Appendix C) contained a few questions which related to the level of interest subjects had toward
the experiment. In particular, questions #20, 21, 22, and 27 were included for that purpose.

Question #20 asked subjects to indicate the degree to which they identified with their assigned roles (on a four-point scale where "1" meant "did not identify at all" and "4" meant "identified strongly"). Across the 72 subjects, the overall average was 2.2; the averages for the seniors, managers, and partners were 2.0, 2.6, and 1.9, respectively.

Question #21 asked subjects to evaluate the similarity of the experimental task relative to real world audit planning (on a five-point scale where "1" meant "very dissimilar" and "5" meant "very similar"). The overall average was 2.9; for the seniors, managers, and partners, the averages were 3.0, 3.3, and 2.5, respectively.

Question #22 queried the subjects about the realism of the experimental information (on a five-point scale where "1" meant "very unrealistic" and "5" meant "very realistic"). Overall, the average was 3.0; for the seniors, managers, and partners, the averages were 3.2, 3.3, and 2.4, respectively.

Question #27 asked the subjects whether they wished to receive written results following the analysis of the results. A "yes" response was viewed as an indication of relative interest in the experiment. Overall, 64 percent responded affirmatively; for the seniors, managers, and partners, the affirmative responses were 71, 54, and 67 percent, respectively.

Although these responses are not conclusive, they are viewed as indicating a reasonably strong interest by the subjects in the nature of the experiment. Comparisons were made across firms and between the two audit technologies with respect to these survey questions and nothing was noted to suggest systematic differences on these dimensions.

23. A further test regarding the two orders of presentation examined for any tendency among the subjects to become stricter (or more lenient) as they proceeded through the 24 cases. Since order #2 was the reverse of order #1, this test for "drift" was accomplished by arranging the responses in the same order. Then, on a case-by-case basis, the average materiality judgment (based on the "raw" responses) was calculated, separately for the subjects receiving order #1 and order #2. A ratio was calculated for each case where the mean response for the subjects receiving order #1 was divided by the mean response for the subjects receiving order #2. A tendency to become stricter, for example, would be shown by a plot of these ratios that decreases from left to right (where the value of the ratio is
represented by the y-axis and the case number is represented by the x-axis).

The plot of the calculated ratios did not indicate any tendency for the subjects to become systematically stricter (or more lenient) as they completed the experiment. Further, Kendall's nonparametric test for independence (Hollander and Wolfe [1973, pp. 185-6]) indicated that the ratios were not significantly associated with the order of case presentation at the .05 level.
CHAPTER V -- REVIEW AND SUMMARY OF RESULTS, CONCLUSIONS, LIMITATIONS, AND EXTENSIONS

Review and Summary of Results

A judgment-capture experiment was conducted to examine (Big Eight) auditors' planning stage materiality judgments. The study focused on certain factors believed to be at least potentially associated with differences in these judgments. Specifically, the judgments were examined for systematic effects associated with level of responsibility (i.e., senior, manager, or partner), firm affiliation, and audit technology, related to firm affiliation. A secondary research issue investigated whether there were attitudinal differences, with respect to risk attitudes, associated with differences in auditors' planning stage materiality judgments.

A total of 72 auditors (from the Columbus, Ohio, offices of six of the Big Eight CPA firms) participated in the study. There were 24 subjects each at the senior, manager, and partner levels. Six Big Eight firms participated; each firm provided twelve participants (four from each of the three levels of responsibility). Three of the firms were viewed as highly structured and were classified as "structured"; the other three firms were viewed as less structured and were classified as "unstructured." These classifications as to audit technology relied on Kinney [1986].
SAS No. 47 indicates that the auditor should consider the preliminary judgment about materiality at the planning stage of an audit engagement. Further, the SAS suggests that the judgment may be made at the aggregate financial statement level; the preliminary judgment about materiality would then be the smallest amount material to any one of the financial statements. On that basis, subjects were asked to specify the appropriate planning stage materiality threshold at the aggregate financial statement level, separately with respect to the income statement and balance sheet, for each of 24 different cases. The 24 cases represented a full factorial design related to three financial statement variables--income (loss) before taxes (with four levels), net sales (with two levels), and total assets (with three levels).

Based on prior research, primarily involving auditors' evaluation stage materiality judgments, and based on the professional literature with respect to auditors' planning stage materiality judgments, the income statement materiality judgments constituted the dependent variable of principal interest. Accordingly, the results discussed in Chapter IV dealt with these income statement materiality judgments. (Similar analyses were conducted with respect to the balance sheet materiality judgments for exploratory purposes, and those results were presented in Appendix D for the interested reader.)

In order to mitigate the conspicuous heterogeneity of variance across the individual cases, and to provide a single, independent
measure of the overall extremeness of each individual's materiality judgments, subjects' responses were "standardized." On a case-by-case basis, the responses were transformed to have a distribution with mean zero and variance one, across all 72 subjects. The standardized scores were summed across all 24 cases for each subject to obtain the subject's standardized materiality score.

Consensus was emphasized, due to the absence of a criterion variable by which to assess the "correctness" of these judgments. Two separable dimensions of consensus were examined (1) location (regarding central tendency) and (2) variation (regarding dispersion).

Two hypotheses were stated with respect to overall effects associated with level of responsibility. H1.1 dealt with differences among the three levels as to location, and H1.2 dealt with differences as to variation.

In order to separate the possible simultaneous effects of responsibility and firm affiliation, the analysis of responsibility effects was conducted on a firm-by-firm basis. The results associated with the individual firms were then combined into an overall test statistic, using Fisher's method for combining separate tests.

With respect to location, six separate one-way ANOVAs (one for each firm) indicated that level of responsibility was significant (at the .05 level) for only one firm. In that apparently isolated instance, the seniors of firm #2 had higher materiality thresholds than the managers and partners. As a result, the overall test
statistic, based on Fisher's method, was significant at the .05 level (but was not significant at the .01 level). Across the firms, differences among levels of responsibility were not pervasive. The results did not suggest any real differences among the three levels of responsibility, and, accordingly, H1.1 was not rejected.

With respect to variation, no significant differences were detected among levels of responsibility for any of the firms, using Cochran's test for homogeneity of variance. Not surprisingly, therefore, the overall test statistic, based on Fisher's method, was not significant at any reasonable level. Accordingly, H1.2 was not rejected, either. Consequently, the seniors, managers, and partners were viewed as homogeneous for purposes of investigating firm effects and, related to that, audit technology effects.

Two hypotheses were stated with respect to overall firm effects. H2.1 dealt with differences among the six firms as to location, and H2.2 dealt with differences as to variation.

With respect to location, overall firm differences were conspicuous. Based on a one-way ANOVA, firm affiliation was highly significant (F=10.07, p ≤.0001). In particular, firm #1 was identified as having a significantly higher mean materiality threshold than each of the other firms. On that basis, the null hypothesis for H2.1 (that there were no differences in location among the six participating firms) was rejected. Interestingly, each of the
structured firms had a higher mean than each of the unstructured firms.

With respect to variation, overall firm differences were also identified. Cochran's test for homogeneity of variance, applied to the sums of squares associated with each of the firms, was significant at the .01 level. In addition, similar results obtained from all pairwise comparisons, i.e., F-tests, using ratios of the firms' sums of squares, based on conservative Bonferroni simultaneous confidence intervals. Accordingly, H2.2 was rejected; there were real differences in variation at the firm level.

To better understand differences at the firm level, the study considered the firms' differing audit technologies. Audit technology dealt with the philosophical differences among these Big Eight firms in the extent to which the national offices should impose relatively specific guidelines (i.e., "structure") on the individual practice offices and auditors throughout the firm. Two hypotheses were stated with respect to audit technology effects, related to firm affiliation. H3.1 dealt with differences between the structured and unstructured firms with respect to location; H3.2 dealt with differences between the two audit technology groups with respect to variation. The investigation of audit technology effects involved partitioning the subjects into two groups, the structured firms and the unstructured firms.
With respect to location, a difference in the mean thresholds between the two audit technology groups was found. Based on a one-way ANOVA, audit technology was highly significant ($F=32.69$, $p \leq .0001$).

On that basis, the null hypothesis for H3.1 (that there were no differences in location between the structured and unstructured firms) was rejected. That the structured firms had significantly higher materiality thresholds was not surprising, since the analysis of firm effects indicated that each of the structured firms had higher mean standardized materiality scores (although the differences were not necessarily significant) relative to each of the unstructured firms.

In view of the relatively uncomplicated judgment context (i.e., relative to the "real world"), this result is particularly striking.

With respect to variation, a difference between the two audit technology groups was also identified. Cochran's test for homogeneity of variance was applied to the sums of squares associated with each of the audit technology groups. The participants affiliated with structured firms were significantly more variable in their judgments than the participants affiliated with unstructured firms ($p \leq .01$).

Further compelling evidence of audit technology differences came from all pairwise comparisons using ratios (i.e., $F$-tests) of the firms' sums of squares. Of the fifteen possible pairwise comparisons, none of the six comparisons within the same audit technology group was significant. However, seven of the nine comparisons between firms
having different technologies were significant at the .05 level (based on Bonferonni simultaneous confidence intervals), and another could be viewed as borderline. Accordingly, H3.2 was rejected.

A secondary purpose of this study was to explore why subjects may specify different materiality thresholds. In that regard, a questionnaire was developed, based on the expectancy-valence construct, to identify systematic differences among subjects with respect to risk attitudes, hypothesized to be related to differences in their materiality judgments. The notion of risk attitude was based on the tradeoff of Type I and Type II errors. Subjects were questioned as to their "beliefs" (regarding likelihoods on a seven-point scale) and their "evaluations" (regarding importance on a five-point scale) for each item (a specified outcome) associated with Type I and Type II errors. To measure each subject's relative sensitivity to Type II versus Type I errors, the expectancy-valence measure for Type II errors was divided by the measure for Type I errors.

Two specific hypotheses were stated. H4.1 dealt with systematic differences among the participants with respect to risk attitudes. H4.2 dealt with the association between the risk attitude measures and the extremeness of the auditors' materiality thresholds. In particular, auditors who were relatively more sensitive to Type I errors (e.g., practice development considerations) were expected to have relatively high materiality thresholds. Auditors who were relatively more sensitive to Type II errors (e.g., litigation from
financial statement users) were expected to have relatively low materiality thresholds.

To construct a fairly powerful test to identify differences among subjects as to their risk attitudes and to measure the association with their materiality judgments, subjects were partitioned into thirds, based on the extremeness of their standardized scores. The subjects in the upper and lower thirds were compared for systematic differences in risk attitudes, and their risk attitudes were correlated with their standardized scores. However, the subjects did not differ systematically with respect to the measure of risk attitude adopted. (Further, across all 72 subjects, risk attitudes did not differ when comparisons were made in terms of level of responsibility, firm affiliation, and audit technology.) Hence, H4.1 was not rejected.

