Scholar Commons
USF University of South Florida

University of South Florida
Scholar Commons

Graduate Theses and Dissertations | Graduate School

2008

The effect of multidimensional information presentation on the effectiveness and efficiency of a spatial accounting judgment

John K. Tan
University of South Florida

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

Part of the American Studies Commons

Scholar Commons Citation

Tan, John K., "The effect of multidimensional information presentation on the effectiveness and efficiency of a spatial accounting judgment" (2008). *Graduate Theses and Dissertations.*
http://scholarcommons.usf.edu/etd/525

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

The Effect of Multidimensional Information Presentation on the Effectiveness and Efficiency of a Spatial Accounting Judgment

by

John K. Tan

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

Co-Major Professor: Uday Murthy, Ph.D.
Co-Major Professor: Jacqueline Reck, Ph.D.
Stephanie Bryant, Ph.D.
Brad Schafer, Ph.D.
Rosann Collins, Ph.D.

Date of Approval:
July 10, 2008

Keywords: three-dimensional, trend analysis, pattern recognition, cognitive fit.

© Copyright 2008, John K. Tan

Dedication

I would like to dedicate this dissertation to my wonderful parents for their unconditional love, support, and encouragement.

Acknowledgments

First and foremost, I would like to thank the co-chairs of my dissertation committee, Dr. Uday Murthy and Dr. Jacqueline Reck, for their guidance, feedback, and support throughout the process.

I would like to thank the other members of my dissertation committee, Dr. Stephanie Bryant, Dr. Brad Schafer, and Dr. Rosann Collins for their support and comments.

I would also like to express my appreciation for the valuable learning environment, especially the Friday workshops, at the University of South Florida. I acknowledge the support from all the faculty and staff of the College of Business throughout my doctoral program.

Table of Contents

List of Tables vii

List of Figures x

Abstract xi

Chapter 1: Introduction 1

1.1 Introduction and Significance of the Issue 1

1.2 The Need for Research on the Issue of Visualization 3

1.3 Research Objective and Question 4

1.4 Results and Contribution 5

Chapter 2: Background, Theory, and Hypotheses Development 8

2.1 What is the Problem? 8

2.1.1 Configural Information Processing 8

2.1.2 Analytical Procedures in Auditing 10

2.2 Visualization 14

2.2.1 Accounting Research on Visualization 15

2.2.2 Multidimensional Accounting Information Research 16

2.2.3 Research in Computer Science and Human Factors 17

2.2.4 Conceptual Limitation on Visualization Using 2-D Graphs with Three Components 18

2.2.5 Informationally Equivalent Presentations 19

2.3 Theory and Hypotheses 27

2.3.1 The Need for a Fit between Presentation Format and Task 27

2.3.2 Trend Analysis and Hypotheses 30

2.3.3 Pattern Analysis and Hypotheses 34

Chapter 3: Research Design 39

3.1 Research Model 39

3.2 The Independent Variables 42

3.3 The Dependent Variables 42

3.4 Experimental Setup and Procedures 47

3.5 Training 48

3.5.1 Training for the Trend Analysis Task 49

3.5.2 Training for the Pattern Recognition Task 52

3.6 Measuring the Dependent Variables 55

3.6.1 Measuring the Dependent Variables for the Trend Analysis Task 55

3.6.2 Measuring the Dependent Variables for the Pattern Recognition Task 60

3.7 Covariates 65

3.7.1 Practice Questions 65

3.7.2 Mental Rotations Test 65

3.7.3 Gender and Age 67

3.7.4 Mental Workload 67

3.7.5 Demographics Data 69

3.8 Post Hoc Analysis (Survey Questions) 70

3.9 Student Participants 70

3.10 Manipulation Check 73

Chapter 4: Pilot Study 82

4.1 Research Design of the Pilot Study 82

4.2 Results of the Pilot Study 83

4.2.1 Descriptive Statistics, Test of Assumptions and Outliers 83

4.2.2 Hypotheses Testing of the Trend Analysis Task 94

4.2.2.1 Results of H1a 94

4.2.2.2 Results of H1b 103

4.2.2.3 Results of H1c 110

4.2.2.4 Results of H1d 111

4.2.3 Hypotheses Testing of the Pattern Recognition Task 113

4.2.3.1 Results of H2a 113

4.2.3.2 Results of H2b 118

4.2.3.3 Results of H2c 119

4.2.3.4 Results of H2d 124

4.3 Lessons Learned from the Pilot Results and How to Improve 129

4.3.1 Rank Order Effects 129

4.3.2 Insufficient Training on Display Formats 129

4.3.3 Insufficient Training on ROE 130

4.3.4 Scrolling Up and Down the Screen 130

Chapter 5: Main Experiment 133

5.1 Sample Size, Inter-Coder Reliability 133

5.1.1 Sample Size 133

5.1.2 Inter-Coder Reliability 133

5.2 Testing of Raw Data 140

5.2.1 Testing of Outliers and Influential Observation 140

5.2.2 Testing of Assumptions 141

5.2.3 Plan of Statistical Analysis 146

5.2.4 Cronbach's Alpha 146

5.3 Manipulation Check 147

5.3.1 Results of the Manipulation Check Questions 148

5.4 Results of the Trend Analysis Task 152

5.4.1 Descriptive Statistics 152

5.4.2 Pearson's Product-Moment Correlation 161

5.4.2.1 Correlation between Dependent Variables 161

5.4.2.2 Multicollinearity 163

5.4.2.3 Correlation between Treatment or Covariate and Dependent Variables 163

5.4.3 Regression Analysis of Possible Covariates 173

5.4.4 Results of H1a 175

5.4.5 Results of H1b 183

5.4.6 Results of H1c 191

5.4.7 Results of H1d 198

5.5 Results of Pattern Recognition Task 201

5.5.1 Descriptive Statistics 201
5.5.2 Pearson's Product-Moment Correlation 209
5.5.2.1 Correlation between Dependent Variables 209
5.5.2.2 Multicollinearity 210
5.5.2.3 Correlation between Treatment or Covariate and Dependent Variables 210
5.5.3 Regression Analysis of Possible Covariates 218
5.5.4 Results of H2a 220
5.5.5 Results of H2b 226
5.5.6 Results of H2c 232
5.5.7 Results of H2d 237
5.6 Post Hoc Analysis 246
5.6.1 Trend Analysis Post Hoc Analysis 247
5.6.2 Pattern Recognition Post Hoc Analysis 251
Chapter 6: Discussion 255
6.1 Summary of Hypothesized Results 255
6.2 Discussion of Results of Hypothesis H1 264
6.2.1 Discussion of Results of Hypothesis H1a 264
6.2.2 Discussion of Results of Hypothesis H1b 268
6.2.3 Discussion of Results of Hypothesis H1c 269
6.2.4 Discussion of Results of Hypothesis H1d 270
6.3 Discussion of Results of Hypothesized H2 271
6.3.1 Discussion of Results of Hypothesis H2a 271

6.3.2 Discussion of Results of Hypothesis H2b 272

6.3.3 Discussion of Results of Hypothesis H2c 273

6.3.4 Discussion of Results of Hypothesis H2d 274

6.4 Comments on the Implication of Significant Covariates 275

6.5 Discussion on the Responses of the Survey Questions 276

6.6 Contribution 277

6.7 Limitations 282

6.8 Future Research 283

References 284

About the Author End Page

List of Tables

Table 1: Tabular Display of a Company's ROE for a Period of Five Years (Case I) 20

Table 2: Tabular Display of Companies with Same Size ROE in a Year (Case II) 23

Table 3: Overview of Dependent Variables Used in the Trend Analysis Task 44

Table 4: Overview of Dependent Variables Used in the Pattern Analysis Task 46

Table 5: Training Questions and Grading Schema of the Trend Analysis Task (Pattern Recognition Task) 54

Table 6: Measurement of the Dependent Variables for the Trend Analysis Task 58

Table 7: Measurement of the Dependent Variables for the Pattern Recognition Task 63

Table 8: Demographic Questions 69

Table 9: Mental Workload Questions and Survey Questions 72

Table 10: Manipulation Checks Questions 75

Table 11: Definition of Variables Used in the Statistical Analysis 76

Table 12: Pilot Study Descriptive Statistics 87

Table 13: Multivariate Tests of H1a 95

Table 14: Pilot Test Results of H1a 97

Table 15: Pilot Test Results of H1a 99

Table 16: Pilot Test Results of H1a 101

Table 17: Multivariate Tests of H1b 104

Table 18: Pilot Test Results of H1b 107

Table 19: Pilot Test Results of H1b 109

Table 20: Multivariate Tests of H1c 111

Table 21: Multivariate Tests of H1d 112

Table 22: Multivariate Tests of H2a 114

Table 23: Pilot Test Results of H2a 117

Table 24: Multivariate Tests of H2b 118

Table 25: Multivariate Tests of H2c 120

Table 26: Pilot Test Results of H2c 122

Table 27: Multivariate Tests of H2d 125

Table 28: Pearson's Product-Moment Correlation Coefficient of the Scores Assigned by Coder One and Coder Two to the Trend Analysis Task 136

Table 29: Pearson's Product-Moment Correlation Coefficient of the Scores Assigned by Coder One and Coder Two to the Pattern Recognition Task 138

Table 30: Results of Testing of Normality, Testing For Homogeneity of Variance 143

Table 31: Results of the Manipulation Check Questions of the Trend Analysis Task 151

Table 32: Results of the Manipulation Check Questions of the Pattern Recognition Task 151

Table 33: Descriptive Statistics for the Trend Analysis Task 156

Table 34: Pearson's Product-Moment Correlation Coefficient of the Variables of the Trend Analysis Task 165
Table 35: Test Results of H1a 180
Table 36: Test Results of H1b 188
Table 37: Test Results of H1c 195
Table 38: Test Results of H1d 200
Table 39: Descriptive Statistics for the Pattern Recognition Task 205
Table 40: Pearson's Product-Moment Correlation Coefficient of the Variables of the Pattern Recognition Task 212
Table 41: Test Results of H2a 224
Table 42: Test Results of H2b 229
Table 43: Test Results of H2c 235
Table 44: Test Results of H2d 242
Table 45: Mean, Standard Deviation, Range and ANOVA Pairwise Comparison (Mean) of the Survey Questions Assessing Display Usefulness and Ease of Use, for the Trend Analysis Task 250
Table 46: Mean, Standard Deviation, Range and ANOVA Pairwise Comparison (Mean) of the Survey Questions Assessing Display Usefulness and Ease of Use, for the Pattern Recognition Task 254
Table 47: Summary of the Results of the Trend Analysis Task 262
Table 48: Summary of the Results of the Pattern Recognition Task 263

List of Figures

Figure 1: 2-D Display of a Company's ROE for a Period of Five Years (Case I) 21

Figure 2: 3-D Perspective Display of a Company's ROE for a Period of Five Years (Case I) 22

Figure 3: 2-D Display of Companies with Same Size ROE in a Year (Case II) 25

Figure 4: 3-D Perspective Display of Companies with Same Size ROE in a Year (Case II) 26

Figure 5: Independent and Dependent Variables at Conceptual and Operation Level 40

Figure 6: Research Model 41

The Effect of Multidimensional Information Presentation on the Effectiveness and Efficiency of a Spatial Accounting Judgment

John K. Tan

ABSTRACT

This study is the first in a series of planned studies on the application of multidimensional visualization of business information and data within the context of accounting. The study's research question is: When is multidimensional visualization of information a better problem representation, improving both the effectiveness and efficiency of a spatial accounting judgment?

To examine when multidimensional visualization can assist auditors in configural cue pattern recognition, the study employs the traditional DuPont analysis as the three pieces of key information to be represented on the X, Y, and Z axes of a single 3-D perspective display. To help determine when use of 3-D perspective display is beneficial in combining pieces of information, I rely on Vessey's (1991) Cognitive Fit Theory, and the Proximity Compatibility Principle (PCP) proposed by Wickens and Carswell (1995).

The study has two hypotheses. Hypothesis H1 predicted that participants viewing a set of 2-D displays will be the most effective or most efficient in generating hypotheses for what caused the changes in the trend of accounting data or in estimating values. Hypothesis H2 predicted that participants viewing a single 3-D perspective display will be the most effective or most efficient in recognizing patterns of accounting data or in generating hypotheses for what caused the emerged pattern.

To test the hypotheses of the study a 3 x 2 between-subjects design (display format x task) is used. The independent variables are display types and task types. Graphical display was manipulated at three levels: no graphical display (table only), 2-D display, and 3-D perspective display. Task was manipulated at two levels: trend analysis and pattern recognition task.

The need for a fit between different types of spatial tasks and display formats is demonstrated by the findings of this study: 1) that 2-D displays appear to be more suitable for spatial tasks involving the generation of hypotheses for causes of trends in accounting data, while 2) 3-D perspective displays appear to be more suitable for spatial tasks involving pattern recognition in accounting data.

Chapter 1: Introduction

1.1 Introduction and Significance of the Issue

In the late 1950s, Miller (1956) reported that a human can enlarge his or her problem solving capabilities by using visualization to chunk information. Information visualization is a process that transforms data into a form that allows viewers to visually perceive the meaning of the information without having to rely on their cognitive powers to perform the necessary transformations (Zhang, 1996). A recent article in the Wall Street Journal by Totty (Sep 24, 2007) described how the corporate world is using visualization techniques, including animated graphics and three-dimensional charts, to reveal underlying patterns in complicated data and to make quicker decisions. Moriarity (1979) and Dull and Tegarden (1999) have also demonstrated the potential benefits of visual representation of multidimensional accounting information.

Research in computer science, human factors, and aviation engineering has moved beyond two-dimensional (2-D) analysis and found positive effects using three-dimensional (3-D) visualization of objects. Kolata (1982) reported that statisticians and computer scientists, through computer-motion graphical displays of multidimensional data, were able to see patterns in data that would not have been picked up with statistical techniques. As explained by Kolata (1982, p.919), three-dimensional displays of data enable scientists to make use of the uniquely human ability to recognize meaningful patterns in the data.

Humans have distinctive visual abilities, such as the ability to accurately identify twenty-four data positions in a square (Miller, 1956 p.87) or the ability to retrieve information and recognize patterns from visual cues (Kosslyn, 1994). With visual abilities, humans can solve problems through the associative system instead of the rule-based system (Sloman, 1996). According to Sloman (1996), computations based on the rule-based system reflect rules and logical content, while computations based on the associative system reflect similarity structure and relations of temporal contiguity. This suggests that problem solvers have the ability to make a judgment or decision through the use of three-dimensional charts. Three-dimensional charts highlight temporal contiguity and facilitate usage of the associative system, reducing the decision makers' cognitive load, and allowing for better decision making.

Three-dimensional visual representation has interested accountants for several reasons. Wright (1997, p. 66) indicates that one "can convey far more information in 3-D than with 2-D bar charts, or rows and columns of numbers." Wright (1997) provides examples of three-dimensional visualization of business data, ranging from a fixed-income portfolio to a retail analysis tool showing aggregate regional and individual store performance. The fact that three-dimensional charts can convey additional information suggests that decision makers can use three-dimensional charts to improve both their effectiveness (accuracy) and efficiency (time on task) in decision making.

1.2 The Need for Research on the Issue of Visualization

Three-dimensional visual representation has also interested auditors. Representing business data three-dimensionally through the X, Y and Z axes not only transforms the data into an analogy of the problem space, but also encourages the viewer to simultaneously consider information from all three dimensions. Therefore, if an auditor's task requires the combination of cues and configural information processing, a decision aid could assist the auditor by visually displaying cues in a 3-D perspective. Bedard and Biggs (1991) report that, when performing analytical procedures, auditors failed not only to reason with the combination of all crucial cues but also to generate an accurate hypothesis to explain the causal agent underlying the pattern and relationships among pieces of financial information. As Bedard and Biggs (1991) suggest, audit efficiency and effectiveness depend on competency in recognizing patterns in financial data as well as in hypothesizing likely causes of those patterns to serve as a guide for deploying appropriate audit procedures. The ability of auditors to recognize patterns and make hypotheses may be negatively affected given that auditors working under time constraints tend to make inferior judgments, as highlighted by the reports of the American Institute of Certified Public Accountants (1978, 1987). Indeed, Choo and Firth (1998) find that auditors did not invoke configural information processing under high time pressure.

1.3 Research Objective and Question

This study is the first in a series of planned studies on the application of multidimensional visualization of business information and data within the context of accounting (broadly defined to include financial, auditing, and managerial). It is the objective of the study to explore when the use of a single 3-D perspective display is beneficial in combining pieces of financial information. In the study, I investigate the following research question: When is multidimensional visualization of information a better problem representation, improving both the effectiveness and efficiency of a spatial accounting judgment?

As will be discussed, two streams of literature in auditing have called for the development of a decision aid to assist auditors in recognizing configural cue patterns or relationships between pieces of financial information. With a decision aid that combines all crucial cues, auditors should find it easier to generate hypotheses to explain the causes of identified patterns, and make more accurate global judgments or predictions.

To examine when multidimensional visualization can assist in configural cue pattern recognition, the study employs the traditional DuPont analysis as the three pieces of key information to be represented on the X, Y, and Z axes of a single 3-D perspective display. Return on Equity (ROE) is the multiplicative function of profitability (income/sales), turnover (sales/average total assets), and leverage (average total assets/average total equity) (Robinson, Munter and Grant 2003). DuPont analysis requires that individuals consider all three components of ROE (i.e., process the information configurally) to generate hypotheses regarding what caused the changes in ROE. When making accounting judgments, individuals also need to discern any emerging pattern or

trend in comparing the company to its peers' ROEs or the industry ROEs, in order to make accounting judgments.

To help determine when use of 3-D perspective display is beneficial in combining pieces of information, I rely on Vessey's (1991) Cognitive Fit Theory, and the Proximity Compatibility Principle (PCP) proposed by Wickens and Carswell (1995); both of which are discussed in Chapter 2.

1.4 Results and Contribution

Rather than being the most effective (accurate), the results suggest that participants viewing the 2-D displays can sometimes be more effective in generating hypotheses for what caused the changes in the trend of accounting data when compared to those participants viewing the 3-D perspective display.

Contrary to expectation, results suggest that participants viewing the 2-D displays failed to be more effective (accurate) or more efficient (used less time) than participants viewing the tabular display or participants viewing the 3-D perspective display in estimating values from existing data on hand. In fact, results show that the static three-dimensional perspective representation of DuPont analysis can, under certain conditions, improve participants' accuracy of prediction or estimation of values such as the return on equity (ROE).