Similarly, the measure of risk attitude adopted in the study was not significantly associated with auditors' materiality thresholds at the .05 level. Hence, H4.2 was not rejected. As a result, the measure of risk attitude adopted was not helpful in explaining systematic differences among the subjects with respect to their planning stage materiality judgments.

To further investigate the issue of individual differences with respect to the materiality judgments, an ANOVA model was constructed for each subject across the 24 cases. Even in the structural forms of the ANOVA models (encompassing each main effect and each two-way
interaction), substantial individual differences were noted. Income
(loss) before taxes ("income") was significant for about two-thirds of
the subjects. Total assets was a significant variable for one-third
of the subjects; and net sales was significant for about twenty
percent of the subjects. Of the 47 subjects for whom only one
variable was significant, income was significant for 29, net sales was
significant for eleven, and total assets was significant for seven. A
firm difference was identified with respect to an emphasis on total
assets. All seven subjects for whom total assets was the only
significant variable were affiliated with one firm; and six of these
had the highest standardized materiality scores of the 72 subjects.
This suggested that differences in subjects' materiality judgments
were influenced by differences in the structural models with respect
to the cues they utilized.

The experimental materials asked the subjects to identify the
business risk (i.e., exposure) to the auditors which they associated
with each case. The purpose was to identify a potential confounding
factor, since differences in the perception of business risk could
affect subjects' materiality thresholds. Specifically, all else
equal, it is believed that perceptions of greater exposure will
influence auditors to establish lower materiality thresholds than they
would under conditions of perceived lower exposure. No systematic
differences in these business risk assessments were identified with
respect to level of responsibility, firm affiliation, or audit technology.

Two other analyses were conducted to assess the meaningfulness of subjects' responses. The first involved test-retest consistency and the second checked for order effects. No significant differences were detected, in the aggregate, between subjects' responses to the original cases and those to the four repeated cases. Further, no systematic differences were identified with respect to level of responsibility, firm affiliation, or audit technology. Similarly, no differences were detected between the two orders of case presentation, either in terms of location or variation. Apparently, the subjects were able to make meaningful responses; at least, no evidence was obtained to the contrary.

Finally, comparisons were made, on a case-by-case basis using subjects' untransformed responses, between the income statement and balance sheet materiality judgments to explore the extent to which, on average (in terms of level of responsibility, firm affiliation, and audit technology), the auditors' preliminary judgments about materiality were driven by income statement considerations in the context of the experiment. For the preponderance of the cases, the average income statement judgment for each level of responsibility, for each firm, and for both audit technology groups, was lower than the corresponding balance sheet judgments. These observations were consistent with the expectation that the income statement would be the
usual frame of reference for most auditors and that income statement considerations would drive the preliminary judgments about materiality. This provided further (albeit _ex post_) support for viewing the income statement judgments as the dependent variable of primary interest.

**Conclusions**

The analyses reported in Chapter IV document substantial individual differences among the participants in this study, involving a decision context much less complex than real-world settings. Individual differences were expected, based on prior studies of auditors' (evaluation stage) materiality judgments. This study looked for systematic differences in terms of level of responsibility, firm affiliation, audit technology, and risk attitude.

Within the context of this study, the seniors, managers, and partners did not differ overall in their materiality judgments, either as to location or variation. Accordingly, seniors, managers, and partners may be viewed as largely homogeneous groups with respect to their planning stage materiality judgments.

Apparently, subjects have developed their mechanisms (e.g., "rules of thumb") for making preliminary judgments about materiality by the time they have responsibility for supervising audit engagements as seniors. There is apparently little modification of these mechanisms following promotion to manager and partner. This result
may not be surprising, since SAS No. 47 implies that the planning stage judgments depend primarily on quantitative considerations, while the evaluation stage judgments are influenced to a greater extent by qualitative considerations. Others (e.g., Woolsey [1954a] and Pattillo [1976]) have suggested that, in an evaluation stage context, there may be a range such that an item below some threshold is clearly immaterial, an item above some other threshold is clearly material, and an item between those two thresholds must be evaluated with respect to its qualitative characteristics in order to assess materiality. The experimental task may have dealt with such quantitative thresholds, and subjects' responses may have been determined by their heuristics adopted to assist them in their tentative materiality judgments. Perhaps level of responsibility (presumably related to experience) plays a greater role in identifying and incorporating the various qualitative considerations in an evaluation stage materiality judgment when the quantitative considerations are ambiguous.

Substantial differences were identified at the firm level among the six participating Big Eight firms, both as to location and variation. Prior studies of auditors' (evaluation stage) materiality judgments have pointed out differences between national and non-national (regional) firms or, alternatively, between Big Eight and non-Big Eight firms. The Big Eight firms have been viewed as so similar as to be regarded as homogeneous organizations. This study
demonstrates empirically that the Big Eight firms are not homogeneous with respect to their planning stage materiality judgments. This conclusion has important implications to auditing researchers. Studies of audit-judgment involving Big Eight firms should consider firm differences and differing audit technologies as factors which may influence subjects' judgments. If not considered, these factors may confound the results of such studies. Obviously, the consequences of assuming that the Big Eight firms are homogeneous depend on the specific focus of the research involved.

These differences among the Big Eight firms underscore the difficulties of the standard-setting bodies concerned with establishing more specific materiality guidelines. Achieving widely accepted guidelines with reasonable specificity would indeed be a challenging task, when there are distinct differences even among the supposedly homogeneous Big Eight firms. Yet, if the ASB develops specific materiality guidelines, they would most likely be directed at the planning stage. The professional literature suggests that the planning stage judgments are primarily based on quantitative considerations, whereas the evaluation stage judgments are influenced to a greater extent by qualitative considerations.

The differences identified at the firm level were themselves systematic and related to the firms' audit technologies. The structured firms differed from the unstructured firms both as to
location and variation. The structured firms had higher thresholds and greater variability relative to the unstructured firms.

That the structured firms had higher materiality thresholds was not surprising. By imposing firm-wide guidelines throughout the organization, the structured firms were expected to manage the risks associated with Type II errors (i.e., a portfolio approach to risk management), while allowing them to have relatively higher materiality thresholds. However, to employ that explanation of their higher materiality thresholds, the structured firms must have lower variability at the same time.

Therefore, the greater variability of the structured firms relative to the unstructured firms was surprising, and one can only speculate about why the structured firms were more variable in their judgments. Perhaps the nature of the judgment task was not amenable to the structured firms' guidelines. The experimental task required subjects to make their materiality judgments at the aggregate financial statement level, separately for the income statement and balance sheet. The structured firms' guidelines do not make the distinction between the income statement and balance sheet judgments. However, the contact people at the structured firms were explicit that such a distinction would not be problematic for participants from their firms.

Alternatively, the participants from the structured firms may have departed from their firm guidelines and responded as they
individually deemed appropriate, since the context for the judgments was a research project and not an actual client situation (and therefore not subject to review, etc.). However, such an explanation would make more sense if the structured and unstructured firms were equally variable. That explanation does not clarify why the structured firms were more variable than the unstructured firms, who do not provide their auditors with detailed guidelines in the first place.

A related possibility is that, within the structured firms, some of the subjects followed the firm guidelines, while others departed substantially from the guidelines. Those that followed the firm guidelines may have specified relatively high materiality thresholds (consistent with the earlier explanation regarding the higher means expected for the structured firms), while yielding greater within-firm variability in comparison with those who deviated from the firm guidelines and specified relatively lower materiality thresholds. Why some subjects would adhere to the firm guidelines and others would not, in the context of the experiment, is unclear. Figure 9, which plots the standardized materiality scores by firm affiliation, does not suggest particular clusters of judgments within the structured firms. Instead, the standardized scores were fairly widely dispersed within each of the structured firms.

Whatever the cause (e.g., whether caused by a judgment task not amenable to firm guidelines, or by subjects who, for other reasons,
departed from those guidelines), that subjects from structured firms may become even more variable in their judgments than other auditors from firms not having such detailed guidelines in the first place is an interesting consideration. This issue has important practical consequences to the structured firms, since such a result is obviously inconsistent with the ostensible purpose of imposing a more structured approach within a firm.

The notion of differing risk attitudes (as an intervening variable) was not helpful in understanding why subjects differed in their materiality thresholds. The overall risk attitude measures were not significantly correlated with subjects' standardized materiality scores. Similarly, the separate measures related to Type I and Type II errors were not significantly associated with the materiality judgments. Nevertheless, there were a few significant differences (regarding specific items incorporated into the aggregate risk attitude measure) identified between the subjects having the highest and lowest materiality thresholds, but the differences were not clear-cut.

One can only speculate about why there was a tendency for subjects' risk attitude measures, related to Type I and Type II errors, to be in directions opposite to that hypothesized. (Subjects with relatively low materiality thresholds were expected to have higher measures related to Type II errors; instead, the subjects with high materiality thresholds had higher expectancy-valences related to
Type II errors.) One explanation is that the observed directionality resulted from chance, since the risk attitude measure was not significant at the .05 level.

Perhaps risk attitude was insignificant because of the way it was operationalized. Risk attitude was defined in terms of the tradeoff of Type I and Type II errors. These tradeoffs may not be salient issues with respect to auditors' planning stage materiality judgments, particularly if these judgments are tentative and precede such qualitative considerations as those implied by Type I and Type II errors. In that regard, the tradeoff between Type I and Type II errors may have stronger implications with respect to differences among auditors' evaluation stage materiality judgments (e.g., in instances of disagreement with the client).

Another possibility is that the lack of significance and curious directionality may have resulted from the nature of the investigation. Palmerino, et al. [1984, p. 186] cautioned, "Attitudes ... have greater influence upon behavior when a person is mindless .... Laboratories may induce self-consciousness in our subjects and may be thought-provoking and habit-inhibiting." Therefore, the questionnaire may have failed to elicit subjects' true attitudes.

The next section identifies the major limitations of the study. These limitations, although not unusual relative to other experimental studies involving audit-judgment, represent caveats which should be kept in mind when interpreting the results of the study.
Limitations of the Experiment

Experience was operationalized as level of responsibility, i.e., senior, manager, or partner. The null hypotheses related to H1.1 (regarding differences as to location) and H1.2 (regarding differences as to variation) could not be rejected. Overall, level of responsibility was apparently unrelated to differences among the subjects' materiality judgments. Since experience could be operationalized in other ways (for example, based on years as an auditor, or based on the number of times actual planning stage materiality judgments have been quantified), there can be no assurance that other experience-related constructs would lead to the same results.

Firm affiliation involved the (six) individual Big Eight firms participating in the study. The tests of audit technology effects compared the three firms viewed as the most structured (the "structured" firms) relative to the three firms viewed as less structured (the "unstructured" firms). The classifications with respect to structure relied substantially on Kinney [1986]. However, the classification of Arthur Andersen among the "unstructured" firms was somewhat subjective, since the cut-off between the "structured" and "unstructured" firms along the continuum of firms' degrees of structure was, to some extent, arbitrary.

Moreover, the results cannot be extrapolated to the two Big Eight firms who did not participate, and the effect on the results of
including those firms (both having "intermediate" structure) is unknown. Further, the notion of "structure" is certainly a dynamic concept, and the rankings of the participating firms on that dimension may change over time. The specific audit technology effects identified by this dissertation may be altered if, for example, relatively unstructured firms impose greater degrees of structure in the future.

This dissertation emphasized systematic differences among Big Eight auditors in their planning stage materiality judgments, in terms of responsibility, firm affiliation, and audit technology. A secondary issue dealt with the notion of differing risk attitudes to explore why auditors might establish different materiality thresholds.

Factors other than those investigated in this dissertation may result in systematic differences among auditors in their planning stage materiality judgments. No pretense is made that level of responsibility, firm affiliation, audit technology, and risk attitudes are the only possible dimensions on which auditors' materiality judgments might systematically differ. There may be other factors, omitted from this study, which could conceivably influence the results obtained. The greater variability within the structured firms (when less variability was expected relative to the unstructured firms) and perhaps the positive correlation of the risk attitude measure (when a negative correlation was expected) could be consequences of such an omitted variable.
Similar to other behavioral experiments, this study may be criticized because it is not necessarily generalizable to the "real world" of auditing. Not all the variables potentially relevant to an auditor's planning stage materiality judgment could be incorporated, and no claim is made that the selected independent variables are comprehensive. In this study, the auditors were limited to three financial statement cues (although percentage information was provided about the composition of the income statement and balance sheet). In addition, an auditor would have much more detailed knowledge about an actual client than can be effectively conveyed about a hypothetical client.

Moreover, since a fixed-effects ANOVA model was used instead of a random-effects model, there can be no generalization beyond the specific levels chosen for the independent variables. Further, external validity may be sacrificed by employing an orthogonal experimental design. In real auditing situations, auditors deal with financial statement variables that are correlated; in the experimental materials, the cues were independent. Cues which have realistic magnitudes individually may not necessarily be realistic when combined across all levels of the other variables. However, subjects' comments on the survey questionnaire suggested that most subjects viewed the experimental materials as reasonably realistic. Only one subject, a partner with an unstructured firm, remarked that certain ratios were rather extreme for some combinations of the financial statement
variables. Based on their responses regarding business risk for each case, subjects apparently were not affected by the combinations which might be viewed as unusual.

Further limiting the generalizability of the study, the subjects were not randomly selected. Consequently, the results may not generalize to others in the participating offices. The subjects were obtained at the pleasure of the firms and were probably selected based on availability. However, there is no reason to believe that the participants were in any way unrepresentative of their offices. Similarly, since all the participants were from the Columbus offices of their firms, the results may not extend to other offices. However, the demographic information collected indicates a variety of educational backgrounds and a diversity of audit experience, encompassing other geographical areas.

In order to minimize the inconvenience to the participating CPA firms and thereby achieve the desired sample of twelve subjects per firm, the experimental materials were delivered to the contact person at each firm for distribution to the participants. As a result, some control over the conduct of the experiment was sacrificed. Participants were assumed to adequately understand the instructions to the experiment in order to provide meaningful responses. In addition, there can be no guarantee that the subjects completed the experiment independently, although they were instructed to work alone and no evidence of collusion was found.
There were a number of assumptions associated with analysis of variance and Cochran's test for homogeneity of variance, which were heavily used in this study. With respect to ANOVA, the assumptions focused on the residuals of the ANOVA models. The residuals were assumed to be independent across observations, and to be distributed normally with mean zero and constant variance. However, these assumptions were not viewed as critical to the interpretation of the results. The strong corroboration from the nonparametric analyses suggested that the results were not dependent on the distributional assumptions of the parametric analyses. Similarly, Cochran's test primarily assumed that the data were normally distributed. The strong corroboration obtained from F-tests (robust with respect to non-normality) involving all pairwise comparisons among firms' sums of squares and similar results based on nonparametric Wilcoxon rank sums, strongly suggested that the results were not dependent on the distributional assumptions underlying the parametric methods.

Finally, whether (or how) differences among auditors in their preliminary judgments about materiality will affect their performance of an audit (or their "final" materiality judgments) is unclear. Conceivably, differences among auditors in their materiality judgments at the planning stage may be reduced or eliminated at the evaluation stage, following fieldwork. Therefore, such differences may or may not imply "quality" differences. This issue is among numerous areas for future research.
Extensions

A number of suggestions for future research follow from this study, which is among the first to examine auditors' planning stage materiality judgments. Two suggestions deal with exploring whether the results reported above were artifacts of the experimental task. First, future research should replicate this experiment, but limit the number of cases which subjects must address to one or a very few. (This would also permit a richer information set to be provided to the respondents.) Second, another study should replicate this experiment, without distinguishing between the income statement and balance sheet materiality judgments. Such a follow-up study should elicit subjects' single "preliminary judgments about materiality" to investigate whether these subjects (particularly those affiliated with structured firms) may have been influenced by asking them to make separate materiality judgments with respect to the income statement and balance sheet. These two suggestions would provide evidence as to whether subjects affiliated with structured firms in this study departed from their firms' guidelines because of the nature of the experimental task.

Future studies might use subjects from different offices of the Big Eight firms to investigate whether the results of this dissertation extend to other locations. Although there is no reason to expect that different results would obtain, these results could not be generalized to other geographical areas. Similarly, future
research should include participants from other national firms and, perhaps, from non-national firms. Such an extension would provide insights into the influence of differing audit technologies (i.e., structured versus unstructured approaches) across a wider population of CPA firms than just the Big Eight firms.

Additional investigation is warranted with respect to understanding why auditors establish different materiality thresholds. As reported above, there may have been some connection in this study between the cues utilized and the extremeness of the thresholds. However, the notion of differing risk attitudes, as operationalized here, was not helpful in clarifying why auditors' judgments differed. This is still an area about which little is known.

Subsequent research should include additional quantitative and qualitative variables as factors influencing auditors' planning stage materiality judgments. As mentioned, this could be accomplished in conjunction with limiting the number of cases to one or a very few. For example, a potentially interesting variable involves the risk characteristics of the client organization. This could encompass the distinction between public and private companies, since there is evidence that public companies are perceived as yielding greater exposure to auditors, all else equal.

An important topic not addressed in this dissertation concerns the impact which different planning stage materiality judgments might have on the conduct of an audit. Prior research has not explored what
the practical consequences of different materiality thresholds, particularly at the planning stage, might be. Whether (and how) differences among auditors in their preliminary judgments about materiality affect their audit planning and their evaluation stage materiality judgments would be interesting topics for future research.

In general, more research is needed to identify whether the differences discussed above in the planning stage materiality judgments between structured and unstructured firms extend to other judgmental areas of auditing. Virtually nothing is known about the pervasiveness of these differences. The consequences within a structured firm when the firm's detailed guidelines are not amenable to a specific decision context would be particularly interesting to explore further. Conceivably, even greater variation might result among participants from the structured firms under those circumstances, relative to subjects from the unstructured firms who did not have such detailed guidelines in the first place.
APPENDIX A

INSTRUCTIONS AND BACKGROUND INFORMATION

Introduction

This experiment deals with the quantification of planning stage materiality judgments (termed "preliminary judgments about materiality") as described by SAS No. 47, "Audit Risk and Materiality in the Conduct of an Audit." As discussed in the SAS, materiality judgments arise in planning the audit engagement and in evaluating the accumulated audit evidence. This experiment focuses on the materiality judgment at the planning stage. Although prior "laboratory" experiments have reported differences in auditors' materiality judgments, no attempt has been made to understand why, within the context of those experiments, the judgments were different. Therefore, this experiment and the related questionnaires address some reasons why auditors may specify different materiality thresholds in such experiments.

In addition to planning stage materiality judgments, this experiment asks you to estimate the required audit hours (by staff, senior, manager, and partner) based on the information provided. Further, you are asked to express your perception of "business risk" (on a seven-point scale where "1" represents very low and "7" represents very high) for each of the individual cases. You may
recall that SAS No. 47 (in footnote #1) referred to business risk when it stated:

...the auditor is also exposed to loss or injury to his professional practice from litigation, adverse publicity, or other events arising in connection with financial statements that he has examined and reported on. This exposure is present even though the auditor has performed his examination in accordance with generally accepted auditing standards and has reported appropriately on those financial statements.

The materials for the experiment are presented in two booklets. This introduction is in the first booklet, which also contains the instructions and some background information about a hypothetical client and the client's industry. This same background information is applicable to each of the specific cases included in the second booklet. The background information is based largely on current trade journals and financial publications, and reflects industry conditions based on publicly available data through 1985. This information is included to provide a common reference point. You may refer to it as often as you wish. In addition, you may refer to any professional or firm materials that you desire to consult.

The second booklet presents the specific situations for which you are asked to provide certain audit-related judgments (regarding materiality, audit hours, and business risk). Also, two questionnaires are included to gather additional information.

It was not feasible to include in these materials all the information that might possibly influence your judgments. (For example, prior years' financial statements and audit hour information
is not provided; and the information that is provided lacks much of the detail that would be available with respect to an on-going audit client.) However, in planning actual engagements, auditors may sometimes find that some desired information is unavailable. Moreover, in this experiment, the relative levels of your judgments are of more interest than the absolute levels. Consequently, these materials invoke an assumption of "all else constant" to the extent some information has been excluded. It is understood that judgments, such as those you are being asked to make, are almost always tentative and are revised as additional relevant information becomes available.

Even though the information presented for the hypothetical client has been modeled from real companies, there are no "right" answers, since the study deals with judgments (e.g., about materiality). Consequently, your responses will not be evaluated for "correctness." Nevertheless, if the study is to be valid, your responses must be conscientious, and it is essential to complete each of the requirements presented in the second booklet. Your responses will be held in the strictest confidence; neither your identify, nor your firm's, will be divulged in the discussions of the experiment's results. Your assistance is greatly appreciated.