Results of this study empirically demonstrated the benefits of using the 3-D perspective display to recognize patterns in accounting data. Benefiting from the emergent features within a three-dimensional space, participants viewing the 3-D perspective display were more effective (accurate) than either those participants viewing a tabular display and or those participants viewing the 2-D displays while performing the

task of recognizing patterns in accounting data. Furthermore, results suggest that participants viewing the 3-D perspective display can sometimes be more efficient or use less time while performing the task of recognizing patterns in accounting data.

Contrary to expectations, results suggest that participants viewing the 3-D perspective failed to be more effective (accurate) than participants viewing the tabular display or participants viewing the 2-D perspective displays in generating hypotheses for what caused the emerged patterns in the accounting data. In terms of efficiency in generating hypotheses for what caused the emerged patterns in the accounting data, mixed results were found.

The study contributes to the literature in several ways. The 3-D perspective display of the DuPont Analysis, newly developed for this study, is the first of its kind in the accounting literature. The study is the first research that technically develops a 3-D perspective display to represent a company by a point in a three-dimensional space in terms of the DuPont analysis. As shown by this study, advances in technology have made it feasible to provide and study the impact of 3-D perspective displays on information processing and decision-making.

One contribution of the study is the testing of findings from human factors research in an accounting context. This study contributes to the literature on decision aids by demonstrating how a 3-D perspective display can integrate information from the X, Y and Z axes to help decision makers invoke configural information processing. In this study DuPont analysis is used as an example. However, accountants and auditors frequently are called upon to integrate information that could benefit from a display perspective that causes users to invoke configural information processing.

The study contributes to the literature by demonstrating how Cognitive Fit Theory applies to different subtypes of spatial tasks. The need for a fit between different types of spatial tasks and display formats is demonstrated by the findings of this study: 1) that 2-D displays appear to be more suitable for spatial tasks involving the generation of hypotheses for causes of trends in accounting data, while 2) 3-D perspective displays appear to be more suitable for spatial tasks involving pattern recognition in accounting data. Thus, when developing decision aids, it is important to align the task and the aid to maximize the effectiveness of the decision process.

Chapter 2: Background, Theory, and Hypotheses Development

2.1. What is the Problem?

As discussed in the next two sections, two streams of auditing research, configural information processing and analytical procedures, have called for the development of a decision aid to help auditors more quickly recognize configural cue patterns, or patterns and relationships between pieces of financial information. Several types of decision aids can potentially assist in the recognition of configural cue patterns and relationships. As explained in later sections of this chapter, certain types of aids may be more appropriate for certain types of tasks. Of major interest in the current study is whether a three-dimensional perspective, new in this study, can provide greater benefit than more traditional types of aids in certain decision situations.

2.1.1 Configural Information Processing

Prior to the publications of Brown and Solomon (1990, 1991), auditing studies like Ashton's (1974) were unable to provide evidence about auditors' configural information processing. Employing ANOVA to model subjects' judgments, Ashton (1974) documented that the majority of the variance in judgments was due to information cue main effects; the single interaction effect explained less than four percent of the variance in auditors' judgments. Studies such as Ashton's resulted in auditors' judgments being characterized primarily as relying on independent rather than patterned (configural) cue usage (Brown and Solomon, 1990).

However, Brown and Solomon (1990, p. 19) indicate that, "Configural information processing is cognition in which the pattern (or configuration) of stimuli is important to the subsequent judgment/decision." The key contribution of Brown and Solomon (1990) is in successfully documenting that auditors do invoke configural information processing. In the Brown and Solomon (1990) study, auditors had drawn on their task-specific knowledge in the context of the cash cycle to assess control risk related to cash disbursements procedures. Of the seventy-four auditors, the variance in judgments of thirty (40.50%) was attributable to configural information processing strategies.

Both the Cohen Commission (American Institute of Certified Public Accountants, 1978) and the Treadway Commission (American Institute of Certified Public Accountants, 1987) reported that time pressure is a major cause of dysfunctional auditor judgments. Choo and Firth (1998) extended Brown and Solomon (1991) by investigating the effect of time pressure on auditors' configural information processing. Choo and Firth (1998) replicated one of the cases from Brown and Solomon (1991) on fifty-one auditors who were randomly assigned to one of three levels of time pressure: high (twenty seconds per case), moderate (one minute per case), and low (two minutes per case). Comparing the individual auditor's judgment variance across the three time-pressure conditions, Choo and Firth (1998) found that the interaction terms were significant only in the low time pressure condition. The interaction terms had little or no explanatory power in the moderate and high time pressure situations; under these constraints auditors did not invoke configural information processing. Choo and Firth (1998) explained that if auditors do not have sufficient time to retrieve specific patterns

of cue interactions from their knowledge structure, auditors will invoke a cognitive simplification strategy instead of employing configural information processing. In light of these findings, citing Sen and Wallace (1991), Choo and Firth (1998, p. 29) suggested that "sophisticated systems of expert measurement and mechanical combination could be developed as a decision aid that helps auditors to more quickly recognize configural cue patterns for expert judgment under time constraint."

Choo and Firth's (1998) call for a mechanical combination to help auditors to recognize configural cue patterns for decision making provides motivation for the current study which seeks to examine whether decision aids in the forms of two dimensional perspective displays and three-dimensional perspective displays improve the effectiveness and efficiency of decisions.

2.1.2 Analytical Procedures in Auditing

Bedard and Biggs (1991) investigated whether auditors were able to recognize patterns in accounting data and generate hypothesized causal explanations while performing analytical procedures in accordance with SAS No. 56; the auditing standard requiring the use of analytical procedures in the planning and overall review stages of all audits. While prior accounting studies of pattern recognition have focused on time-series extrapolation of financial trends, Bedard and Biggs (1991) define pattern recognition in analytical procedures to include not only recognizing relationships among pieces of financial information, but also to explaining the concepts or causal agent that underlie such relationships through hypothesis generation. Defining and recognizing such financial trends is critical because the failure to generate a correct hypothesis for

explaining discrepancies in account relationships could result in considerable wasted audit hours.

Bedard and Biggs (1991) tested twenty-one auditors with a case involving an error of capitalizing some part of selling, general and administrative expenses to inventory, which caused four discrepancies: increase in inventory, increase in income, decrease in gross margin, and no change in sales. This case was designed in such a way that a participant could not identify an error that caused the discrepancies unless those discrepancies were considered as a pattern. The auditors were asked to view projected (prepared by audit firm) and unaudited (prepared by client) financial information (six ratios and four year-end balances at 1987). Subjects were instructed that only one error caused discrepancies between projected and unaudited financial information. The client explanation for the cause of discrepancies, as included in the case, was a large year-end purchase at a favorable price. The subjects' task was to generate the most likely hypothesis about an error that could have caused the observed discrepancies.

Out of the twenty-one auditors participating in the Bedard and Biggs (1991) experiment, three auditors made acquisition errors, while four auditors failed to combine crucial cues into a pattern. Fourteen auditors recognized the pattern, yet only six auditors proposed a hypothesis consistent with the pattern. The most important finding of Bedard and Biggs (1991), in relation to this study was in documenting that, while performing pattern recognition, subjects tended to reason with one or two cues at a time instead of attempting to reason with the combination of all crucial cues. Another important finding of Bedard and Biggs (1991) was in documenting significant differences between seniors and managers in correctly generating hypotheses for the recognized pattern. Managers

had better performance due to their ability to organize diverse pieces of information into meaningful chunks. When performing analytical procedures, Bedard and Biggs (1991) highlighted the fact that auditors faced difficulties not only in combining all crucial cues to recognize patterns and relationships among pieces of financial information, but also in explaining the concept or causal agent that underlies such relationships through hypothesis generation. According to Bedard and Biggs (1991, p.636), "These subjects' lack of pattern recognition seems plausible given Kinney's (1987) conclusion that analytical procedures as performed in practice do not stress the use of data in combination."

Kinney (1987) documented the importance of the ability of auditors to recognize patterns or trends in relation to audit efficiency and effectiveness. One of the decision rules suggested by Kinney (1987) for detecting accounting errors was crude pattern analysis for the monthly cross-sections of three financial ratios: receivables turnover, inventory turnover, and cost of sales ratios. Kinney (1987) explained that if both receivables turnover and the cost of sales ratio indicate anomalies, crude pattern analysis can alert auditors to the possibility of material fictitious credit sales.

Given the fact that the auditors in the Bedard and Biggs (1991) study performed poorly despite having sufficient knowledge of accounting principles to determine which cues should be combined, Bierstaker, Bedard and Biggs (1999) extended Bedard and Biggs (1991) to provide possible explanations of why auditors were unable to correctly solve the analytical procedures and generate correct hypotheses for financial statement discrepancies. The reason why auditors were unable to use their accounting knowledge to solve the analytical procedures, as tested in Bierstaker, Bedard and Biggs (1999), was

that auditors' initial representation of the problem was unproductive and they were unable to shift to a productive representation that contained the knowledge relevant to a correct solution. Within the context of analytical procedures, Bierstaker, Bedard and Biggs (1999) defined a problem representation as the knowledge structure of client, double-entry accounting, industry patterns, error frequencies, and search processes that are held in working memory to guide current problem-solving activity. Bierstaker, Bedard, and Biggs (1999) replicated the task of Bedard and Biggs (1991) using twelve auditors to examine whether the auditors' initial representation of the problem was unproductive and next whether auditors would shift toward a productive representation after receiving prompts to activate relevant knowledge. The results revealed that all auditors initially formulated an unproductive problem representation. Before prompts were given, only one of twelve auditors shifted to a productive problem representation and correctly solved the task. After prompts were given, nine of the remaining eleven auditors correctly solved the task. These findings suggest that problem representation shifts are necessary to achieve effective decision processes.

In relation to this study, the most important finding of Bierstaker, Bedard and Biggs (1999) was the discovery of pattern recognition as an enhancement factor for the shift in problem representation; i.e., if one can see the pattern, one can then form a productive problem representation. In the Bierstaker, Bedard and Biggs (1999) study, the eleven auditors did not shift to a productive problem representation until several key discrepancies in the client financial data had been combined into a single pattern to be explained. It was critical for the eleven auditors to recognize the pattern of discrepancies before prompts could activate the knowledge that would in turn yield shifts toward

productive problem representations. Bierstaker, Bedard and Biggs (1999) suggest that the finding that prompts were not helpful for most auditors until they understood the pattern of financial discrepancies has implications for research on decision aids in auditing. The findings, of Bedard and Biggs (1991) and Bierstaker, Bedard and Biggs (1999), help inform the development of a decision aid for applying analytical procedures.

2.2. Visualization

This study explores the application of visualization techniques to the development of decision aids that can assist auditors with configural information processing and with performing analytical procedures. The study develops a decision aid that can help auditors (1) reason with the combination of all crucial cues available, (2) recognize the patterns and relationships between pieces of financial information, and (3) generate correct hypotheses to explain the causes of those patterns. The remainder of this section first provides background on visualization and multidimensional presentations by discussing the research that has been conducted. Subsequently, the section illustrates how the study creates the three-dimensional display, representing DuPont analysis that will be used in the experiment.

2.2.1 Accounting Research on Visualization

Wickens et al. (1994) defined visualization as both a display technique and a cognitive operation. This cognitive operation is recognized as the forming of a mental image about the relationships and constraints of the data. Tegarden (1999) indicates that the purpose of visualization is not to replace good solid quantitative analysis, but to allow decision makers to use their natural spatial/visual abilities to identify structure, patterns, trends, anomalies, and relationships in data such that decision makers can determine the course of further exploration. Auditors regularly perform ratio analysis to discern the possible risk of misstatement of the financial statements. If auditors can also visualize financial data and identify anomalies, auditors can then plan audit procedures accordingly.

In a well-cited study, Wright (1995) studied the effects of 2-D (two-dimensional) line and bar graphs on financial institution auditors' loan collectibility judgment. As Wright and Willingham (1997) suggest, loan collectibility is a function of the borrower's current financial condition, future cash flow from operations, and quality of borrower's past performance in meeting loan obligations. The integration of three measures that have differential weighting makes the loan collectibility judgment a complex task. When performing a complex task decision makers' cognitive effort will increase, unless information relationships are presented effectively and efficiently.

Wright (1995) tested auditors' loan collectibility judgment by evaluating the differences between tabular and tabular plus graphical presentation of three financial measures: liquidity, long-term leverage, and profitability. Control subjects received tabular only information while treatment subjects received tabular plus graphical

summarization of financial measures. Wright (1995) provided six sets of line graphs and bar charts−two sets for each of the liquidity, long-term leverage and profitability categories−to the treatment groups. Wright (1995) found a significant interaction between the complexity of the task and the incremental benefit of 2-D graphs when analyzing a complex task such as the loan collectibility judgment, which demands the integration of several sources of information−liquidity, long-term leverage, and profitability. In Wright (1995), the treatment group not only had less judgment bias but also had significantly better judgment accuracy. These findings suggest that when the number of pertinent information relationships is high and the task demands information integration, the availability of graphs improves auditors' loan collectibility judgment. The study extends the findings of Wright (1995) by investigating whether, under certain situations, multidimensional visualization technologies, such as a 3-D (three-dimensional) representation, can further improve the efficiency and effectiveness of auditors' decision making involving three financial ratios, as in the case of loan collectability judgments.

2.2.2. Multidimensional Accounting Information Research

Tegarden (1999) described business information as multidimensional and introduced visualization technologies such as 3-D scattergrams, 3-D line graphs, and volume renderings, which require a 3-D data set. Moriarity (1979) and Dull and Tegarden (1999) investigated the potential benefits of multidimensional representation of accounting information. Moriarity (1979) employed the graphic technique of Chernoff (1971) to display a company's financial situation, and demonstrated that subjects relying

on schematic faces outperformed a statistical bankruptcy prediction model. However, unlike line or bar graphs, subjects cannot retrieve actual quantitative data from the multidimensional Chernoff faces.

Dull and Tegarden (1999) studied the relationship between three visual representations of multidimensional data and the subjects' ability to make predictions based on the data. In constructing the experiment, Dull and Tegarden relied on momentum accounting, which studies companies' monthly market values over an extended period of time. Using the same companies' monthly market values over a period of 120 months, Dull and Tegarden (1999) compared the effect of 2-D and 3-D line graph presentations of the data on subjects' predictions of companies' future wealth. Dull and Tegarden (1999) found that subjects using 3-D line graphs that could be rotated provided the most accurate predictions. This finding implies that as the complexity of variables increases in terms of dimensions, decision-making accuracy can be enhanced by employing a representation with the capability of displaying the relationships among variables.

2.2.3 Research in Computer Science and Human Factors

Multidimensional visualization has been extensively researched in disciplines such as computer science and human factors. For example, through computer-motion graphical displays of multidimensional data, computer scientists were able to see patterns in data that would not have been detected with statistical techniques (Kolata, 1982).

In a human factors study, Wickens et al. (1994) conducted two experiments to compare comprehension effectiveness between 3-D perspective displays and 2-D planar

displays of multidimensional data. During the experimental session, participants were asked to view data points representing a corporation's normalized values of three variables—price, earnings, and debt. The control group viewed two 2-D orthogonal graphs, one plotting the relationship between price (Y axis) and earnings (X axis) and the other defining points on the price (Y) and debt (X) axes (Wickens et al. 1994). The objective was to determine whether participants could discern the relationships between earnings and debt. Participants in the control group were asked to compare the two 2-D graphs in order to integrate related information through the Y axis (price). The treatment group viewed a 3-D perspective representation of a cubic space of price (Y axis), earnings (X axis), and debt (Z axis). Each company was identified by one sphere in the 3-D condition and by two circles in the 2-D display (one in each 2-D graph).

Wickens et al. (1994) demonstrated that 3-D perspective displays had the advantage of dimensional integrality over 2-D planar displays. The result is that viewers of 3-D representations spend less effort in searching, scanning and comparing information because all information is displayed in a single panel. Viewers of 3-D representations also benefit from the emergence of perceptual features such as the surface created by the spatial integration of the dimensions in the 3-D perspective display.

2.2.4 Conceptual Limitation on Visualization Using 2-D Graphs with Three Components

Kumar and Benbasat (2004) suggested that information with two ratio-interval (continuous) components and one nominal component can be represented as a single 2-D line graph because the third nominal component can be represented on a 2-D plane using an appropriate visual variable (size, value, texture, color orientation and shape). Further,

Bertin (1981) suggested that in the case wherein all three components are interval or ratio (continuous) scale, the information can be represented by a pair of 2-D graphs. The information must be split between two 2-D graphs because the third continuous component cannot be represented on a 2-D plane using visual variables such as legends (Kumar and Benbasat, 2004).

Instead of representing information with three components, this study involves the plotting of the DuPont analysis [1] which has four continuous components (return on equity (ROE), turnover, profitability, and leverage). Return on equity (ROE) and profitability can be plotted on a single line graph or bar chart, as they can be scaled as percentages. Turnover and leverage can be plotted on another line graph or bar chart, as they can be scaled as multiples. A single 3-D bubble plot, however, can represent turnover (X axis), profitability (Y axis), leverage (Z axis) and ROE (bubble) simultaneously. Using a set of 2-D graphs, problem solvers must mentally combine all variables from different 2-D graphs, which invariably increases cognitive load.

2.2.5 Informationally Equivalent Presentations

One additional concept that needs to be discussed before the theory and hypothesis development is the concept of informationally equivalent representations. Since the study relies on three different visual displays, it is important that the displays be informationally equivalent. The three informationally equivalent representations used in the study are−a tabular display, a set of 2-D displays, and a single 3-D perspective

1Recall from Chapter 1 that DuPont analysis is the analysis of the components of ROE: profitability (income/sales), turnover (sales/average total assets), and leverage (average total assets/average total equity).

display. The next paragraph uses the two spatial tasks of this study (trend analysis and pattern recognition) to illustrate the concept of informational equivalency as applied to ROE, and as used in this study.

Representations are informationally equivalent if all of the information in one is also inferable from the other, and vice versa (Larkin and Simon, 1987; Simon, 1978). In this study information is presented equivalently through different display formats using two different cases of DuPont analysis (modified from White, Sondhi, and Fried, 1988) presented in a tabular display, a set of 2-D displays and a 3-D perspective display, respectively. Table 1 shows the DuPont analysis of a company's ROE for a period of five years (trend analysis task). Equivalent information can be presented in a set of 2-D displays (Figure 1), and a 3-D perspective display (Figure 2).

Table 1
Tabular Display of a Company's ROE for a Period of Five Years (Case I).