Instructions

SAS No. 47 does not require auditors to quantify their preliminary judgments about materiality (i.e., planning stage
materiality judgments), but implies that it may be reasonable to do so. In addition, paragraph 12 of the SAS suggests, "...the auditor ordinarily considers materiality for planning purposes in terms of the smallest aggregate level of errors that could be considered material to any one of the financial statements." Therefore, you are asked to quantify the planning stage materiality threshold (before taxes) which you deem appropriate for, first, the balance sheet and, second, the income statement. In addition, you are asked to estimate the required audit hours, by level, for the staff, senior, manager, and partner. Further, you are asked to make a general assessment of business risk (i.e., "exposure") on a seven-point scale, where "1" refers to very low and "7" refers to very high.

The materiality judgments and risk assessments are requested for each of the given combinations of (a) total assets, (b) net sales, and (c) income (loss) before taxes. The audit hour estimations are requested only for selected combinations of those financial statement amounts. Please assume that these values have been obtained from the client and are unaudited.

Except that you may temper your judgments by your perception of general economic conditions, please base your judgments only on the information presented in the background information in this booklet and the specific values provided in the second booklet for total assets, net sales, and income (loss) before taxes. For each of the 28 cases, please specify a single value for your preliminary judgment
about materiality for the balance sheet and a corresponding single value for the income statement. If you are accustomed to providing a range of values for these judgments, please specify the range which you believe to be applicable; in addition, please specify the one value for the balance sheet and corresponding one value for the income statement which you deem most appropriate.

In addition, you are asked to complete a questionnaire that deals with auditors' attitudes toward qualified and unqualified opinions and their perceptions of the consequences associated with those opinions. Finally, you are asked to respond to a "debriefing" questionnaire that asks for some additional information which is commonly requested in experimental studies such as this. It is very important that you respond independently of others who may be participating; please do not discuss the experiment with others, in order to maintain the integrity of individuals' responses.

**Background Information**

**Your role**

You are the partner in charge of your firm's audit of the financial statements of Amicus Incorporated as of January 31, 1986, and for the year then ended. As you approach the planning of the engagement, a question has been raised about the appropriate planning level materiality threshold (for both the balance sheet and the income statement) which will be specified in the planning memorandum.
Your firm's audit work must be performed during the "busy season." Due to staffing difficulties, no significant interim work can be done; the fieldwork will be performed entirely at year-end. The client requests that your report be issued by April 30. The audit team consists of two staff level auditors (one has six months of experience and the other has one year), a senior, a manager, an engagement partner, and a pre-issuance review partner who will review the audit work and the firm's report.

**Private company**

Amicus, Inc., is a department store company with its primary operations in Columbus, Ohio. It was founded in the 1920's as a dry goods store and evolved into a department store company during the 1950's. It has enjoyed continued growth and profitability. Initially a sole proprietorship, Amicus was incorporated not long after it was founded. The founding family has maintained control of Amicus, and trading of the shares is restricted. The company continues to be closely-held.

**Scope of operations**

Amicus carries most of the merchandise associated with a traditional department store. Amicus has two stores in Columbus--one near the downtown area and one in the rapidly growing north-end of the city. In addition, there are two branch stores--one each in two
nearby communities. The relative sales volume for the stores is: downtown (48%), north-end (36%), branch #1 (10%), and branch #2 (6%).

**Competition**

Competition is strong in the retail industry and there are few barriers to entry. The major competitors are other department stores, some of which are large national companies. Additional pressure comes from numerous specialty stores, which are increasing in importance in the retail industry. Nevertheless, Amicus remains well-positioned, due in part to strong customer identification. Good planning by management led to the development of strength in several areas associated with the changing demographics and tastes of its customers.

**Audit history**

Your firm has audited the financial statements of Amicus since the fiscal year ended 1/31/79. That first audit was a consequence of a bank's requirement for an annual audit when Amicus borrowed $6 million for ten years in March of 1979. Prior to that first audit, your firm had provided tax services to Amicus for nearly 15 years.

Your firm has given an unqualified opinion each year, except for a qualified opinion the first year due to inability to verify the beginning inventory. Management has been willing to record your firm's proposed adjustments; in your view, the net effect of these adjustments has been minor in the past. Adjustments have usually
related to tax accruals, the allowance for uncollectibles, and insignificant cut-off matters such as unrecorded liabilities, which have not varied much from year to year.

You regard management as "control conscious." Similar procedures, adequate in your view, are in effect at each of the stores to control transactions at the sales registers. The registers are part of a centralized computer network that updates customers' accounts and maintains the perpetual inventory records. The administrative offices and the accounting records for each store are located in the downtown store. Although there is no internal auditing department, numerous control procedures are in place within the computer system, and others are performed by accounting department employees. Your firm has concluded from prior years' reviews of the system of internal controls that there are adequate control procedures and a reasonable degree of segregation of duties in the major transaction cycles--purchasing, cash disbursements, sales, cash receipts, and payroll.

Because you believe that reliance on the system of internal controls is not cost beneficial, you plan to use entirely a substantive audit approach. This was the approach used in prior years. There are reportedly no major changes in internal controls this year. Your review of the internal control system will consist primarily of "walking through" selected transactions to verify that the controls are operating as documented.
As in prior years, the two most significant areas of audit concern are inventory and receivables. Inventory is about 41 percent of total assets and is accounted for by the LIFO retail method. The client plans to close the four stores for two days (February 1 and 2) in order to conduct the annual physical inventory. You plan to verify, on a test basis, the client's physical counts, unit costs, and extensions.

Amicus has its own charge card, and trade receivables are its own charge accounts. About 21 percent of total assets at year-end consist of these receivables. Major planned audit procedures include confirmation of selected accounts and review of cash receipts subsequent to year-end.

Fees

You have compared the audit and tax fees paid by Amicus with those paid by other clients of comparable size and audit effort, and have found that the Amicus fees are about average, as a percentage of standard (desired) fees, relative to those of the other clients. Due to management's moderate resistance to fee increases, it is expected that this condition will continue in future years.

Composition of financial statements

As explained earlier, the cases presented in the second booklet differ as to the magnitudes of total assets, net sales, and income
(loss) before taxes. However, the following balance sheet, expressed in percentages, and the income statement information are applicable to each case.

Assets:
- Cash & equivalents: 6.5%
- Trade receivables (net): 21.1%
- Inventory: 41.2%
- Other current: .9%
- Fixed assets (net): 24.1%
- Intangibles (net): .2%
- Other non-current: 6.0%
- Total assets: 100.0%

Equities:
- Trade payables: 18.6%
- Current portion of long-term debt: 2.3%
- Accrued expenses: 7.8%
- Other current: 2.8%
- Long-term debt: 33.6%
- Net worth: 34.9%
- Total equities: 100.0%

The gross profit rate in each case is 35.7 percent of sales.

Since both sales and income (loss) before taxes vary in the cases in the second booklet, only the cost of goods sold as a percentage of sales is given. In recent years, Amicus' effective federal income tax rate has tended to be about 38 percent.
APPENDIX B
ATTITUDE MEASUREMENT QUESTIONNAIRE

The purpose of this questionnaire is to elicit your attitudes toward qualified audit opinions relative to unqualified audit opinions. In the context of this study, qualifications related to GAAP departures (i.e., regarding "material" misstatements) are considered.

PART (A)

Please identify the likelihood which you believe to be associated with each of the following statements. Please circle the appropriate number associated with the likelihood which is most consistent with your views.

If you would prefer to think of the likelihood statements used in this section in terms of probabilities, you may use the following guidelines:

<table>
<thead>
<tr>
<th>Likelihood Statements</th>
<th>Corresponding Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. practically no chance that this would occur;</td>
<td>0-10%</td>
</tr>
<tr>
<td>2. very unlikely that this would occur;</td>
<td>11-25%</td>
</tr>
<tr>
<td>3. somewhat unlikely that this would occur;</td>
<td>26-45%</td>
</tr>
<tr>
<td>4. there is about a 50-50 chance that this would occur;</td>
<td>46-55%</td>
</tr>
<tr>
<td>5. somewhat likely that this would occur;</td>
<td>56-75%</td>
</tr>
<tr>
<td>6. very likely that this would occur;</td>
<td>76-90%</td>
</tr>
<tr>
<td>7. this almost certainly would occur;</td>
<td>91-100%</td>
</tr>
</tbody>
</table>

Consider a client whose financial statements do not actually contain any "material" misstatements. For me to conclude that there are "material" misstatements, and issue a qualified opinion (perhaps inadvertently by error—e.g., due to judgmental areas such as warranties or sales returns, or by making extensions of statistical samples, etc.) would:
a. result in losing the firm as a client

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

b. result in litigation by the client against my accounting firm

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

c. damage my accounting firm's reputation within the business community

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

d. result in deteriorated relations with the client

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.
e. damage my own standing within my accounting firm

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

f. demonstrate auditing negligence

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

g. damage the reputation of the client within the business community

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

Consider a client whose financial statements contain material misstatements. For me to conclude that there are no "material" misstatements, and issue an unqualified opinion (i.e., not qualified for a GAAP departure; perhaps inadvertently by error—e.g., due to judgmental areas such as warranties or sales returns, or by making extensions of statistical samples, etc.), would:

a. result in litigation by affected parties against my accounting firm

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.
b. demonstrate auditing negligence

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

c. result in losing the firm as a client

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

d. damage my accounting firm's reputation within the business community

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

e. damage my own standing within my accounting firm

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.
f. result in deteriorated relations with the client

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.


g. damage the reputation of the client within the business community

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

Consider a client whose financial statements contain "material" misstatements. For me to conclude that there are "material" misstatements and issue an opinion qualified for a GAAP departure would:

a. alert financial statement users to a reporting problem and, thereby, increase the credibility of the audit report to the public

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

b. enhance my accounting firm's reputation within the business community

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.
c. result in losing the firm as a client

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

d. enhance my own standing within my accounting firm

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

e. damage the reputation of the client within the business community

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

f. result in deteriorated relations with the client

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.
g. avoid litigation against my accounting firm

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

Consider a client whose financial statements do not contain any "material" misstatements. For me to conclude that there are no "material" misstatements and issue an unqualified opinion (i.e., not qualified for a GAAP departure) would:

a. fulfill my accounting firm's responsibility to the public

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

b. result in stronger relations with the client

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

c. enhance my accounting firm's reputation within the business community

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.
d. avoid the undue concern by financial statement users that might occur if a qualified opinion were issued

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

e. enhance my own standing within my accounting firm

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

f. avoid litigation against my accounting firm

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.

g. avoid losing the firm as a client

1. practically no chance that this would occur;
2. very unlikely that this would occur;
3. somewhat unlikely that this would occur;
4. there is about a 50-50 chance that this would occur;
5. somewhat likely that this would occur;
6. very likely that this would occur;
7. this almost certainly would occur.
PART (B)

For each of the following statements, please circle one number which most accurately reflects your feelings about the statement.

a. My accounting firm losing a client is
   1. neither bad nor good (indifferent)
   2. slightly bad
   3. moderately bad
   4. bad
   5. extremely bad

b. Litigation against my accounting firm by a client is
   1. neither bad nor good (indifferent)
   2. slightly bad
   3. moderately bad
   4. bad
   5. extremely bad

c. Damage to my accounting firm's reputation within the business community is
   1. neither bad nor good (indifferent)
   2. slightly bad
   3. moderately bad
   4. bad
   5. extremely bad

d. Causing deteriorated relations with a client is
   1. neither bad nor good (indifferent)
   2. slightly bad
   3. moderately bad
   4. bad
   5. extremely bad

e. Damaging my own standing within my accounting firm is
   1. neither bad nor good (indifferent)
   2. slightly bad
   3. moderately bad
   4. bad
   5. extremely bad
f. A demonstration of auditing negligence is

1. neither bad nor good (indifferent)
2. slightly bad
3. moderately bad
4. bad
5. extremely bad

g. Damaging the reputation of the client within the business community is

1. neither bad nor good (indifferent)
2. slightly bad
3. moderately bad
4. bad
5. extremely bad

h. Litigation against my firm by affected financial statement users is

1. neither bad nor good (indifferent)
2. slightly bad
3. moderately bad
4. bad
5. extremely bad

i. Alerting financial statement users to a reporting problem and, thereby, increasing the credibility of the audit report to the public is

1. neither bad nor good (indifferent)
2. slightly good
3. moderately good
4. good
5. extremely good

j. Fulfilling my accounting firm's responsibility to the public is

1. neither bad nor good (indifferent)
2. slightly good
3. moderately good
4. good
5. extremely good
k. Enhancing my accounting firm's reputation within the business community is
1. neither bad nor good (indifferent)
2. slightly good
3. moderately good
4. good
5. extremely good

l. Enhancing my own standing within my accounting firm is
1. neither bad nor good (indifferent)
2. slightly good
3. moderately good
4. good
5. extremely good

m. Causing stronger relations with the client is
1. neither bad nor good (indifferent)
2. slightly good
3. moderately good
4. good
5. extremely good

n. Avoiding the undue concern by financial statement users that might occur is a qualified opinion were issued is
1. neither bad nor good (indifferent)
2. slightly good
3. moderately good
4. good
5. extremely good

o. Avoiding litigation against my accounting firm is
1. neither bad nor good (indifferent)
2. slightly good
3. moderately good
4. good
5. extremely good

p. Avoiding the loss of a client is
1. neither bad nor good (indifferent)
2. slightly good
3. moderately good
4. good
5. extremely good
PLEASE PROCEED TO THE SURVEY QUESTIONNAIRE
APPENDIX C
SURVEY QUESTIONNAIRE

Please provide the following information about your own background and opinions.

1. To the nearest half year, how long have you been employed as an auditor? _____ years

2(a). Have you worked for your present firm for your entire auditing career? _____ (yes, no)

2(b). If you answered "no" to #2(a), what firm(s) did you previously work for, and how long (to the nearest half year) did you work for the firm(s) in an auditing capacity? (Please list the most recent prior firm first.)

<table>
<thead>
<tr>
<th>Firm name</th>
<th>Office (city)</th>
<th>How long</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most recent:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Next most recent:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Next most recent:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Across your entire auditing career, what was the length of time (to the nearest half year) you spent as a staff level auditor, senior, manager, and partner? (Since the hierarchy is somewhat different across firms, please include the time spent as an "assistant" along with "staff" and include time spent as a "supervisor" along with "manager.")

<table>
<thead>
<tr>
<th></th>
<th>_____ years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staff</td>
<td></td>
</tr>
<tr>
<td>Senior</td>
<td></td>
</tr>
<tr>
<td>Manager</td>
<td></td>
</tr>
<tr>
<td>Partner</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>_____ years (should agree to #1 above)</td>
</tr>
<tr>
<td></td>
<td>=========</td>
</tr>
</tbody>
</table>

4. Are you a CPA? _____ (yes, no)

5. Have you received a college/university degree? _____ (yes, no) [If you answered "no" to #5, please skip to #8(a).]
6(a). What was your undergraduate major? (Circle one number)

1. Accounting
2. Business (other than accounting)
3. Other (please specify) ______________________________

6(b). What institution granted your undergraduate degree?

________________________________________________________________________

7(a). Have you received a graduate degree? _____ (yes, no)
[If you answered "no" to #7(a), please skip to #8(a).]

7(b). If you answered "yes" to #7(a), which graduate degree(s) have you received? (Circle all the numbers that apply)

1. MBA with an accounting concentration
2. MBA with a concentration in a non-accounting area
3. M.S. in accounting
4. Other(s) (please specify) ______________________________

7(c). If you answered "yes" to #7(a), what institution(s) granted the degree(s)?

________________________________________________________________________

8(a). Do you consider yourself an industry specialist? _____ (yes, no)

8(b). If you answered "yes" to #8(a), what industry (industries)?

________________________________________________________________________

9. Have you ever audited a department store client? _____ (yes, no)

10. Have you ever audited a retail client other than a department store? _____ (yes, no)

11(a). Have you read SAS No. 47 ("Audit Risk and Materiality in Conducting an Audit")? _____ (yes, no)

11(b). If you answered "yes" to #11(a), has SAS No. 47 affected your approach to audit planning? _____ (yes, no)

11(c). If you answered "yes" to #11(b), please explain briefly.

________________________________________________________________________
12(a). Do you believe that, under the same circumstances, different auditors should establish the same materiality threshold at the planning stage? (Circle one number)

1. Strongly yes;
2. Moderately yes;
3. Neutral (neither yes nor no);
4. Moderately no;
5. Strongly no;

12(b). Do you believe that, under the same circumstances, different auditors should establish the same materiality threshold at the evaluation stage? (Circle one number)

1. Strongly yes;
2. Moderately yes;
3. Neutral (neither yes nor no);
4. Moderately no;
5. Strongly no;

13(a). Do you believe that more detailed guidelines are needed to assist auditors in quantifying their materiality judgments (in order to achieve a greater degree of consensus across the profession) at the planning stage of the audit (termed the "preliminary judgment about materiality" per SAS No. 47)? (Circle one number)

1. Strongly yes;
2. Moderately yes;
3. Neutral (neither yes nor no);
4. Moderately no;
5. Strongly no;

13(b). Do you believe that more detailed guidelines are needed to assist auditors in quantifying their materiality judgments (in order to achieve a greater degree of consensus across the profession) at the evaluation stage of the audit (i.e., the "final" materiality judgment)? (Circle one number)

1. Strongly yes;
2. Moderately yes;
3. Neutral (neither yes nor no);
4. Moderately no;
5. Strongly no;
14. In conjunction with audit planning, how often do you quantify (i.e., specify a dollar magnitude for) the preliminary judgment about materiality? (Circle one number)

1. Almost always;
2. Usually;
3. Occasionally;
4. Seldom;
5. Almost never;

15(a). Compared with other auditors within your firm, how do you view your willingness to accept risks? (Circle one number)

1. Much more willing than most;
2. Somewhat more willing than most;
3. As willing as most;
4. Somewhat less willing than most;
5. Much less willing than most;
6. No opinion--cannot say;

15(b). Compared with other auditors from other firms, how do you view your willingness to accept risks? (Circle one number)

1. Much more willing than most;
2. Somewhat more willing than most;
3. As willing as most;
4. Somewhat less willing than most;
5. Much less willing than most;
6. No opinion--cannot say;

16(a). Compared with other auditors within your firm, how do you view your own materiality thresholds? (Circle one number)

1. Much higher than most;
2. Somewhat higher than most;
3. About the same as most;
4. Somewhat lower than most;
5. Much lower than most;
6. No opinion--cannot say;

16(b). Compared with other auditors from other firms, how do you view your own materiality thresholds? (Circle one number)

1. Much higher than most;
2. Somewhat higher than most;
3. About the same as most;
4. Somewhat lower than most;
5. Much lower than most;
6. No opinion--cannot say;
17(a). Compared with other auditors within your firm, in similar engagements, do you believe you tend to perform more or less extensive audit tests? (Circle one number)

1. Much more extensive than most;
2. Somewhat more extensive than most;
3. About the same as most;
4. Somewhat less extensive than most;
5. Much less extensive than most;
6. No opinion--cannot say;

17(b). Compared with other auditors from other firms, in similar engagements, do you believe you tend to perform more or less extensive audit tests? (Circle one number)

1. Much more extensive than most;
2. Somewhat more extensive than most;
3. About the same as most;
4. Somewhat less extensive than most;
5. Much less extensive than most;
6. No opinion--cannot say;

18(a). Compared with other auditors within your firm, do you believe you tend to spend relatively more or less time related to audit planning activities? (Circle one number)

1. Much more time than most;
2. Somewhat more time than most;
3. About the same as most;
4. Somewhat less time than most;
5. Much less time than most;
6. No opinion--cannot say;

18(b). Compared with other auditors from other firms, do you believe you tend to spend relatively more or less time related to audit planning activities? (Circle one number)

1. Much more time than most;
2. Somewhat more time than most;
3. About the same as most;
4. Somewhat less time than most;
5. Much less time than most;
6. No opinion--cannot say;
19. Based upon your audit program planning experience, do you believe you tend to overestimate or underestimate the time required to actually complete the planned audit work? (Circle one number)

1. Considerably overestimate;
2. Somewhat overestimate;
3. Accurately estimate;
4. Somewhat underestimate;
5. Considerably underestimate;
6. No opinion--cannot say;

20. How strongly did you identify with your assigned role as an auditor of the hypothetical client? (Circle one number)

1. Did not identify at all;
2. Identified a little;
3. Identified moderately;
4. Identified strongly;

21. How similar were the tasks you were asked to perform in the experiment relative to tasks you must perform in the planning of an actual engagement? (Circle one number)

1. Very dissimilar;
2. Somewhat dissimilar;
3. Neutral (neither similar nor dissimilar);
4. Somewhat similar;
5. Very similar;

22. How realistic did you find the experimental information? (Circle one number)

1. Very unrealistic;
2. Fairly unrealistic;
3. Neutral (neither realistic nor unrealistic);
4. Fairly realistic;
5. Very realistic;

23(a). Would you have found additional information helpful in quantifying the preliminary judgment about materiality for the 28 cases? _____ (yes, no)

23(b). If you answered "yes" for #23(a), please specify. 