Year	Turnover	Profitability	Leverage	ROE
1	1.10	5.77%	2.26	14.34%
2	1.04	3.63%	2.24	8.46%
3	1.01	-10.37%	2.63	-27.55%
4	0.98	1.49%	3.31	4.83%
5	1.06	4.93%	3.07	16.04%

Figure 1: 2-D Display of a Company's ROE for a Period of Five Years (Case I)

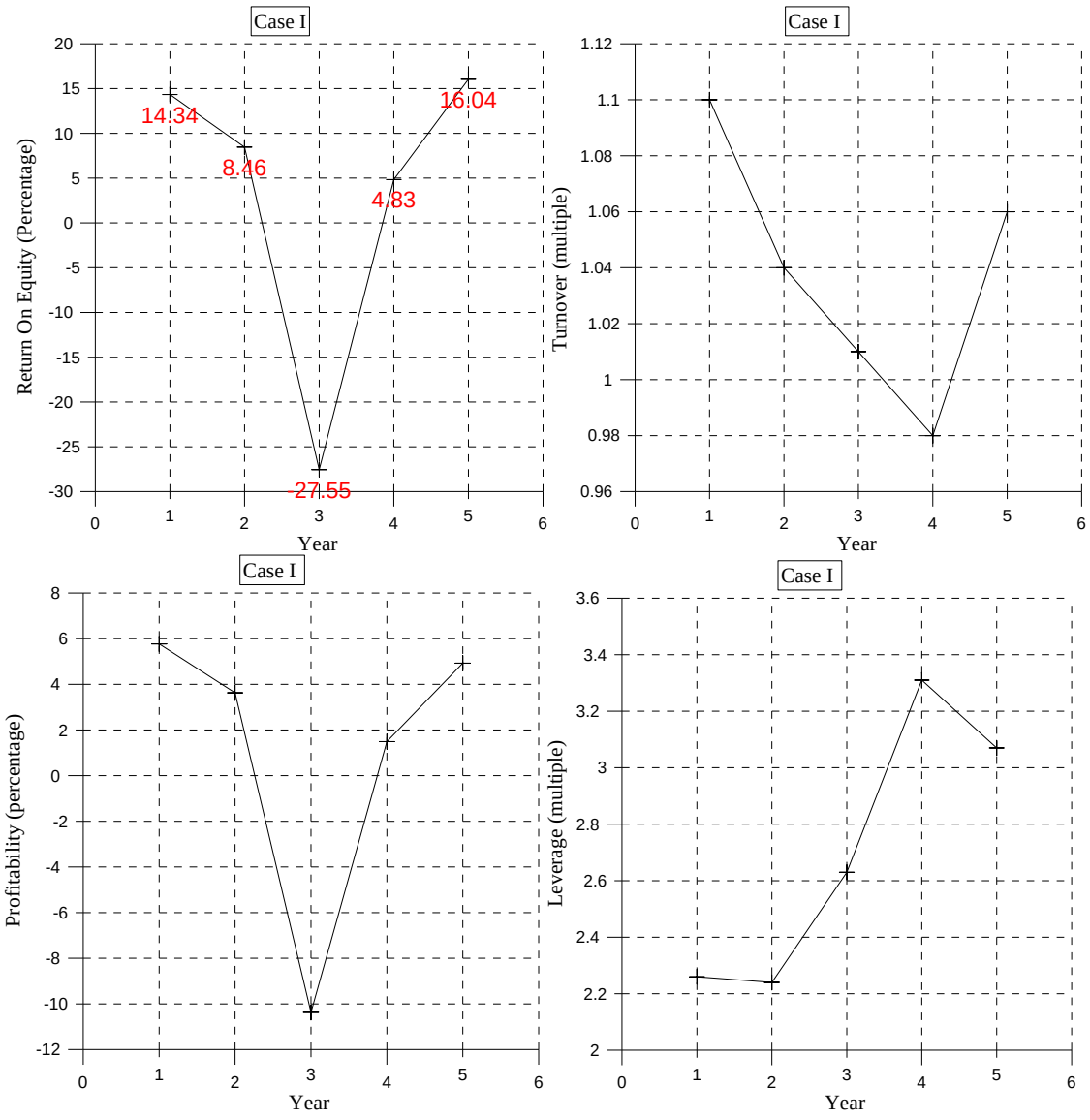


Figure 2: 3-D Perspective Display of a Company's ROE for a Period of Five Years (Case I)

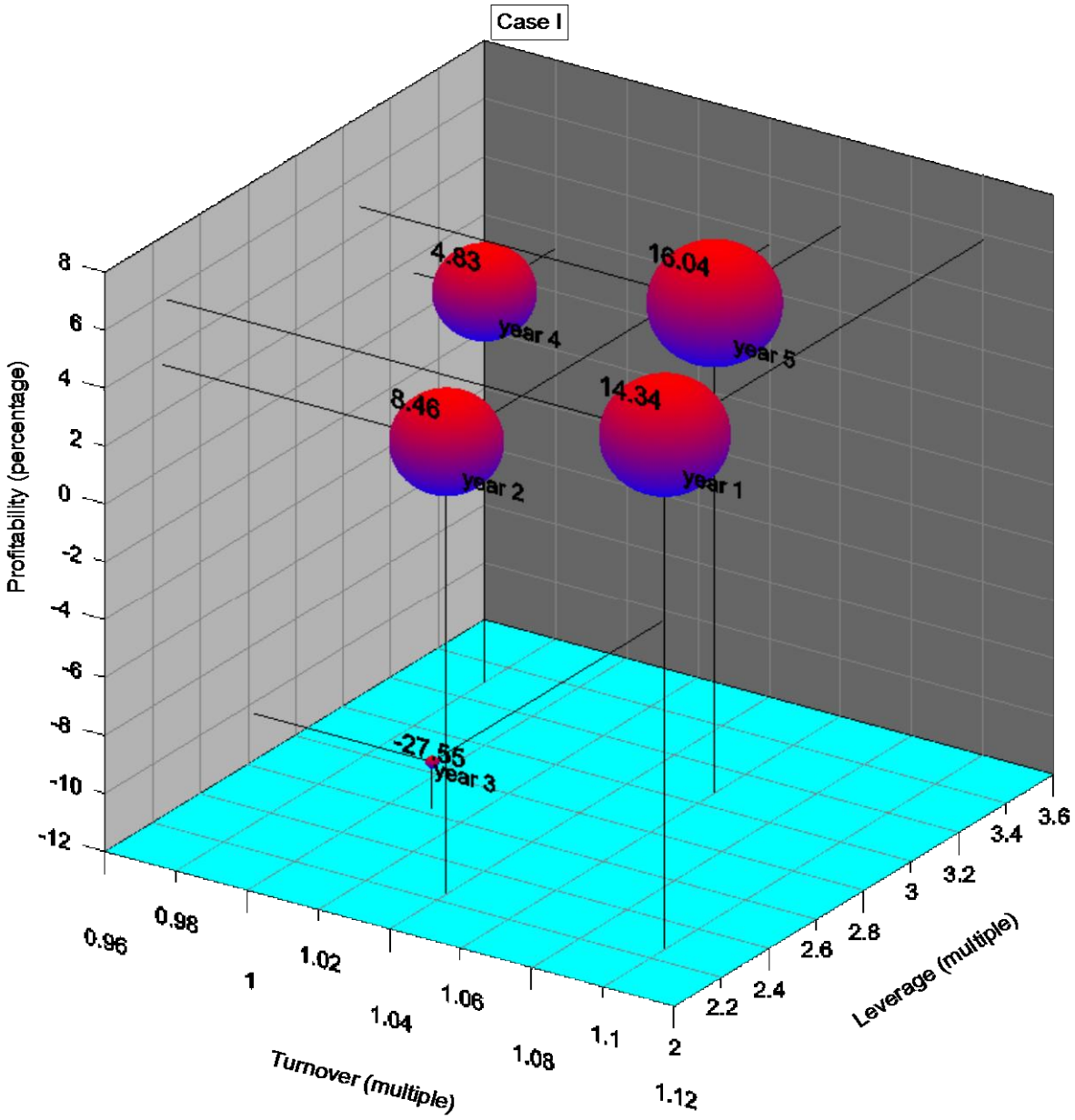


Table 2 shows the DuPont analysis of companies with the same size ROE in a year (pattern recognition task). Equivalent information can be presented in a set of 2-D displays (Figure 3), and a 3-D perspective display (Figure 4).

Table 2
Tabular Display of Companies with Same Size ROE in a Year (Case II)

Company	Turnover	Profitability	Leverage	ROE
1	1.16	2.898	2.05	6.891
2	1.56	2.231	1.98	6.891
3	0.91	3.506	2.16	6.891
4	0.79	3.826	2.28	6.891
5	1.47	2.59	1.81	6.891
6	1.04	3.012	2.20	6.891

In this study only a static 3-D graphic display will be tested; the effects of 3-D rotation and animation will be left for future research. The three components of return on equity (ROE) will be in a 3-D perspective display with the X axis as turnover (a multiple measure), Y axis as profitability (a percentage measure) and Z axis as leverage (a multiple measure). In the 3-D perspective display ROE will be shown as a bubble of varying size that indicates the multiplicative function of turnover, profitability and leverage. The actual value of the ROE will be labeled, while actual values of each of the components of ROE can be read by following the drop lines from the bubbles linking to X, Y and Z axes, respectively.

Graphical software from the Golden Software, Inc. (www.goldensoftware.com) was used to draw figures 1 to 4. According to Golden Software, Inc, the graphical software will by default draw and show the data in the best viewing position. Figures 1, 2 and 4 do not have their axes start from zero, since the graphical software, by default, does not believe that starting from zero shows the data to the viewers in the best viewing position.

If the 3-D perspective displays of Figures 2 and 4 were redrawn with the axes starting from zero, the bubbles would appear much deeper into the back wall or farther away from the viewers, making perception difficult. Although the reason is unknown, the 3-D graph used by Kuamr and Benhasat (2004), also, did not have axes start from zero. Tufte (1983) showed a 2-D line graph that did not have axes start from zero, because the range of the data were distant from the point of zero.

Since the aim of the study is to explore when the use of a single 3-D perspective display is beneficial in combining pieces of financial information, it is crucial that graphical displays are drawn to show the data to the viewers in the best viewing position. Further, the nature of this study's experimental task is spatial, which demands integration of information rather than extraction of discrete data points (symbolic task). Therefore, Figures 1, 2, 3 and 4 are drawn to display the data in the best viewing position to facilitate the integration of information.

Figure 3: 2-D Display of Companies with Same Size ROE in a Year (Case II)

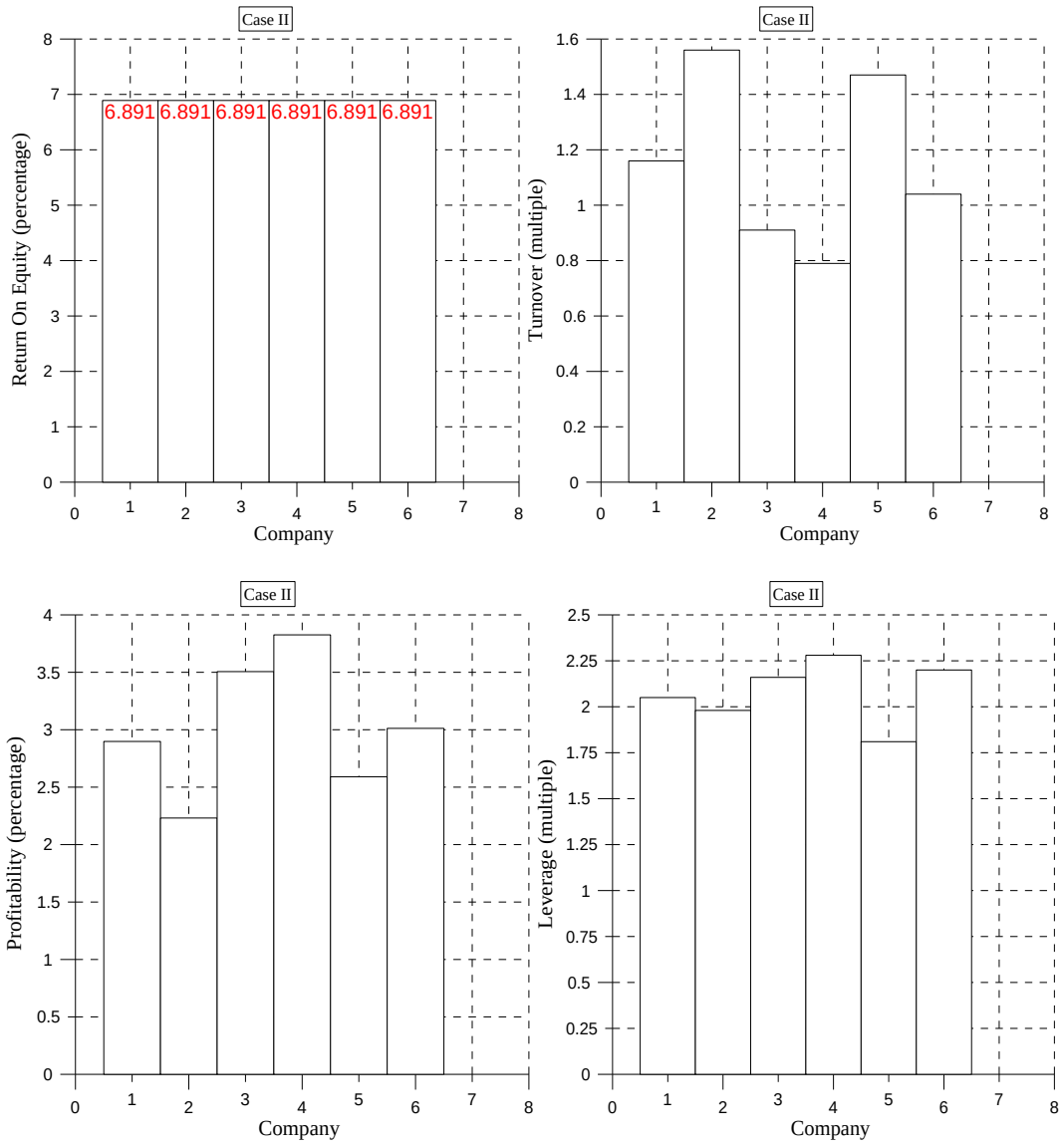
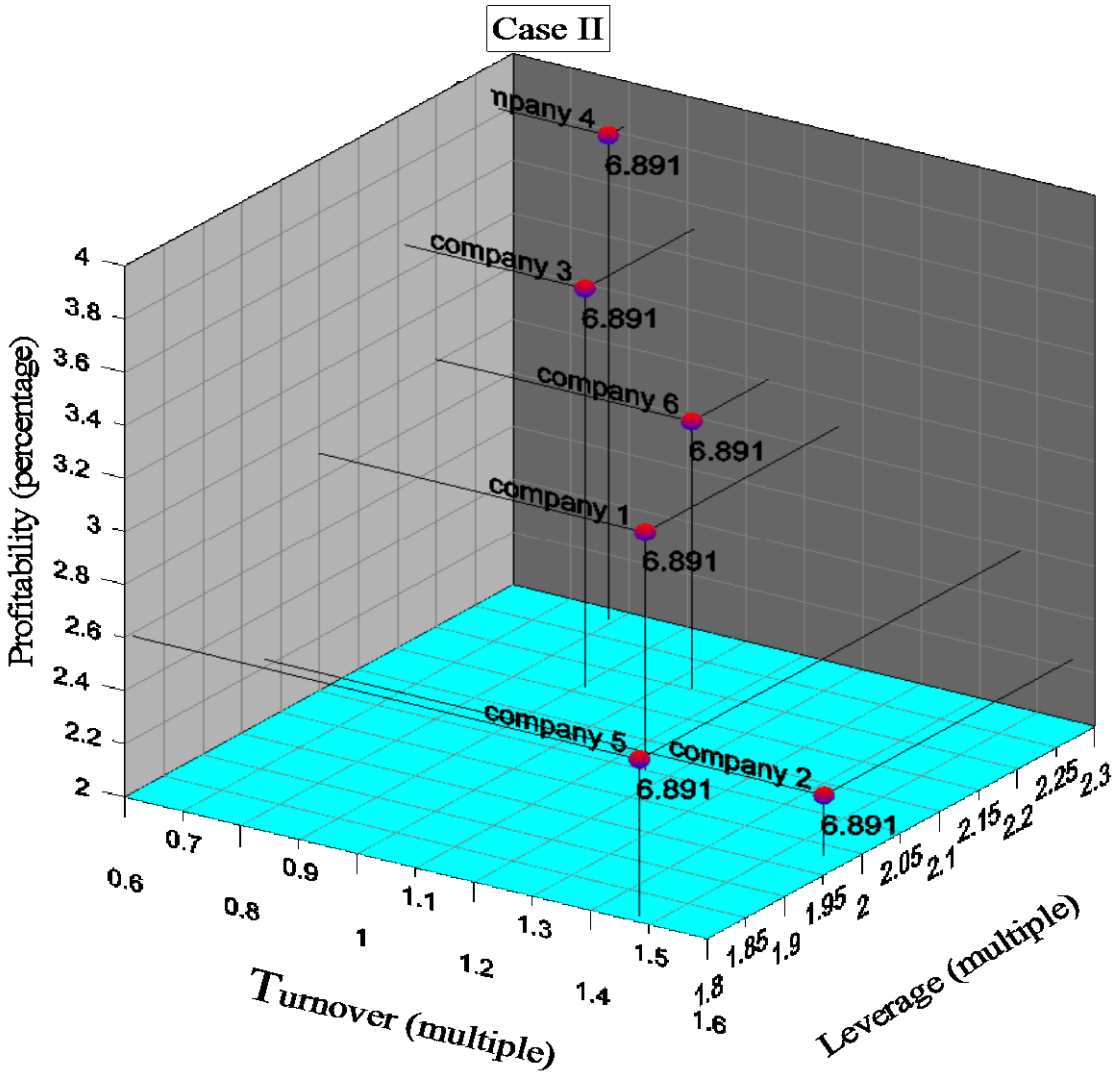


Figure 4: 3-D Perspective Display of Companies with Same Size ROE in a Year (Case II)



2.3 Theory and Hypotheses

2.3.1 The Need for a Fit between Presentation Format and Task

The Standards of Field Work of General Accepted Auditing Standards (GAAS) includes the following statement, "The auditor must obtain a sufficient understanding of the entity and its environment, including its internal control, to assess the risk of material misstatement of the financial statements whether due to error or fraud, and to design the nature, timing and extent of further audit procedures" (AICPA Professional Standards June 1, 2007, AU Section 150). Following these GAAS guidelines, participants in this study performed two tasks−one trend analysis task and one pattern recognition task.

In this study the trend analysis task is the identification and generation of hypotheses about what is happening between years in terms of the DuPont analysis of a company. The trend analysis task also includes projecting or calculating a new value from a data set on hand. The pattern recognition task is the categorization of companies with the same ROE into different groups and the generation of hypotheses to explain the differences between groups in terms of each group's financial characteristics.