____________________________________________________________________
____________________________________________________________________
____________________________________________________________________
24(a). Would you have found additional information helpful in estimating the audit hours for the selected cases? _____ (yes, no)

24(b). If you answered "yes" to #24(a), please specify.

<table>
<thead>
<tr>
<th>Factor Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total assets</td>
<td></td>
</tr>
<tr>
<td>Net sales</td>
<td></td>
</tr>
<tr>
<td>Income (loss) before taxes</td>
<td></td>
</tr>
<tr>
<td>Other (e.g., some combination of factors)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

25(a). Please allocate 100 points to the three factors manipulated in the 28 cases according to how important each was in arriving at your materiality judgments regarding the income statement (i.e., important factors should be assigned more points than the less important factors).

<table>
<thead>
<tr>
<th>Factor Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total assets</td>
<td></td>
</tr>
<tr>
<td>Net sales</td>
<td></td>
</tr>
<tr>
<td>Income (loss) before taxes</td>
<td></td>
</tr>
<tr>
<td>Other (e.g., some combination of factors)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

25(b). Please allocate 100 points to the three factors manipulated in the 28 cases according to how important each was in arriving at your materiality judgments regarding the balance sheet (i.e., important factors should be assigned more points than the less important factors).

<table>
<thead>
<tr>
<th>Factor Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total assets</td>
<td></td>
</tr>
<tr>
<td>Net sales</td>
<td></td>
</tr>
<tr>
<td>Income (loss) before taxes</td>
<td></td>
</tr>
<tr>
<td>Other (e.g., some combination of factors)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>

25(c). Please allocate 100 points to the three factors manipulated in the 28 cases according to how important each was in arriving at the estimated audit hours (i.e., important factors should be assigned more points than the less important factors).

<table>
<thead>
<tr>
<th>Factor Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total assets</td>
<td></td>
</tr>
<tr>
<td>Net sales</td>
<td></td>
</tr>
<tr>
<td>Income (loss) before taxes</td>
<td></td>
</tr>
<tr>
<td>Other (e.g., some combination of factors)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
</tr>
</tbody>
</table>
26. How interesting did you find this experiment? (Circle one number)
   1. Very dull;
   2. Fairly dull;
   3. Neither dull nor interesting;
   4. Fairly interesting;
   5. Very interesting;

27. Would you like a written summary of the results? _____ (yes, no)

28. If you have any other comments about the nature of the experiment
    and/or the experimental materials, please record them here.

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

29. How long did it take you to complete the experimental cases (both
    the materiality judgments and audit hour estimates) and
    questionnaires?

    _____ hour(s) and _____ minutes

YOUR GENEROUS COOPERATION IS GREATLY APPRECIATED. THANK YOU.
APPENDIX D
RESULTS OF ANALYSES (BALANCE SHEET JUDGMENTS)

Introduction

As pointed out previously, SAS No. 47 suggests that auditors should consider their preliminary judgments about materiality at the aggregate financial statement level. The SAS indicates (paragraph 12) that these judgments may be made appropriately from either an income statement or balance sheet perspective. In general, prior research has scarcely examined auditors' planning stage materiality judgments. In particular, there has been no distinction made in the academic or professional literatures between materiality judgments from an income statement versus balance sheet point of view.

In this study, the income statement materiality judgments are deemed to be the dependent variable of primary interest. Virtually all of the prior studies of auditors' materiality judgments (nearly all of which focused on the evaluation stage) have shown income considerations to dominate the judgments. Consequently, it was expected that income (loss) before taxes would be the primary cue influencing auditors' judgments in this experiment. Along the same line, income statement considerations are generally believed to drive auditors' planning stage materiality judgments, in practice (e.g., Zuber, et al. [1983, p. 43]).
The balance sheet materiality judgments were elicited for exploratory purposes. Analyses, virtually identical to those reported in Chapter IV for the income statement materiality judgments, were performed on the balance sheet judgments; and the results are reported here. However, since the purpose of the analyses of the balance sheet judgments was exploratory, formal hypotheses with respect to these judgments were not considered necessary or appropriate. Therefore, although the analyses follow those reported in Chapter IV with respect to the income statement judgments, the discussion below is not tied to formal hypotheses.

The organization of this appendix is similar to that of Chapter IV with respect to the income statement judgments. This begins with consideration of effects associated with level of responsibility, and is followed by discussions of firm and audit technology effects. For each of these effects, the results with respect to location are presented first, followed by the results with respect to variation. (Analyses of risk attitudes with respect to the balance sheet materiality judgments are omitted from this appendix. The income statement, rather than the balance sheet, materiality judgments were of primary interest and were expected to be influenced by the notion of differing risk attitudes. Thus, risk attitudes were examined only with respect to the income statement judgments and were reported in Chapter IV.)
These discussions are followed by an examination of the participants' individual ANOVA models for the balance sheet judgments. After that, test-retest consistency and order effects are considered.

**Level of Responsibility Effects**

As a starting point, subjects' responses were standardized on a case-by-case basis, across all 72 subjects. For each individual case, the responses of all the subjects on that case were transformed to a distribution with mean zero and variance one. (Thus, an individual with an above average response on that case would have a positive standardized score for that case; an individual with a below average response on that case would have a negative standardized score for that case.) These standardized scores were summed for each participant across all 24 cases to obtain an overall measure of the extremeness of each subject's balance sheet judgments. Figure 10 plots these standardized score, by level of responsibility, for the seniors, managers, and partners (levels 2-4, respectively).

**Location**

Figure 10 suggests that there are no systematic differences among the seniors, managers, and partners in their balance sheet materiality judgments with respect to location. Further, no outliers can be identified. As described in Chapters III and IV, this study
Figure 10

Standardized Scores for the Balance Sheet Materiality Judgments Plotted by Level of Responsibility
investigates two distinct potential sources of systematic differences among the participants (i.e., experience-related and firm-related effects). To simultaneously consider both possible effects in the analysis of effects associated with level of responsibility, the analyses for differences among the seniors, managers, and partners were performed on a firm-by-firm basis.

Therefore, six separate one-way ANOVAs were conducted to identify whether there were systematic differences in location across the three levels of responsibility. As summarized in Table 20, level of responsibility was not significant for any of the firms at any reasonable level. Fisher's method for combining tests was used to derive an overall test statistic regarding differences associated with level of responsibility (Hedges and Olkin [1985, p. 37]). The resulting test statistic was approximately 8.87 and was not significant at any reasonable level (the critical value for the .05 level was 21.03). Thus, there were no real differences in location among the three levels of responsibility.

One further type of comparison was made to gain possible insights regarding differences associated with level of responsibility. Effects were examined within each of the two audit technologies. (Firms #1-3 were the relatively structured firms and firms #4-6 were the relatively unstructured firms.) Two separate one-way ANOVAs were run to identify whether there were differences among the three levels of responsibility for either audit technology group. Again,
Table 20

Standardized Balance Sheet Materiality Scores
By Level of Responsibility and By Firm Affiliation

<table>
<thead>
<tr>
<th>Firm #</th>
<th>Seniors</th>
<th>Managers</th>
<th>Partners</th>
<th>Average</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-5.73</td>
<td>-5.46</td>
<td>-1.83</td>
<td>-4.34</td>
<td>.80</td>
</tr>
<tr>
<td>2</td>
<td>-3.09</td>
<td>-1.80</td>
<td>5.09</td>
<td>.07</td>
<td>.77</td>
</tr>
<tr>
<td>3</td>
<td>-6.76</td>
<td>-1.98</td>
<td>-9.72</td>
<td>-6.16</td>
<td>.86</td>
</tr>
<tr>
<td>4</td>
<td>-23.10</td>
<td>-8.15</td>
<td>-11.91</td>
<td>-14.39</td>
<td>.14</td>
</tr>
<tr>
<td>5</td>
<td>26.94</td>
<td>30.71</td>
<td>12.27</td>
<td>23.31</td>
<td>.42</td>
</tr>
<tr>
<td>6</td>
<td>-12.02</td>
<td>.94</td>
<td>15.61</td>
<td>1.51</td>
<td>.38</td>
</tr>
<tr>
<td>Average</td>
<td>-3.96</td>
<td>2.38</td>
<td>-3.57</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p-value based on one-way ANOVA for level of responsibility calculated on a firm-by-firm basis.
level of responsibility was not significant at the .10 level for either of the two audit technologies.

Variation

Figure 10 does not portray any conspicuous differences in variability among the seniors, managers, and partners. The following sums of squares were associated with each level of responsibility:

<table>
<thead>
<tr>
<th>Level</th>
<th>Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seniors</td>
<td>9,832</td>
</tr>
<tr>
<td>Managers</td>
<td>9,521</td>
</tr>
<tr>
<td>Partners</td>
<td>10,630</td>
</tr>
<tr>
<td>Total</td>
<td>29,983</td>
</tr>
</tbody>
</table>

Effects associated with level of responsibility with respect to variation were examined within each of the six firms. Six separate applications of Cochran's test for homogeneity of variance were conducted, using the sums of squares for each level of responsibility for each firm, to identify whether there were any significant differences among the levels of responsibility for any of the firms. Using the method described by Cochran [1941], the specific p-values were inferred with respect to level of responsibility for each firm. These p-values were then combined into an overall test statistic using Fisher's method for combining tests.

No differences associated with level of responsibility were found for any of the firms at the .05 level. When the six p-values were combined using Fisher's method, an overall test statistic of approximately 11.09 resulted. This was not significant at any
reasonable level. (The critical value associated with the .05 level was 21.03.) A supplemental type of comparison with respect to variation was made. Effects associated with level of responsibility were examined within each of the two audit technologies. The sums of squares associated with the seniors, managers, and partners were calculated within each audit technology group. Cochran's test for homogeneity of variance was applied separately to each of the two audit technologies. No difference was found among levels of responsibility for either of the two audit technologies at the .05 level.

Summary of Level of Responsibility Effects

The results of the analyses of the balance sheet materiality judgments for differences associated with level of responsibility were consistent and clear. No systematic differences were detected among levels of responsibility, either as to location or variation (overall, for any of the firms, or for either of the two audit technologies). These results are summarized in Table 21.