Based on the discussion in the next sections I hypothesize that participants using a set of 2-D displays will be the most effective (accuracy) and the most efficient (less time) when performing the task of trend analysis. Further, I also hypothesize that participants using a single 3-D perspective display will be the most effective (accuracy) and most efficient (less time) when performing the task of pattern recognition.

Since this study is investigating performance on a spatial integration task, it is important to understand the need for a fit between presentation format and task. For this reason the definition and differences of spatial and symbolic tasks are explained. Spatial

tasks require making associations or perceiving relationships in the data (Vessey, 1991). An example of a spatial task that requires perceiving relationships in the data is: "Did sales exceed the cost of goods sold?" (Dickson, et. al. 1986) Speier (2006) suggests that spatial tasks assess the problem area holistically and often involve trend analysis or other types of information associations. Symbolic tasks, on the other hand, require a specific amount as the response, which often involves extracting discrete data values. An example of a symbolic task that requires a specific amount as the response is: "What was the company's net income for the past year?" (Dickson, et. al., 1986) Speier (2006) suggests that symbolic tasks require the recall or identification of precise data values.

Based on the preceding definitions of spatial and symbolic tasks, audit analytical procedures, such as trend analysis or pattern recognition, are basically spatial in nature. The association between spatial tasks and analytic procedures is further supported by Bedard and Biggs (1991), who suggest that the objectives of performing analytical procedures are to recognize patterns and relationships among pieces of financial information and to generate hypotheses on what caused the patterns and relationships. It is common for auditors to perform spatial tasks such as: recognizing patterns in accounting data, generating hypotheses for what caused the pattern, identifying trends in accounting data, generating hypotheses for what caused the changes in trend, and calculating a new value from the data on hand. All of these tasks are spatial because they require perceiving relationships in the data and interpolating values (Vessey 1991). Since all of these tasks are spatial, 2-D and 3-D graphic displays should result in better performance than tabular displays. According to Vessey and Galletta (1991) tabular displays are suitable only for symbolic tasks or tasks which involve extracting discrete

data values. Given a spatial task the natural question is whether a single 3-D perspective display (Figures 2 and 4) is a better presentation format than a set of 2-D displays (Figures 1 and 3).

Vessey (1991) suggests that performance on a task will be enhanced when there is a cognitive fit or match between the information emphasized in the representation type and the task type. Cognitive fit occurs through mental representation which pertains to the way problem solving elements, including problem representation and task, are being represented in human working memory. The strategies and processes that problem solvers use to solve the problem at hand are the link between problem representation and task. When a mismatch occurs between problem representation and task, Vessey (1991) suggests that problem solvers first need to formulate a mental representation based either on the problem representation or on the task. The problem solvers then need to transform either the mental representation or the data into a form suitable for solving a particular problem. Such a two-step process consumes more cognitive effort, and can potentially result in an incomplete mental representation that will likely lead to a less accurate decision (Vessey 1994 p.107). Accordingly, cognitive fit occurs when graphs support spatial tasks, and when tables support symbolic tasks. The majority of prior studies have focused on the relative effectiveness of tables versus 2-D graphs (line graphs and bar charts).

As indicated, Vessey's (1991) Cognitive Fit Theory has primarily been studied by comparing the relative fit between tables or 2-D graphs (line graphs and bar charts) and symbolic or spatial tasks. Cognitive Fit Theory has not been used to examine how fit occurs between different presentation formats (a single 3-D perspective display or a set of

2-D displays) and a spatial task. This study applies Cognitive Fit Theory in the context of a comparison of a set of line graphs or bar charts (2-D displays) with a single 3-D perspective display for a spatial task. To help inform the hypotheses presented in the next section other theories and principles are used to help explain how the cognitive fit between a presentation format and a spatial task occurs. These other theories and principles help identify when the match between the representation type (presentation format) and the spatial task are optimal resulting in the least cognitive effort or best cognitive fit.

2.3.2 Trend Analysis and Hypotheses

Line graphs emphasize the direction and volatility of data over time and improve judgments requiring trend analysis (Wright 1995). According to Jarvenpaa and Dickson (1988), viewers of line graphs can see trends and relationships at a glance, avoiding the steps of reading, comparing, and interpreting that are necessary to spot deviations using tabular data. Viewers of line graphs can discern the trend relationship simply by following the changes in the slope of the line (see Figure 1). For discerning trends, viewers of a 3-D perspective display are essentially in the same position as table users (see Figure 2). In Figure 2, despite the fact that the changes in a company's ROE from year one to year five are represented through the different sizes of the bubbles and the relative positions of the bubbles the changes in each of the components of ROE are not shown explicitly.

Viewers of the 3-D perspective display in Figure 2 must first identify the correct bubble that represents year 1 to year 5, and then use the drop lines from the bubbles to

compare the differences between years for each variable of the ROE. In doing so, the cognitive demand for 3-D viewers of Figure 2 is much greater than that for the 2-D line graph viewers using Figure 1, as the latter can discern changes through the slope of the line. The line graphs in Figure 1 permanently display the slope of a trend relationship to viewers. Thus, there is a better cognitive fit between 2-D line graphs and the task of identifying trends in accounting data and generating hypotheses for what caused the changes in trend.

Prior literature such as Triffett and Trafton (2006) suggests that when information is not explicitly shown in a graph, viewers with domain-specific knowledge can infer implicit information through a mental process called spatial transformation. Spatial transformations occur when a spatial object is transformed from one mental state or location to another mental state or location. Triffett and Trafton (2006) provided a number of examples of spatial transformations; among them are creating a mental image, modifying that mental image by adding or deleting features, mental rotation (Shepard & Metzler, 1971), mentally moving an object, animating a static image (Bogacz & Trafton, 2005), making comparisons between different views (Trafton et al., 2005), as well as any other mental operation that transforms a spatial object from one state or location into another.

Triffett and Trafton (2006) described a meteorologist spatially transforming the position of a low pressure system toward a certain direction (by hand gesture) even though actual movement of the low pressure system was not explicitly shown in a graph. Similarly, viewers of the 3-D perspective display in Figure 2 can spatially create a line of slope in their mental mind, while analyzing trends in accounting data or generating

hypotheses for what caused the changes in trend. Whether auditors do spatially transform implicit information, while performing ratio analysis, is an empirical question. Given that 2-D line graphs explicitly display the trend relationship through the slope of the line (see Figure 1) there is less of a need for data to be transformed, relative to 3-D displays, allowing for a better cognitive fit.

Therefore, it is hypothesized that:

H1a: Subjects using a set of 2-D displays will be the most effective (accuracy) in generating hypotheses for what caused the changes in the trend of accounting data when compared to subjects using a single 3-D perspective display or subjects using a table.

H1b: Subjects using a set of 2-D displays will be the most efficient (less time) in generating hypotheses for what caused the changes in the trend of accounting data when compared to subjects using a single 3-D perspective display or subjects using a table.

It is common for accountants or auditors to perform spatial tasks such as estimating or calculating a new value from a data set on hand. Pinker's (1990) Graph Difficulty Principle can help us to understand why a set of 2-D line graphs is a better representation for a task involving interpolation of values. Lohse (1991) suggested that only a small fraction of the information (about 3 chunks) decoded from a graph can be held in short-term memory at one time. Reorganization and reinterpretation of the information decoded from a graph is subject to capacity and duration limitations in short-term memory (Lohse, 1991). Pinker (1990) suggested that limits on short-term memory and on processing resources will make specific sorts of information easier or more

difficult to extract. Pinker's (1990) Graph Difficulty Principle explains that it is easier to extract a particular piece of information from a graph if the graph has attached a specific message flag to that piece of information.

Viewers of line graphs can discern the trend relationship by following the changes in the slope to determine whether the data are linear (straight), not changing (flat), or increasing sharply (steep slope). The 3-D perspective display of the study cannot provide message flags with similar richness about the nature of the data. If a graph can attach a rich message flag to a piece of information, viewers can more easily encode that piece of information from the graph. Therefore, a set of 2-D line graphs is a better representation for a task involving estimation or calculation of new values. By following the slope of the line, viewers of the 2-D line graphs can roughly estimate the new value to be larger or smaller than the existing values or perhaps that it remains the same. Pinker (1990) suggested that elements in a 2-D line graph will be seen as being a smooth continuation of one another. Again, the 2-D line graph should result in less cognitive effort or a better cognitive fit.

Therefore, it is hypothesized that:

H1c: Subjects using a set of 2-D displays will be the most effective (accuracy) in an accounting judgment involving estimation of values when compared to subjects using a single 3-D perspective display or subjects using a table.

H1d: Subjects using a set of 2-D displays will be the most efficient (less time) in an accounting judgment involving estimation of values when compared to subjects using a single 3-D perspective display or subjects using a table.

Hypothesis H1 relies on different measures related to trend analysis which will be discussed in Chapter Three.

2.3.3 Pattern Recognition and Hypotheses

According to Wickens and Carswell (1995) the Proximity Compatibility Principle (PCP) consists of two key concepts: processing proximity and perceptual proximity. Processing, or mental-processing proximity defines the extent to which two or more sources of information are used as part of the same task. If different sources of information must be integrated for a single task, they have close processing proximity. PCP suggests that if there is close processing proximity, then close perceptual proximity, the notion that different sources of information will be perceived as more similar if they are displayed in close proximity, is required. This closeness in space between different sources of information, as illustrated by Wickens and Carswell's (1995) principle of perceptual proximity, decreases the cognitive effort of comparison and integration. For example, turnover, profitability, and leverage are different sources of information but they have close processing or mental proximity as they multiplicatively explain what causes the changes in the return on equity. Since turnover, profitability, and leverage have close processing proximity, they should be plotted in a single 3-D perspective display so that their close perceptual proximity enables problem solvers to visualize and integrate them into a single view.

Human effort in retrieving information or data through graphical display involves information access costs (IAC) such as movement of attention, the eye and the head (Wickens and Carswell, 1995). Displaying return on equity, turnover, profitability, and

leverage in a set of line graphs or bar charts, instead of a single 3-D perspective display, would increase IAC. An increase in IAC will disrupt performance especially on integration tasks, like the DuPont analysis, because such tasks impose additional load on working memory. To perform an integration task, working memory is used to perform multiple processes of computation as required by integration, retaining information from one source (display) while accessing new information from other sources (displays) (Liu and Wickens, 1992). Displaying different but related information in close proximity will provide a great benefit for integration tasks, particularly those that require sequential retrieval of two or more sources of information. Carswell (1992) suggests that integrative cognitive tasks should be supported by integrated perceptual representations, like 3-D perspective displays.

An important property of a 3-D perspective display is the function that allows problem solvers to see emergent features between datum. Figure 1 of Wickens et al. (1994, p.48) illustrates an example of an emergent feature, which is the surface in a 3-D space from plotting the relationship among earnings (X axis), price (Y axis) and debt (Z axis). The aforementioned surface clearly shows that as price increases, earnings also increase, while debt decreases. Without such an emergent feature, as in the case of problem solvers viewing a set of 2-D graphs, recognition of relationships between plotted variables would require the mental computation or comparison of the individual data values. Bennett and Flach (1992) suggested that a mental combination of values, which puts strain on working memory, can be replaced with an automatic perceptual operation as humans have distinctive visual abilities (Miller 1956).

PCP emphasizes that if the problem-solving task requires integration of information with close processing proximity, a 3-D perspective display has better perceptual proximity than a 2-D planar display. The latter requires mental combination while the former invokes automatic perceptual operations.

If each corporation can be defined by three financial ratios, it is possible to represent that corporation by a point in a three-dimensional space (3-D perspective). Altman et al.(1974 p.199) suggests that a 3-D perspective display can separate companies into subgroups of good or bad companies in terms of the three financial ratios as represented by the X, Y and Z axes, respectively.

While 3-D perspectives may be beneficial in identifying emergent features, 2-D perspectives in the form of bar charts are what is commonly used for extracting specific point values. According to Tan and Benbasat (1993), bar charts displaying numeric values at the top of the bar are the best type of graph for answering a wide range of questions. Though the bar charts in Figure 3 display the point values separately for each variable, these variables are not integrated or grouped to show any emergent feature or pattern. In contrast, the 3-D perspective display in Figure 4 is showing an emergent feature. Companies in Figure 4 are automatically categorized into two groups: one group is pursuing high profitability, low turnover and high leverage; the other group is pursuing low profitability, high turnover and low leverage. The 3-D perspective display in Figure 4 is a demonstration of the PCP. The relative position of each bubble or company in Figure 4 is determined by each company's turnover, profitability, and leverage. The 3-D perspective display (Figure 4) has, thus, already categorized the companies in accordance with their relative position within the three-dimensional space. In contrast, bar chart

viewers (Figure 3) must repeatedly view each variable's chart and mentally combine the information to recognize patterns in the accounting data. In doing so the cognitive load is heavy for viewers of Figure 3. Haskell and Wickens (1993) suggest that humans cannot easily integrate information across separate spatial locations. The resultant implication is that viewers of the bar charts (Figure 3) will have difficulty integrating information across four different bar charts in their effort to recognize the patterns in the accounting data. Thus, there is a better cognitive fit between a 3-D perspective display and the tasks of recognizing patterns of accounting data, and generating hypotheses for what caused the pattern.

It is hypothesized that:

H2a: Subjects using a single 3-D perspective display will be the most effective (accuracy) in recognizing patterns of accounting data when compared to subjects using a set of 2-D displays or subjects using a table.

H2b: Subjects using a single 3-D perspective display will be the most efficient (less time) in recognizing patterns of accounting data when compared to subjects using a set of 2-D displays or subjects using a table.

H2c: Subjects using a single 3-D perspective display will be the most effective (accuracy) in generating hypotheses for what caused the emerged patterns when compared to subjects using a set of 2-D displays or subjects using a table.

H2d: Subjects using a single 3-D perspective display will be the most efficient (less time) in generating hypotheses of what caused the emerged patterns when compared to subjects using a set of 2-D displays or subjects using a table.

Hypothesis H2 relies on different measures related to pattern recognition which will be discussed in Chapter Three.

Chapter 3: Research Design

3.1 Research Model

To test the hypotheses of the study a 3 x 2 between-subjects design (display format x task) is used. Two performance constructs are of interest in the study, effectiveness and efficiency. In this study I defined effectiveness as accuracy. Effectiveness is measured as the participants' accuracy in performing the tasks. Efficiency is measured as the response time related to the various tasks the participants are asked to perform. The independent variables are display types and task types. Graphical display was manipulated at three levels: no graphical display (table only), 2-D displays, and 3-D perspective display (see Tables 1 and 2, and Figures 1-3 for examples of display formats used). Task was manipulated at two levels: trend analysis and pattern recognition task. A number of covariates measures are used in this study to test for significant correlations with the dependent variables.

Figure 5 shows the independent and dependent variables at the conceptual and operational levels. Figure 6 shows the research model.

Figure 5

Independent and Dependent Variables at Conceptual and Operational Level

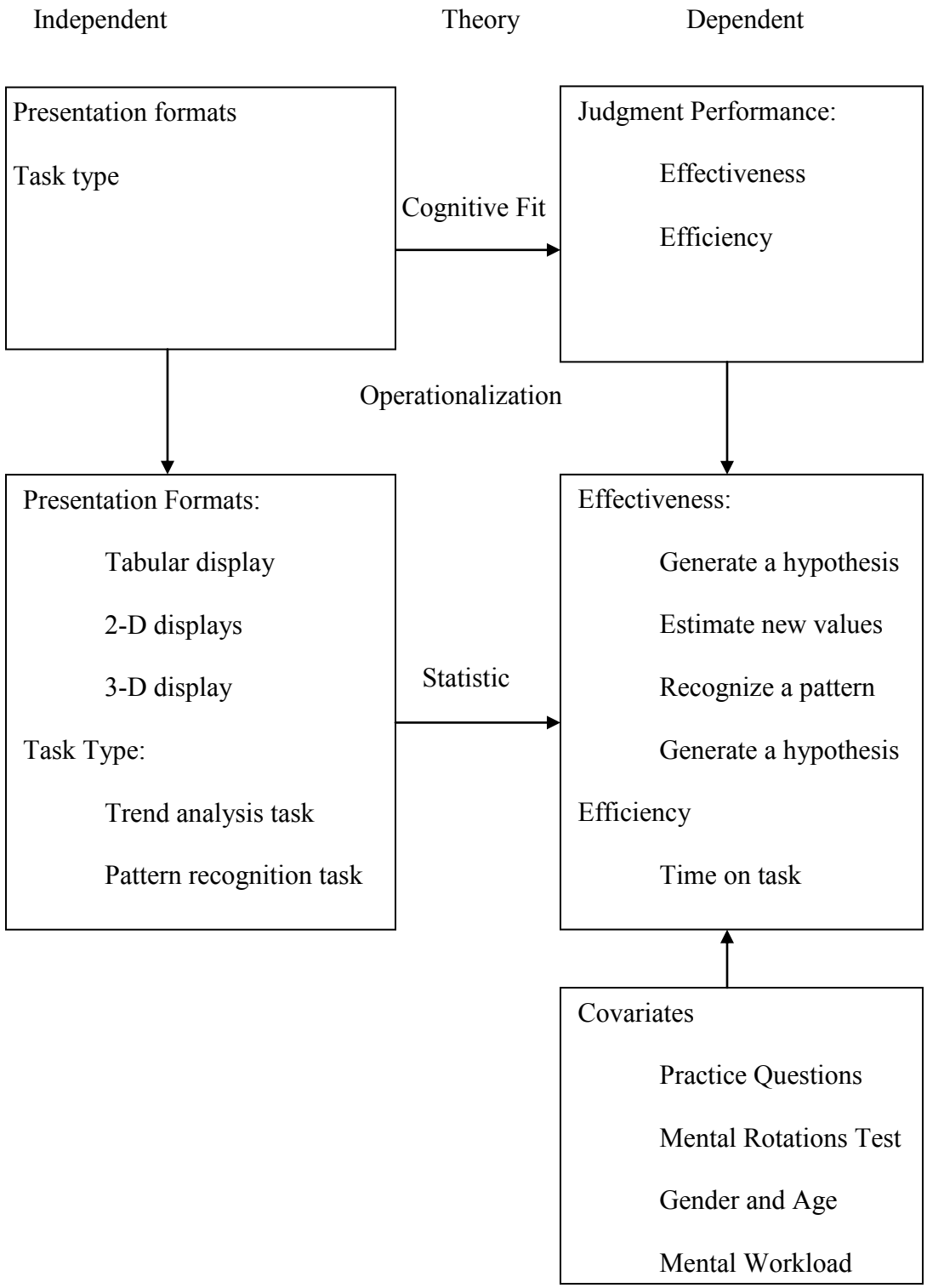
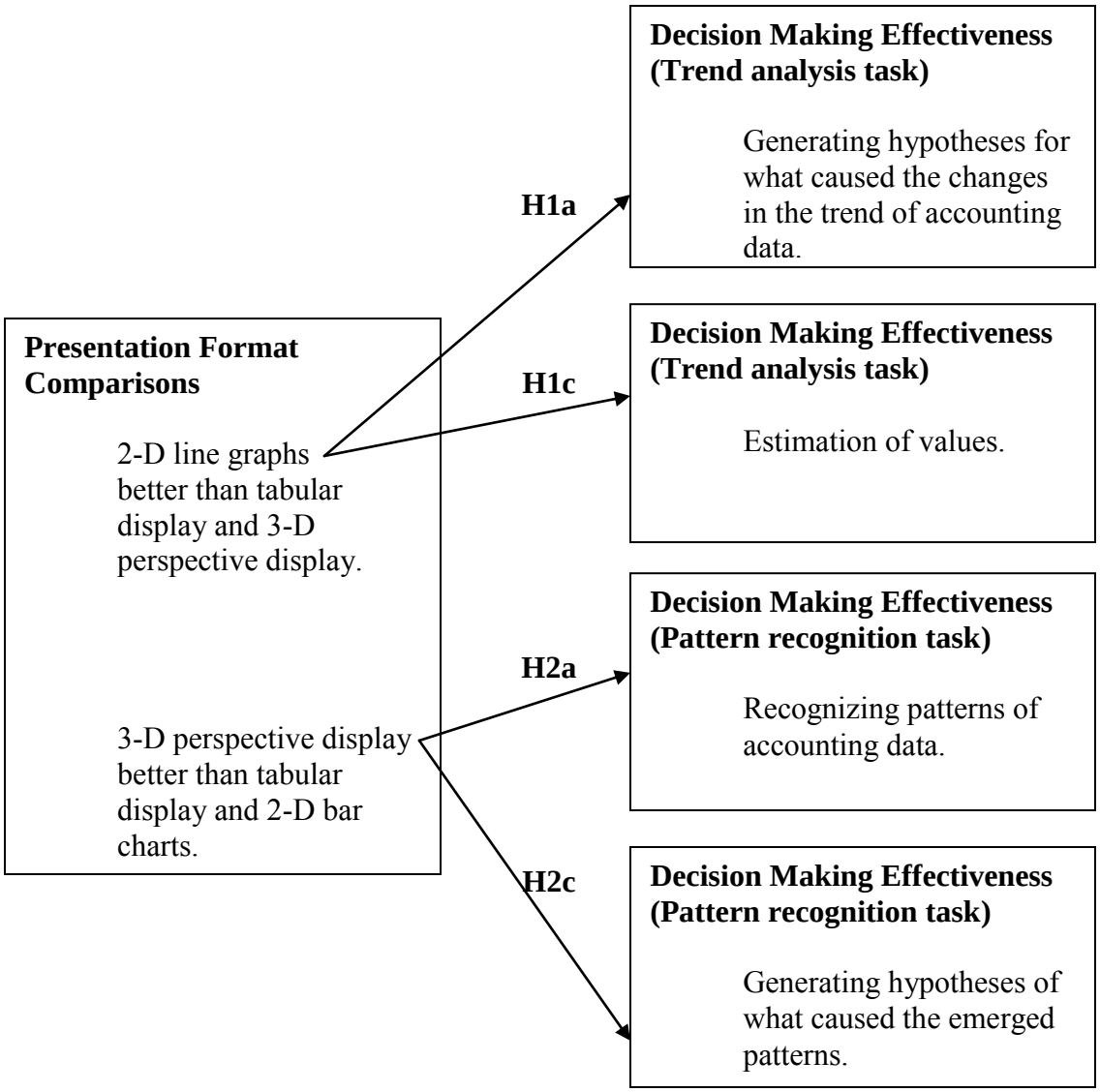