Firm Effects

In view of the preceding section which reports no differences among levels of responsibility, one may appropriately regard the seniors, managers, and partners as homogeneous, and proceed with the
Table 21
Summary of Level of Responsibility Effects
For the Balance Sheet Judgments

<table>
<thead>
<tr>
<th>Comparisons Made</th>
<th>Re: Location</th>
<th>Re: Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Level of responsibility, by firm affiliation</td>
<td>No significant differences were found at the .10 level.</td>
<td>No significant differences at the .05 level.</td>
</tr>
<tr>
<td>B. Overall effect associated with level of responsibility</td>
<td>No differences among the seniors, managers, and partners at the .10 level.</td>
<td>No significant differences at the .05 level.</td>
</tr>
<tr>
<td>C. Level of responsibility, by audit technology</td>
<td>No significant differences for either audit technology at the .10 level.</td>
<td>No significant differences at the .05 level.</td>
</tr>
</tbody>
</table>

Note: This table summarizes the analyses identified in Table 4. Table 4 summarizes the primary method of analysis for differences as to location and variation, identifies the data on which the above conclusions are based, and ties the analyses conducted to the notation introduced in Figure 7.
analysis of firm effects without regard to differences associated with level of responsibility. Differences at the firm level are explored, first, with respect to location and, second, with respect to variation.

Location

Figure 11 plots the standardized scores, by firm affiliation, and suggests the existence of differences among the firms with respect to location. For example, firm #5 has a relatively high mean and firm #4 has a relatively low mean.

A one-way ANOVA compared the standardized scores for systematic differences in the means associated with firm affiliation. The means and standard deviations (sd) of the standardized balance sheet materiality scores, by firm, were as follows:

<table>
<thead>
<tr>
<th>Firm</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-4.34</td>
<td>.07</td>
<td>-6.16</td>
<td>-14.39</td>
<td>23.31</td>
</tr>
<tr>
<td>sd</td>
<td>8.15</td>
<td>15.18</td>
<td>17.56</td>
<td>10.68</td>
<td>19.02</td>
</tr>
</tbody>
</table>

As documented in Table 22, firm affiliation was highly significant (p ≤ .0001). Tukey's multiple comparison test identified firm #5 to be significantly different than each of the other firms; no other differences were significant at the .05 level. The nonparametric Wilcoxon rank sum test corroborated the parametric ANOVA results. The large sample approximation of the Wilcoxon rank sum test's distribution (Hollander and Wolfe [1973, p.68]) indicated that
Figure 11

Standardized Scores for the Balance Sheet Materiality Judgments Plotted by Firm Affiliation
Table 22
ANOVA Table For Firm Effects
For the Balance Sheet Judgments

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
<th>Mean Square</th>
<th>F-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5</td>
<td>9,711.0</td>
<td>1,942.2</td>
<td>6.15</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>66</td>
<td>20,845.2</td>
<td>315.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>30,556.2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Equal to
firm #5 had a higher balance sheet materiality threshold than four of the other firms (p ≤ .003 where .003 = .05/15, since there were 15 between-firm comparisons). There was a real difference in central tendency among the firms with respect to their balance sheet materiality judgments.

Two types of supplemental comparisons were made to gain insights about the overall firm differences identified above. First, firm differences were examined, separately, across each of the three levels of responsibility. Second, firm differences were examined separately within each of the two audit technologies.

Three separate one-way ANOVAs were conducted to identify whether there were any significant differences as to location across the firms at each level of responsibility. Firm effects were highly significant for the seniors (p = .006). Tukey's multiple comparison test identified the seniors from firm #5 to be different from the seniors from firms #4 and 6. Firm affiliation was not significant for the managers (p = .06) or partners (p = .37), although the managers could be viewed as borderline. Despite the non-significance among the managers and partners, an apparently modest firm effect was associated with them, too. The managers from firm #5 had the highest standardized scores, followed by the managers from firm #6; the managers from firm #4 had the lowest scores. The partners from firm #6 had the highest scores, followed by the partners from firm #5; the partners from firm #4 had the lowest scores. As a practical matter, at each level of
responsibility, the participants from firm #5 tended to have relatively high balance sheet materiality thresholds, while the participants from firm #4 had relatively low balance sheet materiality thresholds.

Two separate one-way ANOVAs looked for firm differences as to location within each audit technology group. Within the structured audit technology, firm differences were not significant (p = .58). However, within the unstructured audit technology, firm differences were highly significant (p ≤ .0003). Tukey's multiple comparison test identified firm #5 as significant relative to firms #4 and 6.

Variation

Figure 11 plots the standardized scores, by firm, and suggests differences among the firms with respect to variation. Apparently, firm #6 had relatively high variability, while firm #1 had relatively low variability. The following sums of squares were associated with the firm affiliations:

<table>
<thead>
<tr>
<th>Firm</th>
<th>Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>798</td>
</tr>
<tr>
<td>#2</td>
<td>2,764</td>
</tr>
<tr>
<td>#3</td>
<td>3,699</td>
</tr>
<tr>
<td>#4</td>
<td>1,370</td>
</tr>
<tr>
<td>#5</td>
<td>4,342</td>
</tr>
<tr>
<td>#6</td>
<td>7,873</td>
</tr>
<tr>
<td>Total</td>
<td>20,846</td>
</tr>
</tbody>
</table>

Cochran's test for homogeneity of variance confirmed the existence of firm differences (p ≤ .05). After deleting the sums of
squares associated with firm #6, the most variable firm, Cochran's test was applied to the remaining sums of squares. No difference could be demonstrated among the remaining firms at the .05 level.

To clarify which particular firms differed (and because Cochran's test for homogeneity of variance may be affected by non-normal data), an additional test of variation was conducted. Ratios of the sums of squares (i.e., F-tests) were calculated for each pair of firms for all pairwise comparisons (there were fifteen such comparisons). Significant differences were identified using the conservative Bonferonni simultaneous confidence intervals at the .05 level. Only one such comparison was significant; firm #6 was significantly higher in variation than firm #1. That result is consistent with Cochran's test and clarifies which firms differ.

Similar to the tests regarding location reported above, two types of supplemental comparisons were made to gain insights about the overall firm differences as to variation. First, the variances were examined across the firms, separately, by level of responsibility. These variances are summarized in Table 23. Second, firm differences were examined, separately within each of the audit technology groups.

Cochran's test for homogeneity of variance was applied across firms, separately for each level of responsibility, based on the sums of squares associated with each level within each firm. The test was applied sequentially, deleting the most variable firm when significance was observed and applying the test to the remaining
Table 23

Average Variance By Firm and Level of Responsibility
Based on Standardized (Balance Sheet) Materiality Scores

<table>
<thead>
<tr>
<th>Firm #1</th>
<th>Seniors</th>
<th>Managers</th>
<th>Partners</th>
<th>Total</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>67.86</td>
<td>78.74</td>
<td>106.78</td>
<td>253.38</td>
<td>84.46</td>
</tr>
<tr>
<td>2</td>
<td>307.59</td>
<td>123.17</td>
<td>438.93</td>
<td>869.69</td>
<td>289.90</td>
</tr>
<tr>
<td>3</td>
<td>398.50</td>
<td>464.66</td>
<td>329.13</td>
<td>1192.29</td>
<td>397.43</td>
</tr>
<tr>
<td>4</td>
<td>6.97</td>
<td>177.62</td>
<td>110.67</td>
<td>295.26</td>
<td>98.42</td>
</tr>
<tr>
<td>5</td>
<td>111.11</td>
<td>798.67</td>
<td>284.31</td>
<td>1194.09</td>
<td>398.03</td>
</tr>
<tr>
<td>6</td>
<td>521.85</td>
<td>179.33</td>
<td>1413.67</td>
<td>2114.85</td>
<td>704.95</td>
</tr>
<tr>
<td>Average</td>
<td>235.65</td>
<td>303.70</td>
<td>477.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Image]
firms. For the seniors, firm #6 was the most variable, followed by firms #3 and 2. For the managers, firm #5 was the most variable, followed by firm #3. Firm #6 was the most variable for the partners (p ≤ .05). These were the only significant firm differences, by level, at the .05 level. Firm differences were fairly pervasive across levels of responsibility.

To check for firm differences as to variation within each audit technology group, two separate applications of Cochran’s test for homogeneity of variance were made. The sums of squares for the firms were used (the same as those reported above with respect to overall firm effects), but grouped by audit technology. No differences were found among the firms within either audit technology group at the .05 level.

**Summary of Firm Effects**

Differences were identified at the firm level with respect to both location and variation. These results are summarized in Table 24.

Systematic differences were identified among the firms with respect to location. In particular, the participants from firm #5 had higher balance sheet materiality thresholds than the participants from each of the other firms. The extremeness of the judgments involving firm #5 was further shown by an analysis of firm affiliation, level-by-level. On average, the seniors, managers, and partners from firm
Table 24
Summary of Firm Effects
For the Balance Sheet Judgments

<table>
<thead>
<tr>
<th>Comparisons Made</th>
<th>Re: Location</th>
<th>Re: Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Overall firm effect</td>
<td>Firm affiliation was highly significant (p ≤ .0001). Firm #5 was high relative to all other firms.</td>
<td>Firm differences were found at the .05 level. (Firm #6 was high relative to firm #1.)</td>
</tr>
<tr>
<td>B. Firm affiliation, by level of responsibility</td>
<td>Firm affiliation was highly significant for the seniors (p ≤ .006), borderline for the managers (p=.06), and not significant for the partners (p=.37). Firm #5 was high for the seniors and managers.</td>
<td>Firm differences were found at the .05 level among the seniors, the managers, and the partners.</td>
</tr>
<tr>
<td>C. Firm affiliation, by audit technology</td>
<td>Firm differences were highly significant for the unstructured firms (p ≤ .0003); firm #5 was high relative to firms #4 and 6. No differences were found among the structured firms (p=.58).</td>
<td>No difference was found at the .05 level for either audit technology.</td>
</tr>
</tbody>
</table>

Note: This table summarizes the analyses identified in Table 5. Table 5 summarizes the primary method of analysis for differences as to location and variation, identifies the data on which the above conclusions are based, and ties the analyses conducted to the notation introduced in Figure 7.
all appeared to have high balance sheet judgments relative to their counterparts from other firms (although only the seniors could be identified as significant at the .05 level). The greatest firm differences with respect to location occurred between two unstructured firms--firm #5 had the highest standardized balance sheet materiality scores and firm #4 had the lowest. No firm differences as to location were detected among the structured firms at the .05 level.

Firm differences were also identified with respect to variability. Firm #6 was shown to be significantly more variable than firm #1; no other differences were significant at the .05 level. Firm differences were identified at each level of responsibility. Among the seniors and among the partners, firm #6 was the most variable; among the managers, firm #5 was the most variable. Within each audit technology group, however, no firm differences could be identified. None of the structured firms was distinctly more variable than the others; similarly, none of the unstructured firms was distinctly more variable than the others.

Audit Technology Effects

Based on examination of the standardized materiality scores, plotted by firm affiliation in Figure 11 and the analyses of firm effects reported above, audit technology is apparently not related to either the central tendency or variability of the firms' balance sheet
materiality judgments. However, for completeness and for comparability with the analyses of the income statement judgments reported in Chapter IV, tests of the balance sheet judgments were performed to identify whether there were any statistically significant differences, first, as to location and, second, as to variation between the two audit technologies.

**Location**

As mentioned, Figure 11 does not suggest that there were systematic differences as to location between the structured and unstructured firms. A one-way ANOVA compared the standardized scores for systematic differences in the means associated with audit technology. As documented in Table 25, audit technology was not significant (p=.16). The means of the standardized balance sheet materiality scores were approximately -3.5 (3.5) for the structured (unstructured) firms. That audit technology was not significant was not surprising, since firm #5, an unstructured firm, had the highest standardized scores and firm #4, also an unstructured firm, had the lowest standardized scores.

A supplemental analysis examined each level separately to identify any systematic differences between the structured and unstructured firms for the seniors, managers, and partners. Three one-way ANOVAs tested for location differences between the structured and unstructured firms among the seniors, among the managers, and
Table 25
ANOVA Table For Audit Technology Effects
For the Balance Sheet Judgments

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
<th>Mean Square</th>
<th>F-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>870.4</td>
<td>870.4</td>
<td>2.05</td>
<td>.16</td>
</tr>
<tr>
<td>Residual</td>
<td>70</td>
<td>29,685.8</td>
<td>424.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>30,556.2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
among the partners. Audit technology was not significant for any of the three levels (p= .78, p= .20, and p= .41 for the seniors, managers, and partners, respectively). These results are consistent with the overall result of no difference as to location between the two audit technology groups.

Variation

Cochran's test for homogeneity of variance was applied to the sums of squares associated with the standardized balance sheet materiality scores for the structured and unstructured firms. The following sums of squares were obtained:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured firms</td>
<td>7,507</td>
</tr>
<tr>
<td>Unstructured firms</td>
<td>22,179</td>
</tr>
<tr>
<td>Total</td>
<td>29,686</td>
</tr>
</tbody>
</table>

In the aggregate, the unstructured firms were significantly more variable than the structured firms; almost 75% of the total sums of squares was attributable to the unstructured firms (p ≤ .01). However, examination of Figure 11 suggests that the overall significance of the unstructured firms is driven by the large variability for firms #5 and 6. Firm #4 has relatively low variability. Indeed, to assert that the structured firms unambiguously had greater variability in their balance sheet materiality judgments may be misleading, since firm #4 had relatively low variability.
Moreover, as pointed out above, only one of the fifteen F-tests (based on all pairwise comparisons among the firms' sums of squares) was significant. Nine of the fifteen comparisons involved firms having different audit technologies, and only the comparison between firms #1 and 6 was significant. Consequently, the evidence of differences in variation between the two audit technologies is ambiguous.

A supplemental analysis examined each level of responsibility separately for differences as to variation between the structured and unstructured firms. The sums of squares were calculated for each level of responsibility within each of the two audit technology groups. Three separate applications of Cochran's test for homogeneity of variance were made, one for each level. Although the sums of squares associated with each level were larger for the subjects from unstructured firms, none of the levels of responsibility was significantly different between the two audit technology groups at the .05 level.

**Summary of Audit Technology Effects**

No difference was identified as to location between the structured and unstructured firms. However, with respect to variation, a significant difference was observed; as pointed out, this result should be viewed somewhat cautiously. These results are summarized in Table 26.
Table 26
Summary of Audit Technology Effects
For the Balance Sheet Judgments

<table>
<thead>
<tr>
<th>Comparisons Made</th>
<th>Re: Location</th>
<th>Re: Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Overall audit technology effect</td>
<td>No difference was found between the structured and unstructured firms (p≠.16).</td>
<td>The unstructured firms were more variable than the structured firms, in the aggregate (p ≤.01). However, one unstructured firm was comparable to the structured firms.</td>
</tr>
</tbody>
</table>

| B. Audit technology, by level of responsibility       | No differences were found, for any of the levels, between the two audit technologies (p≠.78, p≠.20, and p≠.41 for the seniors, managers, and partners, respectively). | No differences were found, for any of the levels, between the structured and unstructured firms at the .05 level. |

Note: This table summarizes the analyses identified in Table 6. Table 6 summarizes the primary method of analysis for differences as to location and variation, identifies the data on which the above conclusions are based, and ties the analyses conducted to the notation introduced in Figure 7.
With respect to location, no overall differences were identified between the structured and unstructured firms. Similarly, no differences were found between the two audit technology groups, on a level-by-level basis, for the seniors, managers, and partners. The notion of audit technology was not helpful in understanding the differences observed at the firm level.

With respect to variation, an overall difference was found between the structured and unstructured firms; the unstructured firms, as a group, had greater variability based on the sums of squares associated with the three firms in each technology group. However, only two of the three unstructured firms were visibly more variable than the structured firms. Furthermore, only one of the nine F-tests (involving all possible pairwise comparisons of the firms' sums of squares) for firms having different audit technologies was significant. Therefore, the evidence of audit technology differences as to variation must be viewed as ambiguous. No differences were found between the structured and unstructured firms, on a level-by-level basis, for the seniors, managers, and partners.

Additional Analyses

Similar to the discussion of the income statement materiality judgments in Chapter IV, this section reports the results of several additional types of analyses with respect to the balance sheet
materiality judgments. Three types of additional analyses are reported. First, comments are made about the subjects' individual ANOVA models of their balance sheet judgments. Second, the test-retest consistency among the subjects is discussed. Third, and finally, the tests for order effects (between the two orders of case presentation) are reported.

**Individual ANOVA Models**

As pointed out in Chapter IV, regarding the income statement materiality judgments, individual ANOVA models were constructed for each participant based on the subject's untransformed (or "raw") responses to the 24 unique cases. The purpose of these models was to gain possible insights about differences among the participants, particularly with respect to cue utilization.

Across all 72 subjects, total assets was clearly the dominant variable. This was not surprising, obviously, since the total asset cue was the only independent variable directly applicable to identifying the "size" of the balance sheet, related to making the balance sheet materiality judgment. Total assets was a significant variable for 65 of the subjects; income and sales were significant for nine and five of the subjects, respectively. The latter two variables were significant primarily for individuals who made little or no distinction between the income statement and balance sheet materiality judgments (i.e., they reported the same materiality threshold for the
separate financial statements on many or all of the cases). Two-way interactions were seldom significant; the most frequently significant interaction was income x assets, which was significant for only six subjects. Table 27 summarizes the significant cues (p ≤ .05), by level of responsibility and, separately, by firm affiliation.

Table 28 summarizes the significant cues for the subjects for whom only one variable was significant. Only one cue was significant for 61 of the 72 subjects. Of these, total assets was the sole significant cue for 56 of the subjects. The individual ANOVA models of subjects' balance sheet materiality judgments do not suggest systematic differences among the subjects, either with respect to level of responsibility or firm affiliation. Total assets was the dominant cue for the preponderance of subjects. Compared to the discussion of the income statement judgments in Chapter IV, individual differences were less conspicuous with respect to the balance sheet judgments.

Test-retest Consistency

In order to provide an indication of the meaningfulness of subjects' responses, four cases were repeated for each participant after the 24 original cases were presented. The same cases were repeated for each subject. As described in Chapter IV, the Pearson product-moment correlation coefficient was calculated across all 72 subjects. Secondly, paired t-tests were calculated, by case, to
Table 27

All Significant Cues (p < .05)
For the Balance Sheet Judgments
By Level of Responsibility, By Firm Affiliation

<table>
<thead>
<tr>
<th>By Level of Responsibility:</th>
<th>Seniors</th>
<th>Managers</th>
<th>Partners</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income (loss) before taxes</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Net sales</td>
<td>4</td>
<td>-</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Total assets</td>
<td>21</td>
<td>23</td>
<td>21</td>
<td>65</td>
</tr>
<tr>
<td>Income x sales</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Income x assets</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Sales x assets</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>By Firm Affiliation:</th>
<th>Structured</th>
<th>Unstructured</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm #1</td>
<td>#2</td>
<td>#3</td>
<td>#4</td>
</tr>
<tr>
<td>Income (loss) before taxes</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Net sales</td>
<td>-</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total assets</td>
<td>12</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Income x sales</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Income x assets</td>
<td>1</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Sales x assets</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 28
Only One Significant Cue (p ≤ .05)
For the Balance Sheet Judgments
By Level of Responsibility, By Firm Affiliation

<table>
<thead>
<tr>
<th></th>
<th>Seniors</th>
<th>Managers</th>
<th>Partners</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income (loss) before taxes</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Net sales</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Total assets</td>
<td>16</td>
<td>22</td>
<td>18</td>
<td>56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Firm #1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income (loss) before taxes</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Net sales</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Total assets</td>
<td>11</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>56</td>
</tr>
</tbody>
</table>
identify whether the responses were, on average, the same or similar for the original and repeated cases. Finally, separate one-way ANOVAs were run on the differences between the responses to the original and repeated cases to identify any systematic differences among levels of responsibility, among firms, and between audit technologies.

Across the 72 subjects in the aggregate, the correlation coefficient was .91226. Not surprisingly, as noted with respect to the income statement judgments, larger responses on the original cases tended to be paired with larger responses on the repeated cases.

Based on the paired t-tests for each case across all 72 subjects, one cannot reject the hypothesis of no difference between the responses to the original and repeated cases for any of the cases. The corresponding t-values (p-values), by case (#3, 9, 18, and 22), were as follows: -.69 (.49), -.93 (.35), -1.52 (.13), and -1.28 (.21), respectively.

Finally, three separate one-way ANOVAs (with respect to level of responsibility, firm affiliation, and audit technology) were run on the differences between the original and repeated cases, on a case-by-case basis. Level of responsibility, firm affiliation, and audit technology were not significant at the .10 level for any of the four repeated cases. Hence, no systematic differences were identified in test-retest consistency. These results suggest that the subjects were
able and willing to provide meaningful (i.e., consistent) responses to the experiment's materials.

**Order Effects**

The 24 cases were presented to the subjects in two different randomized orders. The responses to the balance sheet judgments were examined to identify whether the order of presentation significantly affected subjects' responses in terms of location and/or variation. If order of presentation influenced subjects' judgments, their responses could be regarded as artifacts of the experiment.

No differences were detected between the two orders of presentation, either as to location or variation. A one-way ANOVA of the standardized materiality scores indicated no significant difference in means between the two orders (p=.32). Cochran's test for homogeneity of variance did not indicate any difference as to variation between the two orders at the .05 level, using the sums of squares calculated with respect to each order. Apparently, the participants' judgments were not influenced by the order in which the cases were presented.


Chandler, R. "Materiality--Does It Need To Be a Guessing Game?" Accountancy, (February, 1985), pp. 84-86.