Figure 6: Research Model



3.2 The Independent Variables

The independent variables are display types and task types. Graphical display was manipulated at three levels: no graphical display (table only), 2-D displays, and 3-D perspective display (see Tables 1 and 2, and Figures 1-3 for examples of display formats used). Task was manipulated at two levels: trend analysis and pattern recognition task. The trend analysis task is defined as the identification and generating of hypotheses about what is happening between years in terms of the DuPont analysis of a company. The trend analysis task also includes projecting or calculating a new value from a data set on hand. The pattern recognition task is defined as the categorization of companies with the same ROE into different groups and the generation of hypotheses to explain the differences between different groups in terms of each group's financial characteristics.

3.3 The Dependent Variables

Two performance constructs are of interest in the study, effectiveness and efficiency. In this study I define effectiveness as accuracy in performing the tasks. The tasks include generation of hypotheses regarding what caused the changes in the trend of accounting data; estimation of new values for ROE, profitability, turnover, and leverage; recognizing patterns of financial data between companies; and generation of hypotheses regarding what caused the patterns in financial data. These same tasks are used to determine the efficiency of participants which is measured as the response time related to the various tasks the participants are asked to perform.

To test for the trend analysis task (H1a, H1b, H1c, and H1d), six questions are developed (see Table 3) and each participant answers these six questions in the same

order. The first, second, fifth, and sixth questions are used to test for H1a and H1b. The first question asks the participant to write brief sentences describing the differences between years 1 and 4. The second question asks the participants to write brief sentences describing what they perceive to be occurring in the data going from year 2 to year 3 and year 4. The fifth question asks participants to indicate the differences between years 2 and 4 by selecting choices from a given template. The sixth question asks participants to describe what they perceive to be occurring in the data from year 1 to year 2 to year 3 by selecting choices from a given template.

The third and fourth questions are used to test H1c and H1d. The third question asks the participants to estimate what the ROE would be in year 6 if each of the variables comprising ROE in year 5 had doubled. The fourth question asks the participants to estimate the average of turnover, leverage and profitability for the years 1, 2, 4 and 5, and use the estimated average to calculate a new ROE.

Table 3
Overview of Dependent Variables Used in the Trend Analysis Task

Dependent variables	Tests of	Questions developed to test hypotheses
Accuracy of hypotheses generated Time (seconds) spent on task	H1 a H1 b	What are the differences between years 1 and 4? Write short sentences using the words 'higher,' 'lower'
Accuracy of hypotheses generated Time (seconds) spent on task	H1a H1b	What is happening as you go from year 2 to year 3 to year 4? (Hint: Please identify the trends from year 2 to year 4).Write short sentences using the words 'higher,' 'lower'
Accuracy of the new values estimated Time (seconds) spent on task	H1c H1d	Based on the ROE of year 5 (Bubble 5), what would be year 6 ROE if each of the variables of ROE in year 5 had doubled?
Accuracy of the new values estimated Time (seconds) spent on task	H1c H1d	Estimate the average of turnover, leverage and profitability for the years 1, 2, 4 and 5, and use them to calculate a new ROE (Hint 1: you do not actually need to calculate the average, please consider the position of ROE as you attempt to answer. Hint 2: year 3 is not used)
Accuracy of hypotheses generated Time (seconds) spent on task	H1a H1b	What are the differences between years 2 and 4? Please answer as accurately and as fast as possible.*
Accuracy of hypotheses generated Time (seconds) spent on task	H1a H1b	What is happening as you go from year 1 to year 2 to year 3? Please answer as accurately and as fast as possible.*

* Participants selected choices from a given template.

To test for the pattern recognition task (H2a, H2b, H2c and H2d), six questions are developed (see Table 4) and each participant answers these six questions in the same order. The second and fourth questions are used to test H2a and H2b. The second question asks the participants to separate companies 1 through 6 into two groups based on similar characteristics. The fourth question asks the participants to select one of the six companies if the goal is to have high profitability, high turnover but low leverage at the same time.

The first, third, fifth and sixth questions are used to test H2c and H2d. The first question asks the participants to write brief sentences describing the differences between companies 1 and 6. The third question asks the participants to write brief sentences describing the patterns of the financial ratios they perceive in group one. The fifth question asks participants to indicate the differences between companies 4 and 6 by selecting choices from a given template. The sixth question asks participants to describe the patterns of the financial ratios they perceive in group two by selecting choices from a given template.

Each of the six questions of the trend (pattern) analysis task has its own page screen, allowing for the time spent by each participant on each question to be recorded in seconds, thus allowing for measurement of efficiency.

Table 4
Overview of Dependent Variables Used in the Pattern Analysis Task

Dependent variables	Test of	Questions developed to test hypotheses
Accuracy of hypotheses generated Time (seconds) spent on task	H2 c H2 d	What are the differences between companies 1 and 6? Write short sentences using the words 'higher,' 'lower'
Accuracy in recognizing patterns Time (seconds) spent on task	H2a H2b	Please separate companies 1 through 6 into 2 groups based on similar characteristics. Note: please assign each company only once to either group one or group two, but the groups need not have the same number of companies *
Accuracy of hypotheses generated Time (seconds) spent on task	H2c H2d	Group one includes companies 1, 3, 4, and 6, and group two includes companies 2 and 5. Compared to group two, what are the patterns of the financial ratios you are seeing in group one? Write short sentences using the words 'higher,' 'lower'
Accuracy in recognizing patterns Time (seconds) spent on task	H2a H2b	Assuming you cannot select a company solely because of a single variable, for example higher profitability. Comparatively, if it is better to have a higher profitability, higher turnover but lower leverage at the same time, which company you will select? *
Accuracy of hypotheses generated Time (seconds) spent on task	H2c H2d	What are the differences between companies 4 and 6? Please answer as accurately and as fast as possible *
Accuracy of hypotheses generated Time (seconds) spent on task	H2c H2d	Group one includes companies 1, 3, 4 and 6, and group two includes companies 2 and 5. Compared to group one what are the patterns of the financial ratios you are seeing in group two? Please answer as accurately and as fast as possible. *

* Participants selected choices from a given template.

3.4 Experimental Setup and Procedures

Two hundred and fifty eight undergraduate business students participated in the main experiment. Each participant was randomly assigned to one of the six treatment conditions: trend analysis task with tabular display, trend analysis task with 2-D display, trend analysis task with 3-D perspective display, pattern recognition task with tabular display, pattern recognition task with 2-D display, and pattern recognition with 3-D perspective display.

The experiment was conducted in a computer laboratory over a period of two weeks. All participants used the same type of computer with a 16" screen to complete the experiment. When the participants arrived at the computer laboratory they were randomly assigned to a computer. The investigator first announced that if any participant had eyesight problems he or she could not participate in the experiment (one participant with color blindness was dismissed). The investigator further announced that calculators, or any external aids were not allowed to be used during the whole experiment. The investigator then distributed and explained the informed consent forms to the participants.

After the investigator had collected all of the signed informed consent forms from participants, the investigator invited each participant to draw an envelope containing the web link to one of the six treatment conditions. All participants then typed the web link into the computer to access their randomly assigned treatment conditions. All participants started the experiment at the same time, after the investigator was satisfied that each

participant had access to a randomly assigned treatment condition. The following paragraphs describe the experimental procedures.

All participants first completed a training task to familiarize them with their assigned display format. Participants performing the trend analysis task viewed either a table, or a set of 2-D line charts, or a single 3-D perspective display showing changes in the desirability rating of apartments rented in the past six years and practiced answering four questions on trend analysis. Participants performing the pattern recognition task viewed either a tabular display, or a set of 2-D bar charts, or a single 3-D perspective display showing the differences between apartments available for rent and practiced answering four questions involving the recognition of patterns.

After all participants had answered four practice questions with their assigned display format, participants then completed a training task on DuPont analysis. All participants answered practice questions about the concept of DuPont analysis and answered another practice question involving the calculation of ROE.

Following the training on DuPont analysis, participants completed the experimental tasks. After completing the trend analysis task (see Table 3) or pattern recognition task (see Table 4), all participants completed post-experiment questionnaire consisting of a manipulation check question (see Table 24), a demographic survey (see Table 22), and the Mental Rotations Test.

3.5 Training

Training involved evaluating the desirability of apartments—a task to which college students can easily relate. The degree of desirability of an apartment to a potential

renter is represented by a desirability rating, which is a multiplicative function of three variables: income factor, size factor, and distance factor. The training task thus mirrored the experimental task, wherein return on equity is a multiplicative function of three accounting variables (return on assets, profitability, and leverage).

3.5.1 Training for the Trend Analysis Task

Participants performing the trend analysis task viewed either a tabular display, or a set of 2-D line displays or a single 3-D perspective display showing changes in the desirability rating of the apartments rented by the participants between years one to six. Participants were told to assume that for various reasons (for example roommates graduating or the landlord selling the rental property) they had rented and lived in six different apartments, with different desirability ratings in the last six years. Participants were also told to assume that in the past six years, their fixed disposable income had not changed, and their desire to rent a 1,200 square foot apartment with the lowest rent and within 5 miles of campus also had not changed.

Each participant answered four practice questions (see Table 5) by reading from either a tabular display, or a set of 2-D line displays, or a single 3-D perspective display. Bubbles on the 3-D perspective display represented an apartment's desirability rating, while the X, Y and Z axes represented the income factor, size factor and distance factor, respectively. Additionally, participants in the 3-D treatment condition learned how to read values from drop lines.

After completing the training on display format, participants completed a training task on DuPont analysis. All participants answered a practice question about the concept

of DuPont analysis and answered another practice question involving the calculation of ROE (see Table 5).

Four questions were developed to familiarize participants with their assigned display format. Each participant answered these four questions in the same order. The first question asks the participant to write brief sentences describing the values of the factors of the apartment rented in year 2. The correct answer to the first question has four parts: (a) desirability rating = 93.78, (b) size factor = 0.94, (c) income factor = 72.03, and (d) distance factor =1.39. For each correct response to one of the four parts the participant was awarded one point. The score range for the answer to the first question is zero to four points (see Table 5). This score on the first question is one of the covariate measures used to test for significant correlations with the dependent variables. The time spent in seconds by each participant when answering the first question is also one of the covariate measures used to test for significant correlations with the dependent variables.

The second question asks the participant to write brief sentences describing the differences between the apartments rented in year 5 and year 6. The correct answer to the second question has three parts: (a) the year 5 size factor is higher than year 6, (b) year 5 income factor is higher than year 6, and (c) year 5 distance factor is lower than year 6. For each correct response to one of the three parts the participant was awarded one point. The score range for the answer to the second question is zero to three points (see Table 19). This score on the second question is one of the covariate measures used to test for significant correlations with the dependent variables. The time spent in seconds by each participant when answering the second question is also one of the covariate measures used to test for significant correlations with the dependent variables.

The third question asks the participant to write brief sentences describing the values of the factors of the apartment rented in year 4. The correct answer to the third question has four parts: (a) the desirability rating = 26.03, (b) size factor = 0.97, (c) income factor = 71.38, and (d) distance factor = 0.38. For each correct response to one of the four parts the participant was awarded one point, providing a range for the answers to the third question from zero to four points (see Table 5). The score on the third question is used as a covariate measure when testing for significant correlations with the dependent variables. The time spent in seconds by each participant when answering the third question is also one of the covariate measures used to test for significant correlations with the dependent variables.

The fourth question asks the participant to write brief sentences describing the differences between the apartments rented in year 2 and year 4. The correct answer to the fourth question has four parts: (a) year 2 desirability rating is higher than year 4, (b) year 2 size factor is lower than year 4, (c) year 2 income factor is higher than year 4, and (d) year 2 distance factor is higher than year 4. Again, one point was awarded for each correct response. The score range for the fourth question is zero to four points (see Table 5). The score on the fourth question is used as a covariate to test for significant correlations with the dependent variables. The time spent in seconds by each participant when answering the fourth question is also one of the covariate measures used to test for significant correlations with the dependent variables.

The two questions developed to familiarize participants with the concept of DuPont analysis were answered in the same order. The first question asks the participant to calculate what a company's ROE is if its turnover ratios are 2, profitability ratio is 5%

and the leverage ratio is 1.1. The correct answer to the first question is ROE equals 11%. A correct response by the participant was awarded one point, providing a score range of zero to one point (see Table 5). This score is one of the covariate measures used to test for significant correlations with the dependent variables. The time spent in seconds by each participant when answering the first question is also one of the covariate measures used to test for significant correlations with the dependent variables.

The second question asks participants to indicate whether ROE is the sum or multiple of turnover ratio, profitability ratio and leverage ratio. The correct answer to the question is multiple, resulting in the participant receiving one point. The score is either zero or one point (see Table 5). This score on the second question is one of the covariate measures used to test for significant correlations with the dependent variables. The time spent in seconds by each participant when answering the second question is also one of the covariate measures used to test for significant correlations with the dependent variables.

3.5.2 Training for the Pattern Recognition Task

Participants were told to assume that they had a fixed disposable income, and want to rent a 1,200 square foot apartment with the lowest rent and within 5 miles of campus. The degree of desirability of an apartment to a renter is represented by a desirability rating that is a multiplicative function of three variables: income factor, size factor, and distance factor.

Each participant answered four practice questions (see Table 5) by reading from either a tabular display, or a set of 2-D bar displays, or a single 3-D perspective display.

The practice questions were similar to the questions practiced by the participants performing the trend analysis task. For example, instead of describing the values of the factors of the apartment rented in a year, participants performing the pattern recognition task practiced describing the values of the factors of an apartment (see Table 5). Additionally, participants in the 3-D treatment condition learned how to read values from drop lines. Bubbles on the 3-D perspective display represented an apartment's desirability rating, while the X, Y and Z axes represented the income factor, size factor and distance factor, respectively. After completing the training on display format, participants completed a training task on DuPont analysis. The same two practice questions about the concept of DuPont analysis and the calculation of ROE were used to train participants performing the trend analysis task, and the pattern recognition task (see Table 5).

Table 5
Training Questions and Grading Schema for the Trend Analysis Task
(Pattern Recognition Task)

Training Questions Answers to Questions	If Answered Correctly	Grading Schema
What are the values of the factors of (apartment 2) the apartment rented in year 2? Desirability rating 93.78 Size factor 0.94 Income factor 72.03 Distance factor 1.39	 1 point 1 point 1 point 1 point	 0 point to 4 points
What are the differences between (apartments 5 and 6) apartments rented in years 5 and 6? Year (apartment) 5 size factor higher than year (apartment) 6 Year (apartment) 5 income factor higher than year (apartment) 6 Year (apartment) 5 distance factor lower than year (apartment) 6	 1 point 1 point 1 point	 0 point to 3 points
What are the values of the factors of (apartment 4) the apartment rented in year 4? Desirability 26.03 Size factor 0.97 Income factor 71.38 Distance factor 0.38	 1 point 1 point 1 point 1 point	 0 point to 4 points
What are the differences between (apartments 2 and 4) apartments rented in years 2 and 4? Year (apartment) 2 desirability rating higher than year (apartment) 4 Year (apartment) 2 size factor lower than year (apartment) 4 Year (apartment) 2 income factor higher than year (apartment) 4 Year (apartment) 2 distance factor lower than year (apartment) 4	 1 point 1 point 1 point	 0 point to 4 points
What is a company's ROE if its turnover is 2, profitability ratio is 5% and leverage ratio is 1.1? ROE 11%	 1 point	0 point to 1 point
ROE is the sum or multiple of turnover ratio, profitability ratio and leverage ratio? Multiple	1 point	0 point to 1 point

3.6 Measuring the Dependent Variables

3.6.1 Measuring the Dependent Variables for the Trend Analysis Task

To test for the trend analysis task (H1a, H1b, H1c, and H1d), six questions are developed (see Table 3) and each participant answers these six questions in the same order. Questions asked and the scoring of answers is similar to what was used in the training. The first, second, fifth, and sixth questions are used to test for H1a and H1b.

The first question asks the participant to write brief sentences describing the differences between years 1 and 4. The correct answer to the first question has four parts: (a) year 1 ROE is higher than year 4, (b) year 1 turnover is higher than year 4, (c) year 1 profitability is higher than year 4, and (d) year 1 leverage is lower than year 4. Each correct response was awarded one point. The score range for the first question is zero to four points (see Table 6). This score on the first question is the dependent measure (accuracy) used to test H1a (see Table 3). The time spent in seconds by each participant when answering the first question is the dependent measure (efficiency) used to test H1b (see Table 3).

The second question asks the participants to write brief sentences describing what they perceive to be occurring in the data going from year 2 to year 3 and year 4. The correct answer to the second question has eight parts: (a) year 2 ROE is higher than year 3, (b) year 3 ROE is lower than year 4, (c) year 2 turnover is higher than year 3, (d) year 3 turnover is higher than year 4, (e) year 2 profitability is higher than year 3, (f) year 3 profitability is lower than year 4, (g) year 2 leverage is lower than year 3, and (h) year 3 leverage is lower than year 4. Each correct was awarded one point, providing a score range from zero to eight points (see Table 6). This score on the second question is the

second dependent measure (accuracy) used to test H1a (see Table 3). The time spent in seconds by each participant when answering the second question is the second dependent measure (efficiency) used to test H1b (see Table 3).

The fifth question asks participants to indicate the differences between years 2 and 4 by selecting choices from a given template. The correct answer to the fifth question has four parts: (a) year 2 ROE is higher than year 4, (b) year 2 turnover is higher than year 4, (c) year 2 profitability is higher than year 4, and (d) year 2 leverage is lower than year 4. With four parts, the range of possible correct answers is zero to four points (see Table 6). This score on the fifth question is the third dependent measure (accuracy) used to test H1a (see Table 3). The time spent in seconds by each participant when answering the fifth question is the third dependent measure (efficiency) used to test H1b (see Table 3).

The sixth question asks participants to describe what they perceive to be occurring in the data from year 1 to year 2 to year 3 by selecting choices from a given template. The correct answer to the sixth question has eight parts: (a) year 1 ROE is higher than year 2, (b) year 2 ROE is higher than year 3, (c) year 1 turnover is higher than year 2, (d) year 2 turnover is higher than year 3, (e) year 1 profitability is higher than year 2, (f) year 2 profitability is higher than year 3, (g) year 1 leverage is higher than year 2, and (h) year 2 leverage is lower than year 3. Each correct response by the participant to each of the eight parts was awarded one point. The score range for the sixth question is zero to eight points (see Table 6). This score on the sixth question is the fourth dependent measure (accuracy) used to test H1a (see table 3). The time spent in seconds by each participant when

answering the sixth question is the fourth dependent measure (efficiency) used to test H1b (see Table 3).

H1c and H1d are tested with questions three and four. Question three asks the participants to estimate what the ROE would be in year 6 if each of the variables comprising ROE in year 5 had doubled. The correct answer to the third question is 128.34% for year 6 ROE (see Table 6). The difference between each participant's response and the correct answer is the dependent measure (accuracy) used to test H1c (see Table 3). Time spent in seconds by each participant when answering the question is the dependent measure (efficiency) used to test H1d (see Table 3).

Question four asks the participants to estimate the average of turnover, leverage and profitability for the years 1, 2, 4 and 5, and use the estimated average to calculate a new ROE. The correct answer to the fourth question has four parts: a) average turnover = 1.045, (b) average profitability = 3.955%, (c) average leverage = 2.72 and the new ROE = 11.241% (see Table 6). Differences between each participant's responses and the correct answers are the dependent measures (accuracy) used to test H1c (see Table 3). Time spent in seconds by each participant when answering the fourth question is the dependent measure (efficiency) used to test H1d (see Table 3). Table 6 shows the grading schema of each of the six questions of the trend analysis task.

Table 6
Measurement of the Dependent Variables for the Trend Analysis Task

Questions developed to test hypotheses Answers to questions	If answered Correctly	Grading Schema (Score Range)
What are the differences between years 1 and 4? Write short sentences using the words 'higher,' 'lower' Year 1 ROE higher than year 4 Year 1 turnover higher than year 4 Year 1 profitability higher than year 4 Year 1 leverage lower than year 4	 1 point 1 point 1 point 1 point	 0 point to 4 points
What is happening as you go from year 2 to year 3 to year 4? (Hint: Please identify the trends for year 2 to year 4).Write short sentences using the words 'higher,' 'lower' Year 2 ROE higher than year 3 Year 3 ROE lower than year 4 Year 2 turnover higher than year 3 Year 3 turnover higher than year 4 Year 2 profitability higher than year 3 Year 3 profitability lower than year 4 Yea r 2 leverage lower than year 3 Year 3 leverage lower than year 4	 1 point 1 point 1 point 1 point 1 point 1 point 1 point 1 point 1 point	 0 point to 8 points
Based on the ROE of year 5 (Bubble 5), what would be year 6 ROE if each of the variables of ROE in year 5 had doubled? Year 6 ROE 128.34%	+/- differences from answer	
Estimate the average of turnover, leverage and profitability for the years 1, 2, 4 and 5, and use them to calculate a new ROE (Hint 1: you do not actually need to calculate the average, please consider the position of ROE as you attempt to answer. Hint 2: year 3 is not used) Turnover 1.045 Profitability 3.955% Leverage 2.72 New ROE 11.241	 +/- differences from answer	

Table 6 (Continued)
Measurement of the Dependent Variables for the Trend Analysis Task

Questions developed to test hypotheses Answers to questions	If answered Correctly	Grading Schema (Score Range)
What are the differences between years 2 and 4? Please answer as accurately and as fast as possible. Year 2 ROE higher than year 4 Year 2 turnover higher than year 4 Year 2 profitability higher than year 4 Year 2 leverage lower than year 4	 1 point 1 point 1 point 1 point	 0 point to 4 points
What is happening as you go from year 1 to year 2 to year 3? Please answer as accurately and as fast as possible. Year 1 ROE higher than year 2 Year 2 ROE higher than year 3 Year 1 turnover higher than year 2 Year 2 turnover higher than year 3 Year 1 profitability higher than year 2 Year 2 profitability higher than year 3 Year 1 leverage higher than year 2 Year 2 leverage lower than year 3	 1 point 1 point 1 point 1 point 1 point 1 point 1 point 1 point 1 point	 0 point to 8 points

3.6.2 Measuring the Dependent Variables for the Pattern Recognition Task

To test for the pattern recognition task (H2a, H2b, H2c and H2d), six questions are developed (see Table 4) and each participant answers these six questions in the same order. The second and fourth questions are used to test H2a and H2b.

The second question asks the participant to separate companies 1 through 6 into two groups based on similar characteristics. The correct answer to the first question places companies 1, 3, 4, and 6 into group one, and companies 2 and 5 into group two. One point was awarded whenever the participant placed a company into the correct group. The score range for this question is zero to six points (see Table 7). This score on the second question is the dependent measure (accuracy) used to test H2a (see Table 4). Time spent in seconds by each participant when answering the second question is the dependent measure (efficiency) used to test H2b (see Table 4).

Question four asks the participant to select one of the six companies if the goal is to have high profitability, high turnover but low leverage at the same time. Company one is the correct answer to the fourth question. A correct response by the participant was awarded one point. The score range for the fourth question is zero to one point (see Table 7). This score on the fourth question is the second dependent measure (accuracy) used to test H2a (see Table 4). Time spent in seconds by each participant when answering the fourth question is the dependent measure (efficiency) used to test H2b (see Table 4).

The first, third, fifth and sixth questions are used to test for H2c and H2d. The first question asks the participant to write brief sentences describing the differences between companies 1 and 6. The correct answer to the first question has three parts: (a) company 1 turnover is higher than company 6, (b) company 1 profitability is lower than

company 6, and (c) company 1 leverage is lower than year 6. Each correct response by the participant to each of the three parts was awarded one point. The score range for the first question is zero to three points (see Table 7). The score on the first question is the dependent measure (accuracy) used to test H2c (see Table 4). Time spent in seconds by each participant when answering the first question is the dependent measure (efficiency) used to test H2d (see Table 4).

Question three asks the participant to write brief sentences describing the financial ratio patterns they perceive in group one (companies 1, 3, 4 and 6) relative to group two (companies 2 and 5). The correct answer to the third question has three parts: (a) group one turnover is lower than group two, (b) group one profitability is higher than group two, and (c) group one leverage is higher than group two. Each correct response was awarded one point. The score range for the third question is zero to three points (see Table 7). This score on the third question is the second dependent measure (accuracy) used to test H2c (see Table 4). Time spent in seconds by each participant when answering the question is the second dependent measure (efficiency) used to test H2d (see Table 4).

The fifth question asks participants to indicate the differences between companies 4 and 6 by selecting choices from a given template. The correct answer to the fifth question has three parts: (a) company 4 turnover is lower than company 6, (b) company 4 profitability is higher than company 6, and (c) company 4 leverage is higher than company 6. Correct responses were awarded one point, resulting in a score range of zero to three points (see Table 7). The score on the fifth question is the third dependent measure (accuracy) used to test H2c (see Table 4). Time spent in seconds by each

participant when answering the fifth question is the third dependent measure (efficiency) used to test H2d (see Table 4).

Question six asks participants to describe the financial ratio patterns they perceive in group two relative to group one by selecting choices from a given template. The correct answer to the sixth question has three parts: (a) group two turnover is higher than group one, (b) group two profitability is lower than group one, and (c) group two leverage is lower than group one. Again, correct responses by the participant were awarded one point. The score range is zero to three points (see Table 7). The score on the third question is the fourth dependent measure (accuracy) used to test H2c (see Table 4). Time spent in seconds by each participant when answering the sixth question is the fourth dependent measure (efficiency) used to test H2d (see Table 4).

Table 7
Measurement of the Dependent Variables for the Pattern Recognition Task

Questions developed to test hypotheses Answers to questions	If answered Correctly	Grading Schema (Score Range)
What are the differences between companies 1 and 6? Write short sentences using the words 'higher,' 'lower' Company 1 turnover higher than company 6 Company 1 profitability lower than company 6 Company 1 leverage lower than company 6	 1 point 1 point 1 point	 0 point to 3 points
Please separate companies 1 through 6 into 2 groups based on similar characteristics. Note: please assign each company only once to either group one or group two but the groups need not have the same number of companies A group includes companies 1, 3 , 4 and 6 Another group includes companies 2 and 5	 1 point 1 point 1 point 1 point 1 point 1 point	 0 point to 6 points
Group one includes companies 1, 3, 4, and 6, and group two includes companies 2 and 5. Compared to group two, what are the patterns of the financial ratios you are seeing in group one? Write short sentences using the words 'higher,' 'lower' Group 1 turnover lower than group 2 Group 1 profitability higher than group 2 Group 1 leverage higher than group 2	 1 point 1 point 1 point	 0 point to 3 points
Assume you cannot select a company solely because of a single variable, for example higher profitability. Comparatively, if it is better to have a higher profitability, higher turnover but lower leverage at the same time, which company you will select? Company 1	 1 point	 0 point to 1 point

Table 7 (Continued)
Measurement of the Dependent Variables for the Pattern Recognition Task

Questions developed to test hypotheses Answers to questions	If answered Correctly	Grading Schema (Score Range)
What are the differences between companies 4 and 6? Please answer as accurately and as fast as possible Company 4 turnover lower than company 6 Company 4 profitability higher than company 6 Company 4 leverage higher than company 6	 1 point 1 point 1 point	 0 point to 3 points
Group one includes companies 1, 3, 4 and 6, and group two includes companies 2 and 5. Compared to group one what are the patterns of the financial ratios you are seeing in group two? Please answer as accurately and as fast as possible. Group 2 turnover higher than group 1 Group 2 profitability lower than group 2 Group 2 leverage lower than group 1	 1 point 1 point 1 point	 0 point to 3 points

3.7 Covariates

3.7.1 Practice Questions

Tufte (1983 p. 56) suggested that perceptions change with experience; and that perceptions are context dependent. An individual's performance on an experimental task is affected by his or her knowledge of the DuPont analysis and his or her familiarity with the (randomly assigned) display format. Compared to the participants viewing the 3-D display, it is expected that participants viewing the tabular display or the 2-D displays would have higher scores in each of the four practice questions regarding the display format; as the latter would be more familiar with their assigned display formats. The score on each of the six practice questions (four questions on display format and two questions on ROE) is the covariate measure used to test for significant correlations with the dependent variables.

The time spent in seconds by each participant when answering each of the six practice questions is another covariate measure used to test for significant correlations with the dependent variables.

3.7.2 Mental Rotations Test

An individual's performance on an experimental task is affected by his or her spatial visualization ability. Using meta-analytic techniques to compare the effect size of the gender difference in fourteen prior studies that administrated the Mental Rotations Test, Masters and Sanders (1993) confirmed that males generally scored higher than females in the Mental Rotations Tests.

Vandenberg and Crawford (1971) developed the Mental Rotations Test to measure individual's spatial visualization abilities. This test uses three-dimensional objects displayed in two-dimensional drawings to measure spatial visualization. In order to control for differences between individuals' spatial visualization abilities, participants in the study completed the Mental Rotations Test.

This study uses the Mental Rotations Test supplied by the Educational Testing Service (ETS). The Mental Rotations Test presents a three-dimensional drawing on the left hand side and requires participants to indicate which two of the four "rotated" (rotated in three-dimensional space) versions of the drawing on the right hand side (rotated in three-dimensional space) represents the original 3-D drawing, thereby measuring participants' spatial visualization ability (Vandenberg and Crawford 1971).

Participants' responses to the Mental Rotations Test were graded according to the following grading schema: (a) if two drawings were chosen and both choices of drawings are correct, two points are awarded, (b) if two drawings were chosen and one choice of drawings is incorrect, or both choice of drawings are incorrect no points are awarded, (c) if only one drawing was chosen and it is correct, one point is awarded. The score range for the Mental Rotations Test is zero to 40 points. This score on the Mental Rotations Test is the covariate measure used to test for significant correlations with the dependent variables. The total time in seconds spent by each participant when answering the Mental Rotations Test is also used as a covariate measure used to test for significant correlations with the dependent variables.

3.7.3 Gender and Age

Hygge and Knez (2001) used gender as an individual variable to test for cognitive performance on four different tasks−attention, problem solving, long-term recall and recognition, and short-term recall−under varying conditions of heat and light. Their results showed that women demonstrated higher performance than men on problem solving tasks, and women also remembered a greater number of words than men. Within the auditing context, Chung and Monroe (1998) confirmed prior findings that females and males process information differently. Accordingly, gender will be one of the covariates in the study (see Table 8). Age will also be collected and tested for significant correlations with the dependent variables (see Table 8).

3.7.4 Mental Workload

Benford (2000) defines mental workload as an individual's assessment of the difference between their capacity for performance and the demands of the task. Benford (2000) further explained that an individual will perceive increases in his/her mental workload as task demands increase.

In this study, participants in the 2-D and 3-D treatment conditions had to scroll up and down the screen when answering questions, since the display and the response area did not fit on one screen. Participants viewing the tabular display did not have such a problem. It is expected that participants viewing the 2-D displays and the 3-D perspective display will perceive heavier mental workloads than participants viewing the tabular display.

Participants of this study were undergraduate business students who are familiar with reading data from a tabular display, 2-D line graphs, and 2-D bar charts. The 3-D perspective display of DuPont analysis is a newly created presentation format of financial ratios, and had never been seen by the participants. Therefore, it is logical to expect that participants viewing the 3-D perspective display would perceive the highest mental workload (due to unfamiliarity and the need to scroll up and down) when compared to the participants viewing the 2-D displays or the participants viewing the tabular display. Participants viewing the 2-D display could perceive a higher mental workload than the participants viewing the tabular display (due to the need to scroll up and down).

This study adopted the four statements developed by Reid and Nygren (1988) to measure a participant's perceived workload. Participants were asked to select the scale number that indicates the extent to which they agree with each of the four statements. The scale numbers are: 1= strongly disagree, 2 = moderately disagree, 3 = slightly disagree, 4 = neither agree nor disagree, 5 = slightly agree, 6 = moderately agree, and 7= strongly agree (see Table 9).

The four statements are: (1) very little mental effort or concentration was required to complete the tasks, (2) the tasks performed were almost automatic, requiring little or no attention, (3) the tasks were very complex and required total attention, and (4) extensive mental effort and concentration was necessary in the tasks. Reid and Nygren (1988) did not report the Cronbach's Alpha on these four statements, however, Benford (2000) found a high Cronbach's Alpha of 0.82.

The score on mental workload is the covariate measure used to test for significant correlations with the dependent variables.

3.7.5 Demographic Data

In addition to the variables already discussed, participants were asked to provide information about their student status, undergraduate GPA, SAT score, number of years of full time working experience, part time working hours per week (if not working full time), working experience in accounting related jobs, and highest level of education (see Table 8). The data on these additional demographic questions will be tested as possible covariate measures.

Table 8
Demographic Questions

Your age is 18-22 23-27 29-32 33-37 38-42 43-47 48-50
Your Gender is Male Female
Student Status Freshman Sophomore Junior Senior Graduate Student
What was your undergraduate overall GPA?
What was your SAT score?
Number of years of full time work experience?
If you do not work full time now, number of hours you work part time now?
Work experience in accounting related jobs (for example: bookkeeping or auditing)? Full time Part time Both full time and part time None
Highest level of education you already achieved? High school Bachelor's degree Master's degree

3.8 Post Hoc Analysis (Survey Questions)

As indicated earlier, the 3-D perspective display of DuPont analysis is a newly created display format of financial ratios that has never been empirically tested. For this reason, the study adopted six survey questions from Fuller, Murthy, and Schafer (2007) to elicit participants' opinions on the usefulness and ease of use of their randomly assigned display format.

Participants were asked to select the scale number from 1 to 7 that indicates the extent to which they agree with each of the six questions. The six questions are: (1) using the tables (graphs) was frustrating, (2) the tables (graphs) displayed the task information in a readable format, (3) I found the tables (graphs) useful in how they presented the data for decision making, (4) the tables (graphs) helped me to understand the task data to make a better decision, (5) the tables (graphs) fit the way I needed to view the task information to make better decisions, and (6) overall, I am satisfied with the tables (graphs) in providing the information I needed to complete this task (see Table 9).

3.9 Student Participants

Student participants are used in the study for several reasons. DuPont analysis is taught in every principles of accounting class. Wright (2007) demonstrated that students can be good surrogates for real world auditors, as task-specific academic instruction and practice is a good substitute for audit working experience, at least for a task such as evaluating loan collectability. Participants in this study learn and practice DuPont analysis in the classroom, while participants in Wright (2007) also learned about the loan collectability judgment in the classroom. In discussing the use of students as participants

in behavioral auditing research, Peecher and Solomon (2001) contend that it is inappropriate only when theory or prior research indicates that experience interacts with a factor of interest in the study. Given the paucity of research on the effects of multidimensional displays in accounting behavioral research and the lack of widespread use of such displays in practice, it is unknown whether experience would interact with display type. Accordingly, the use of students as participants is deemed appropriate at least in the initial phases of this research stream.

Table 9
Mental Workload Questions and Survey Questions

Please use the following scale as the index for your responses.						
1	2	3	4	5	6	7
strongly disagree	moderately disagree	slightly disagree	Neither agree nor disagree	slightly agree	moderately agree	strongly agree
Please select the scale number which indicates the extent to which you agree with each of the following statements. There is no right or wrong answer to these statements						
Using the tables (graphs) was frustrating.						
Very little mental effort or concentration was required to complete tasks.						
Tasks performed were almost automatic, requiring little or no attention.						
Tasks were very complex and required total attention.						
Extensive mental effort and concentration was necessary in tasks.						
The tables (graphs) displayed the task information in a readable format.						
I found the tables (graphs) useful in how they presented the data for decision making.						
The tables (graphs) helped me understand the task data to make a better decision.						
The tables (graphs) fit the way I needed to view the task information to make a better decision.						
Overall, I am satisfied with the tables (graphs) in providing the information I needed to complete the task.						

3.10 Manipulation Checks

The objective of the manipulation questions is to test whether participants' responses were in the expected directions in accordance with the participants' treatment conditions. Different manipulation questions were designed for the treatment conditions of display format and task type. Another manipulation question was designed to test participants' post experiment knowledge of DuPont analysis. Participants were asked to answer three manipulation questions by selecting the choices of true or false.

To test the manipulation of display formats, participants viewing the tabular display were asked whether: "The tables that you see in this experiment have data points of zero (answer is false)."Participants viewing the 2-D line display were asked whether: "The line graphs that you see in this experiment have their axis started from zero (answer is false)." If participants viewed the 2-D bar display they were asked whether: "The bar charts that you see in this experiment have their axis started from zero (answer is true)." Finally participants viewing the 3-D display were asked whether: "The graphs that you see in this experiment have their axis started from zero (answer is false)" (See Table 10).

To test the manipulation of tasks, participants performing the trend analysis task were asked to respond to the statement: "Within the context of this experiment companies can have negative ROE (answer is true)." Participants performing the pattern analysis task were asked to respond to the statement: "Within the context of this experiment, companies with the same ROE also have the same turnover, profitability, and leverage ratios (answer is false)" (See Table 10). To test participants' post experiment knowledge of DuPont analysis, all participants were asked whether: "Return on equity is the sum of turnover, profitability, and leverage ratios (answer is false)" (See Table 10).

Lastly, due to the fact that the study has many variables, it is beneficial to have a table summarizing the definition and symbols of all the variables to be used in statistical analysis (see Table 11).

Table 10
Manipulation Check Questions

Tabular Display	
Trend Analysis Task	**Pattern Recognition Task**
The tables you see in the experiment, have data points of zero	The tables you see in the experiment, have data points of zero
Return on Equity is the SUM of turnover, profitability, and leverage ratios	Return on Equity is the SUM of turnover, profitability, and leverage ratios
Within the context of this experiment companies can have negative ROE	Within the context of this experiment, companies with the same ROE also have the same turnover, profitability, and leverage ratios.
2-D Line (Bar) Display	
Trend Analysis Task	**Pattern Recognition Task**
The line graphs you see in this experiment, have their axis started from zero	The bar charts you see in this experiment, have their axis started from zero
Return on Equity is the SUM of turnover, profitability, and leverage ratios	Return on Equity is the SUM of turnover, profitability, and leverage ratios
Within the context of this experiment companies can have negative ROE	Within the context of this experiment, companies with the same ROE also have the same turnover, profitability, and leverage ratios.
3-D Display	
Trend Analysis Task	**Pattern Recognition Task**
The graphs you see in this experiment, have their axis started from zero	The graphs you see in this experiment, have their axis started from zero
Return on Equity is the SUM of turnover, profitability, and leverage ratios	Return on Equity is the SUM of turnover, profitability, and leverage ratios
Within the context of this experiment companies can have negative ROE	Within the context of this experiment, companies with the same ROE also have the same turnover, profitability, and leverage ratios.

Table 11
Definition of Variables Used in Statistical Analysis

Variable	Variable Symbol	Definition	Refer To
Trend Analysis Task Question one and the Time Spent on Task Question one	TAQ1 TSTAQ1	Scores and time on Trend Analysis Task Question one, "What are the differences between years 1 and 4?"	Table 3 and 6
Trend Analysis Task Question two and the Time Spent on Task Question two	TAQ2 TSTAQ2	Scores and time on Trend Analysis Task Question two, "What is happening as you go from year 2 to year 3 to year 4?"	Table 3 and 6
Trend Analysis Task Question three and the Time Spent on Task Question three	TAQ3 TSTAQ3	Scores and time on Trend Analysis Task Question three, "Based on the ROE of year 5 (Bubble 5), what would be year 6 ROE if each of the variables of ROE in year 5 had doubled?"	Table 3 and 6
Trend Analysis Task Question four part a	TAQ4a	Scores on Trend Analysis Task Question four part a, "Estimate the average of turnover, for the years 1, 2, 4 and 5."	Table 3 and 6
Trend Analysis Task Question four part b	TAQ4b	Scores on Trend Analysis Task Question four part b, "Estimate the average of leverage for the years 1, 2, 4 and 5."	Table 3 and 6
Trend Analysis Task Question four part c	TAQ4c	Scores on Trend Analysis Task Question four part c, "Estimate the average of profitability for the years 1, 2, 4 and 5."	Table 3 and 6
Trend Analysis Task Question four part d	TAQ4d	Scores on Trend Analysis Task Question four part d, "Estimate the average of turnover, leverage and profitability for the years 1, 2, 4 and 5, and use them to calculate a new ROE."	Table 3 and 6
Time Spent on Task Question four part a, b, c , and d.	TSTAQ4	The time spent on answering Task Question four, parts a, b, c and d.	Table 3 and 6
Trend Analysis Task Question five and the Time Spent on Task Question five	TAQ5 TSTAQ5	Scores and time on Trend Analysis Task Question five, "What are the differences between years 2 and 4?"	Table 3 and 6
Trend Analysis Task Question six and the Time Spent on Task Question six	TAQ6 TSTAQ6	Scores and time on Trend Analysis Task Question six, "What is happening as you go from year 1 to year 2 to year 3?"	Table 3 and 6

Table 11 (Continued)
Definition of Variables Used in Statistical Analysis

Variable	Variable Symbol	Definition	Refer To
Pattern Recognition Task Question one and the Time Spent on Task Question one	PRQ1 TSPRQ1	Scores and time on Pattern Recognition Task Question one, "What are the differences between companies 1 and 6?"	Table 4 and 7
Pattern Recognition Task Question two and the Time Spent on Task Question two	PRQ2 TSPRQ2	Scores and time on Pattern Recognition Task Question two, "Please separate companies 1 through 6 into 2 groups based on similar characteristics."	Table 4 and 7
Pattern Recognition Task Question three and the Time Spent on Task Question three	PRQ3 TSPRQ3	Scores and time on Pattern Recognition Task Question three, "Group one includes companies 1, 3, 4, and 6, and group two includes companies 2 and 5. Compared to group two, what are the patterns of the financial ratios you are seeing in group one?"	Table 4 and 7
Pattern Recognition Task Question four and the Time Spent on Task Question four	PRQ4 TSPRQ4	Scores and time on Pattern Recognition Task Question four, "Assume you cannot select a company solely because of a single variable, for example higher profitability. Comparatively, if it is better to have higher profitability, higher turnover but lower leverage at the same time, which company you will select?"	Table 4 and 7
Pattern Recognition Task Question five & the Time Spent on Task Question five	PRQ5 TSPRQ5	Scores and time on Pattern Recognition Task Question five, "What are the differences between companies 4 and 6?"	Table 4 and 7

Table 11 (Continued)
Definition of Variables Used in Statistical Analysis

Variable	Variable Symbol	Definition	Refer To
Pattern Recognition Task Question six and the Time Spent on Task Question six	PRQ6 TSPRQ6	Scores and time on Pattern Recognition Task Question six, "Group one includes companies 1, 3, 4 and 6, and group two includes companies 2 and 5. Compared to group one what are the patterns of the financial ratios you are seeing in group two?"	Table 4 and 7
Practice Question one and the Time Spent on Practice Question one	PQ1 TSPQ1	Scores and time on Practice Question one, "What are the values of the factors of (apartment 2) the apartment rented in year 2?"	Table 5
Practice Question two and the Time Spent on Practice Question two	PQ2 TSPQ2	Scores and time on Practice Question two, "What are the differences between (apartments 5 and 6) apartments rented in years 5 and 6?"	Table 5
Practice Question three and the Time Spent on Practice Question three	PQ3 TSPQ3	Scores and time on Practice Question three, "What are the values of the factors of (apartment 4) the apartment rented in year 4?"	Table 5
Practice Question four and the Time Spent on Practice Question four	PQ4 TSPQ4	Scores and time on Practice Question four, "What are the differences between (apartments 2 and 4) apartments rented in years 2 and 4?"	Table 5
Practice Question five and the Time Spent on Practice Question five	PQ5 TSPQ5	Scores and time on Practice Question five, "What is a company's ROE if its turnover is 2, profitability ratio is 5% and leverage ratio is 1.1?"	Table 5

Table 11 (Continued)
Definition of Variables Used in Statistical Analysis

Variable	Variable Symbol	Definition	Refer To
Practice Question six and the Time Spent on Practice Question six	PQ6 TSPQ6	Scores and time on Practice Question six, "ROE is the sum or multiple of turnover ratio, profitability ratio and leverage ratio."	Table 5
Score on Mental Rotations Test	SMRT	Score on Mental Rotations Test, which measures an individual's spatial ability	Section 3.6.2
Time Spent on Mental Rotations Test	TSMRT	Time in seconds spent on the Mental Rotations Test	Section 3.6.2
Age	AGE	Participant's age (1 = 18-22, 2 = 23-27, 3 = 28-32, 4 = 33-37, 5 = 38-42, 6 = 43-47, 7 = 48-50).	Table 8
Gender	GEN	Participant's gender(1= Male, 2 = Female)	Table 8
Student Status	SS	Participant's enrollment status in the university (1= Freshman, 2 = Sophomore, 3 = Junior, 4 = Senior, 5= Graduate Student)	Table 8
GPA	GPA	Participant's undergraduate GPA	Table 8
SAT	SAT	Participant's SAT score	Table 8
Full Time Working Experience	FTWE	Number of years of full time working experience	Table 8
Part Time Working Hours	PTH	Number of part time working hours per week	Table 8
Accounting Related Working Experience	ARWE	Accounting related working experience (1= full time, 2= part time 3= full time and part time, 4=none)	Table 8

Table 11 (Continued)
Definition of Variables Used in Statistical Analysis

Variable	Variable Symbol	Definition	Refer To
Highest Level of Education	HE	Participant's highest level of education already achieved (1= High School, 2 = Bachelor's degree, 3 = Master's degree).	Table 8
Survey Question one	S1	Score on the Survey Question one, "Using the tables (graphs) was frustrating."	Table 9
Survey Question two	S2	Score on the Survey Question two, "Very little mental effort or concentration was required to complete tasks."	Table 9
Survey Question three	S3	Score on the Survey Question three, "Tasks performed were almost automatic, requiring little or no attention."	Table 9
Survey Question four	S4	Score on the Survey Question four, "Tasks were very complex and required total attention."	Table 9
Survey Question five	S5	Score on the Survey Question five, "Extensive mental effort and concentration was necessary on Tasks."	Table 9
Survey Question six	S6	Score on the Survey Question six, "The tables (graphs) displayed the task information in a readable format."	Table 9
Survey Question seven	S7	Score on the Survey Question Seven, "I found the tables (graphs) useful in how they presented the data for decision making."	Table 9

Table 11 (Continued)
Definition of Variables Used in Statistical Analysis

Variable	Variable Symbol	Definition	Refer To
Survey Question eight	S8	Score on the Survey Question eight, "The tables (graphs) helped me understand the task data to make a better decision."	Table 9
Survey Question nine	S9	Score on the Survey Question nine, "The tables (graphs) fit the way I needed to view the task information to make a better decision."	Table 9
Survey Question ten	S10	Score on the Survey Question ten, "Overall, I am satisfied with the tables (graphs) in providing the information I needed to complete the task."	Table 9
Mental Workload	MW	Average score on the Survey Question two, three, four and five. (reverse coding on S2 and S3)	Table 9
Manipulation Question one	M1	The tables (graphs) you see in the experiment, have data points of zero	Table 10
Manipulation Question two	M2	Return on Equity is the SUM of turnover, profitability, and leverage ratios	Table 10
Manipulation Question three	M3	Within the context of this experiment, companies can have negative ROE or within the context of this experiment, companies with the same ROE also have the same turnover, profitability, and leverage ratios.	Table 10

Chapter 4: Pilot Study

4.1 Research Design of the Pilot Study

The research design for the pilot study is as is described in Chapter 3, with the following exceptions. For the pilot study, task type is manipulated within-subjects rather than between-subjects. The order of tasks is balanced such that half the participants in each treatment group perform the trend analysis task first and the pattern recognition task second, and vice versa.

A total of eighty-six undergraduate business school students participated in a pilot study designed to test the experimental materials and determine whether the experimental manipulations had the intended effect. Each participant was randomly assigned to one of the three treatment conditions.

Each participant answered two practice questions by reading from either a table, or a set of 2-D bar charts, or a single 3-D perspective display. After participants answered the two practice questions assessing apartment desirability, DuPont analysis was explained to them using the return on equity of Motorola and Nokia in 1997 as examples. These examples were provided to refresh participants' memories regarding DuPont analysis, as it was expected that participants had already learned the concept of DuPont analysis in the principles of accounting class. Neither a practice question on the calculation of the return on equity nor a practice question about the concept of the DuPont analysis had been provided to the participants. Participants in the 3-D treatment

group viewed an additional page of training demonstrating that bubbles can also be used to display how data changes over time (i.e., from year to year).

4.2 Results of the Pilot Study

4.2.1 Descriptive Statistics, Test of Assumptions and Outliers

In terms of effectiveness or accuracy, for both the practice questions one and two, participants viewing the tabular display scored the highest, then followed by those participants viewing the 2-D displays, while those participants viewing the 3-D perspective display scored the lowest (see Table 12).

In term of effectiveness or accuracy, participants viewing the 3-D perspective display scored the lowest in all the six questions of the trend analysis task. Participants viewing the tabular display scored the highest in the first question (which asked what the data differences were between year 1 and 4), the third question (which asked what the ROE would be in year 6 if each of the variables comprising ROE in year 5 doubled), and the fourth question part 'd' (which asked participants to estimate a new ROE using the average of turnover, profitability, and leverage for the years 1, 2, 4 and 5) of the trend analysis task. While those participants viewing the 2-D display scored the highest in the second question (which asked what was occurring in the data when going from year 2 to year 3 and year 4), the fourth question part 'a' (which asked participants to estimate the average turnover for the years 1, 2, 4, and 5), the fourth question part 'b' (which asked participants to estimate the average profitability for the years 1, 2, 4, and 5), the fourth question part 'c' (which asked participants to estimate the average leverage for the years

1, 2, 4, and 5), and the sixth question (which asked what was occurring in the data when going from year 1 to 2 to 3) of the trend analysis task (see Table 12).

In terms of efficiency or time used (in seconds) by each participant when answering each of the question of the trend analysis task, those participants viewing the 2-D displays were the most efficient or used the least amount of time in answering questions, while those participants viewing the 3-D perspective display used the most amount of time in answering questions (see Table 12).

For the pattern recognition task, in term of effectiveness or accuracy, participants viewing the 3-D perspective display scored the highest in the second question (which asked participants to separate companies one through six into two groups), the third question (which participants to describe the pattern of financial ratios in group one), and the fourth question of the (which participants to select one of the six companies if the goal is to have high profitability, high turnover, but low leverage at the same time), while those participants viewing the tabular display scored the lowest in the mentioned above three questions. Those participants viewing the 2-D displays scored the highest in the fifth question (which asked what were the data differences between companies 4 and 6), and question six (which asked participants to describe the pattern of financial ratios in group two) of the pattern recognition task, while those participants viewing the 3-D perspective display scored the lowest in the mentioned above two questions. Finally, those participants viewing the tabular display scored the highest in the first question of the pattern recognition task (which asked what data difference were between companies 1 and 6) with those participants viewing the 2-D displays scored the lowest in this question (see Table 12).

In terms of efficiency or time used (in seconds) by each participant when answering each of the question of the pattern recognition task, those participants viewing the 2-D displays were the most efficient or used the least amount of time in answering five out of the six questions of the pattern recognition task. While those participants viewing the tabular display or those participants viewing the 3-D perspective display used the most amount of time in answering three out of the six questions of the pattern recognition task, respectively (see Table 12).

The scores on the Mental Rotations Test between tabular, 2-D and 3-D treatment groups were 21, 22, and 23 points, respectively. T-testing shows that there is no significant difference in spatial ability between treatment groups. Participants of each of the three treatment groups had similar GPA and scores of SAT. Participants of the tabular display treatment group had the longest full time working experience, while those participants of the 2-D displays treatment group had the shortest. Participants of the 2-D displays treatment group had longest part time working hours while those participants of the tabular display treatment group had the shortest (see Table 12).

Two blind coders coded the responses of each participant. Descriptive statistics (see Table 12) reveal that there is considerable variability in the data. Tests of influential observations and outliers, at two standard deviations, were performed on both the coded responses and response time. The time spent by a participant in the 3-D treatment condition in selecting a choice from a given template, which best describes the patterns of the financial ratios from year 1 to year 2 to year 3 had a Cook's distance greater than one. The aforementioned data point was subsequently dropped.

Tests of normality and homogeneity of variance were also performed on both the coded responses and response time. The coded responses and response time were not normally distributed. Since univariate ANCOVA analysis is robust to data that is not normally distributed no adjustments were made to the data. Multivariate MANCOVA was used first to test whether the manipulation of the presentation formats had significance results. If the manipulation of the presentation formats had significant result, ANCOVA was then used to analysis the univariate results.

Table 12
Pilot Study Descriptive Statistics

Panel A: Mean (Standard Deviation) and Range of Practice Questions.

	Tabular Display (n=27)	**2-D Displays (n=30)**	**3-D Display (n=29)**
Practice question one: What are the values of the factors of apartment one?	3.222 (1.086) 0.000 to 4.000	3.000 (1.389) 0.000 to 4.000	2.966 (0.680) 0.000 to 4.000
Practice question two: What are the differences between apartments five and six?	3.444 (0.891) 0.000 to 4.000	3.400 (0.932) 0.000 to 4.000	3.034 (0.778) 1.000 to 4.000

Panel B: Mean (Standard Deviation) and Range of Trend Analysis Task.

	Tabular Display (n=27)	**2-D Displays (n=30)**	**3-D Display (n=29)**
Question one What are the differences between year 1 and year 4?	3.481 (0.752) 1.000 to 4.000	3.400 (0.968) 0.000 to 4.000	2.793 (0.726) 1.000 to 4.000
Time (seconds) spent on question one	132 (74) 53 to 331	127 (71) 39 to 390	148 (77) 49 to 318
Question two What is happening as you go from year 2 to year 3 to year 4?	5.777 (2.241) 1.000 to 8.000	5.866 (2.849) 0.000 to 8.000	3.862 (2.199) 0.000 to 8.000
Time (seconds) spent on question two	137 (91) 35 to 415	129 (61) 36 to 258	155 (128) 26 to 674

Table 12 (Continued)
Pilot Study Descriptive Statistics

Panel B: Mean (Standard Deviation) and Range of Trend Analysis Task.

	Tabular Display (n=27)	**2-D Displays (n=30)**	**3-D Display (n=29)**
Question three Based on the ROE of year 5, what would be year 6 ROE if each of the variables of ROE in year 5 had doubled?	4.148 (2.918) 0.000 to 10.000	3.200 (1.864) 1.000 to 10.000	3.4483 (2.30816) 0.000 to 10.000
Time (seconds) spent on question three	87 (48) 20 to 250	88 (52) 27 to 193	86 (55) 33 to 253
Question four part a Estimate the average of turnover for the years 1, 2, 4, and 5	6.814 (4.582) 0.000 to 12.000	9.100 (3.467) 0.000 to 12.000	8.413 (4.939) 0.000 to 12.000
Question four part b Estimate the average of leverage for the years 1, 2, 4, and 5	3.481 (4.846) 0.000 to 12.000	3.500 (4,761) 0.000 to 10.000	2.551 (4.264) 0.000 to 10.000
Question four part c Estimate the average of profitability for the years 1, 2, 4, and 5	3.111 (4.870) 0.000 to 12.000	5.400 (3.891) 0.000 to 11.000	3.448 3.869 0.000 to 11.000
Question four part d Estimate the average of turnover, leverage and profitability for the years 1, 2, 4, and 5, and use them to calculate a new ROE	1.111 (2.750) 0.000 to 9.000	0.533 (2.029) 0.000 to 8.000	1.000 2.604 0.000 to 9.000
Time (seconds) spent on question four	205 (123) 29 to 595	172 (78) 49 to 364	836 (149) 47 to 836

Table 12 (Continued)
Pilot Study Descriptive Statistics

Panel B: Mean (Standard Deviation) and Range of Trend Analysis Task.

	Tabular Display (n=27)	**2-D Displays (n=30)**	**3-D Display (n=29)**
Question five What are the differences between years 2 and 4? Please answer as accurately and as fast as possible.	3.407 (0.747) 2.000 to 4.000	3.866 (0.434) 2.000 to 4.000	3.620 (0.621) 2.000 to 4.000
Time (seconds) spent on question five	34 (13) 21 to 79	32 (9) 19 to 54	50 (13) 27 to 86
Question six What is happening as you go from year 1 to year 2 to year 3? Please answer as accurately and as fast as possible.	7.222 (1.368) 3.000 to 8.000	7.666 (0.660) 5.000 to 8.000	6.931 (1.412) 3.000 to 8.000
Time (seconds) spent on question six	69 (27) 31 to 164	54 (17) 29 to 101	100 (38) 15 to 205

Panel C: Mean (Standard Deviation) and Range of Pattern Recognition Task.

	Tabular Display (n=27)	**2-D Displays (n=30)**	**3-D Display (n=29)**
Question One What are the differences between companies 1 and 6?	2.777 (0.640) 0.000 to 3.000	2.566 (0.897) 0.000 to 3.000	2.689 (0.603) 1.000 to 3.000

Table 12 (Continued)
Pilot Study Descriptive Statistics

Panel C: Mean (Standard Deviation) and Range of Pattern Recognition Task.

	Tabular Display (n=27)	**2-D Displays (n=30)**	**3-D Display (n=29)**
Time (seconds) spent on question one	112 (58) 46 to 317	111 (45) 42 to 192	163 (82) 71 to 365
Question two Please separate companies 1 through 6 into 2 groups based on similar characteristics.	4.555 (0.697) 3.000 to 6.000	4.800 (0.996) 3.000 to 6.000	5.069 (0.923) 3.000 to 6.000
Time (seconds) spent on question two	91 (52) 38 to 232	70 (30) 26 to 183	83 (32) 37 to 164
Question three Group one includes companies 1, 3, 4, and 6, and group two includes companies 2 and 5. Compared to group two, what are the patterns of the financial ratios you are seeing in group one?	1.777 (1.012) 0.000 to 3.000	2.333 (0.802) 1.000 to 3.000	2.517 (0.828) 0.000 to 3.000
Time (seconds) spent on question three	121 (46) 26 to 251	97 (44) 37 to 213	105 (42) 54 to 243

Table 12 (Continued)
Pilot Study Descriptive Statistics

Panel C: Mean (Standard Deviation) and Range of Pattern Recognition Task.

	Tabular Display (n=27)	**2-D Displays (n=30)**	**3-D Display (n=29)**
Question four Assuming you cannot select a company solely because of a single variable, for example higher profitability. Comparatively, if it is better to have a higher profitability, higher turnover, but lower leverage at the same time, which company will you select?	0.000 (0.000) 0.000 to 0.000	0.200 (0.406) 0.000 to 1.000	0.724 (0.454) 0.000 to 1.000
Time (seconds) spent on question four	67 (22) 20 to 113	65 (28) 28 to 127	55 (20) 21 to 117
Question five What are the differences between companies 4 and 6? Please answer as accurately and as fast as possible.	2.851 (0.362) 2.000 to 3.000	3.00 (0.000 3.000 to 3.000	2.689 (0.603) 1.000 to 3.000
Time (seconds) spent on question five	30 (11) 16 to 65	27 (9) 15 to 57	43 (19) 21 to 122

Table 12 (Continued)
Pilot Study Descriptive Statistics

Panel C: Mean (Standard Deviation) and Range of Pattern Recognition Task.

	Tabular Display (n=27)	**2-D Displays (n=30)**	**3-D Display (n=29)**
Question six Group one includes companies 1, 3, 4, and 6, and group two includes companies 2 and 5. Compared to group one, what are the patterns of the financial ratios you are seeing in group two? Please answer as accurately and as fast as possible.	2.703 (0.724) 0.000 to 3.000	2.833 (0.592) 0.000 to 3.000	2.655 (0.768) 0.000 to 3.000
Time (seconds) spent on question six	46 (20) 20 to 101	35 (12) 19 to 68	55 (32) 22 to 168

Panel D: Mean (Standard Deviation) and Range of Mental Rotations Test.

	Tabular Display (n=27)	**2-D Displays (n=30)**	**3-D Display (n=29)**
Score on Mental Rotations Test	21.592 (10.123) 2.000 to 36.000	22.366 (12.397) 2.000 to 40.000	23.689 (9.849) 4.000 to 0.000
Time (seconds) spent on Mental Rotation Test	820 (365) 333 to 2031	657 (201) 195 to 1197	771 (263) 306 to 1222

Table 12 (Continued)
Pilot Study Descriptive Statistics

Panel E: Demographic Mean and Range.

	Tabular Display (n=27)	**2-D Displays (n=30)**	**3-D Display (n=29)**
Male/Female	n= 9/n = 18	n = 14/n= 16	n = 8/ n = 21
GPA	3.2048 2.270-4.000	3.2953 2.50-3.96	3.2583 2.00-3.87
SAT Score	1194 980-1750 (n=23)	1158 900-1470 (n=29)	1149 680-1800 (n=28)
Full Time Working Experience in Years	3.7778 0.000-20.000	2.2167 0.00-28.00	2.8793 0.00-27.00
Part Time Working Hours Per Week	15.4815 0.000-40.000	18.667 0.000-52.000	16.5862 0.000-40.000

4.2.2 Hypotheses Testing of the Trend Analysis Task

Results of hypotheses H1a, H1b, H1c and H1d are shown in Tables 13 to 21. Multivariate MANCOVA was first used to test for the significance of the manipulation variable Treatment. If the manipulation variable Treatment was significant, ANCOVA was then conducted to analysis the univariate results. Covariates included in the analysis are the scores on practice questions one and two, gender, age, score on the Mental Rotations Test, time spent on the Mental Rotations Test, and the task order performed by each participant (for example the trend analysis task first and the pattern recognition task second, and vice versa).

The following paragraphs report significant results with tables on top of narration.

4.2.2.1 Results of Hypothesis H1a. Hypothesis H1a has four dependent measures of accuracy − the score on the first question, which asked what the data differences were between years 1 and 4; the score on the second question, which asked what participants perceived to be occurring in the data when going from year 2 to year 3 and year 4; the score on the fifth question, which asked participants to select from a template to indicate the differences in data between year 2 and 4; and the score on the sixth question, which asked participants to select from a template to indicate changes in data when going from year 1 to year 2 to year 3.

The mean results for all dependent variables are hypothesized in H1a. In constructing the models to test H1a, all covariates mentioned in section 4.2.2 were included in the model along with the manipulated variable Treatment.

Prior to presenting ANCOVA results for the four dependent variables used to test hypothesis H1a, a MANCOVA analysis was conducted. As shown (Table 13), the overall F-statistic for the manipulated variable Treatment is significant (p = 0.003) using Pillai's Trace. The results are also significant (p = 0.003) using Hotelling's Trace. These significant results allow for analysis of the univariate results which are provided on Table 14 to 16.

Table 13
Multivariate Tests of H1a.

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.566	23.173	<0.001
PQ1	Pillai's Trace	0.025	0.450	0.772
PQ2	Pillai's Trace	0.343	9.249	<0.001
AGE	Pillai's Trace	0.097	1.907	0.119
GEN	Pillai's Trace	0.087	1.683	0.164
SMRT	Pillai's Trace	0.072	1.373	0.252
TSMRT	Pillai's Trace	0.025	0.460	0.765
Treatment	Pillai's Trace	0.296	3.129	0.003
Rank	Pillai's Trace	0.032	0.592	0.667
Treatment*Rank	Pillai's Trace	0.167	1.639	0.119

PQ1= the score on practice question 1; PQ2 = the score on practice question 2;
AGE = age of participants; GEN = gender of participants.
SMRT = the score on Mental Rotation Test.
TSMRT = time spent on Mental Rotations Test.
Treatment = manipulation variables.
Rank = the order of performing trend analysis task first then pattern recognition task and vise versa.
Treatment * Rank = interaction between the manipulation variables and the order of performing tasks.

The first dependent variable used to test hypothesis H1a was the first question in the trend analysis task, which asked the participants which asked what the data differences were between years 1 and 4. Table 14, Panel A indicates that on average those participants viewing the tabular display (mean score 3.481) were the most accurate (had the highest score) on this first trend analysis task. However, those participants

viewing a 3-D perspective display (mean score 2.793) were 18% less accurate than those viewing the 2-D display (mean score 3.400).

Table 14, Panel A suggests that score on practice question two is significantly ($p < 0.001$) associated with the accuracy of the participants in describing what the data difference were between years 1 and 4. As expected, the results also suggest that the manipulation of the presentation formats (Treatment) has a significant ($p = 0.003$) main effect on the accuracy of the participants in this trend analysis task. A paired comparison test was conducted to determine if the participants receiving the 2-D Treatment were more effective or accurate than those participants receiving the Tabular or 3-D Treatment. Results revealed that the participants viewing the 2-D displays were significantly ($p = 0.012$) more accurate than participants viewing the 3-D perspective display (see Table 14, Panel B). There was no significant difference in the effectiveness or accuracy between participants viewing the 2-D displays and participants viewing the tabular display. Thus, the paired comparison tests provide partial support for hypothesis H1a.

Table 14
Pilot Test Results of H1a
ANCOVA Model on Effectiveness (Accuracy) in Trend Analysis Task
'What are the differences between year 1 and 4?'

Tests of Between-Subjects Effects on Effectiveness

Panel A: Mean Score on the Task and ANCOVA Results Using Score as the Dependent Variable.

	Actual Mean	**ANCOVA Adjusted Mean***
Tabular Display (n = 27)	3.481	3.404
2-D Displays (n = 30)	3.400	3.368
3-D Perspective Display (n = 29)	2.793	2.891

Source of Variation	**Type III SS**	**DF**	**Mean Square**	**F stat**	**P value***
Corrected Model	35.342	11	3.213	8.070	<0.001
Intercept	1.182	1	1.182	2.968	0.089
PQ1	0.077	1	0.077	0.194	0.661
PQ2	14.099	1	14.099	35.413	<0.001
AGE	0.003	1	0.003	0.007	0.931
GEN	0.923	1	0.923	2.317	0.132
SMRT	1.266	1	1.266	3.181	0.079
TSMRT	0.009	1	0.009	0.022	0.884
Treatment	4.233	2	2.116	5.316	0.003
Rank	0.005	1	0.005	0.012	0.915
Treatment*Rank	2.020	2	1.010	2.537	0.086
Error	29.460	74	0.398		
Total	957.000	86			
Corrected Total	64.802	85			

*Adjusted Mean is for the effect of the covariate. Adjusted R-Squared = 0.478.
*Treatment p-values are one-tail, all others are two-tail. See Table 11 for definition of the variables.

Panel B: Bonferroni Pairwise Comparisons for Test H1a

(I) Treatment (J) Treatment	**Mean Difference (I-J)**	**Std Error**	**p value***
2-D Displays Tabular Display	-0.036	0.176	0.500
3-D Display	0.477	0.176	0.012

*p-values are one-tail.

The second dependent variable used to test hypothesis H1a was the second question in the trend analysis task, which asked the participants to write short sentences describing what they perceive to be occurring in the data when going from year 2 to year 3 and year 4. Table 15, Panel A indicates that on average those participants viewing the 2-D displays (mean score 5.866) were the most accurate (had the highest score) on this first trend analysis task. Those participants viewing a 3-D perspective display (means score 3.862) were 35% less accurate than those viewing the 2-D display (mean score 5.866). Those participants viewing the tabular display (mean score 5.777) were 2% less accurate than those viewing the 2-D display (mean score 5.866).

Table 15, Panel A suggests that, as expected, the results suggest that the manipulation of the presentation formats (Treatment) has a significant ($p = 0.003$) main effect on the accuracy of the participants in this trend analysis task. A paired comparison test was conducted to determine if the participants receiving the 2-D Treatment were more effective or accurate than those participants receiving the Tabular or 3-D Treatment. Results revealed that the participants viewing the 2-D displays were significantly ($p = 0.005$) more accurate than participants viewing the 3-D perspective display (see Table 15, Panel B). There was no significant difference in the effectiveness or accuracy between participants viewing the 2-D displays and participants viewing the tabular display. Thus, the paired comparison tests provide partial support for hypothesis H1a.

Table 15
Pilot Test Results of H1a
ANCOVA Model on Effectiveness (Accuracy) in Trend Analysis Task
'What is happening as you go from year 2 to year 3 to year 4?'

Tests of Between-Subjects Effects on Effectiveness

Panel A: Mean Score on the Task and ANCOVA Results Using Score as the Dependent Variable.

	Actual Mean	ANCOVA Adjusted Mean*
Tabular Display (n = 27)	5.777	5.656
2-D Displays (n = 30)	5.866	5.897
3-D Perspective Display (n = 29)	3.862	3.895

Source of Variation	**Type III SS**	**DF**	**Mean Square**	**F stat**	**p value***
Corrected Model	152.995	11	13.909	2.435	0.012
Intercept	4.617	1	4.617	0.808	0.372
PQ1	0.692	1	0.692	0.121	0.729
PQ2	14.800	1	14.800	2.591	0.112
AGE	0.010	1	0.010	0.002	0.967
GEN	11.469	1	11.469	2.008	0.161
SMRT	1.400	1	1.400	0.245	0.622
TSMRT	0.033	1	0.033	0.006	0.939
Treatment	61.640	2	30.823	5.396	0.003
Rank	9.663	1	9.663	1.692	0.197
Treatment*Rank	22.671	2	11.336	1.984	0.145
Error	422.726	74	5.713		
Total	2868..000	86			
Corrected Total	575.721	85			

*Adjusted Mean is for the effect of the covariate. Adjusted R-Squared = 0.157.
*Treatment p-values are one-tail, all others are two-tail. See Table 11 for definition of the variables.

Panel B: Bonferroni Pairwise Comparisons for Test H1a

(I) Treatment (J) Treatment	**Mean Difference (I-J)**	**Std Error**	**p value***
2-D Displays Tabular Display	0.241	0.666	0.500
3-D Display	2.002	0.667	0.005

*p-values are one-tail.

The third dependent variable used to test hypothesis H1a was the fifth question in the trend analysis task, which asked participants to indicate the differences between years 2 and 4 by selecting choices from a given template. Table16, Panel A shows that those participants viewing the 2-D displays (3.866) were on average the most accurate (had the highest score) on this trend analysis task. Those participants viewing a tabular display (mean score 3.407) were 12% less accurate than those viewing the 2-D display (mean score 3.866). Those participants viewing the 3-D perspective display (mean score 3.620) were 7% less accurate than those viewing the 2-D display (mean score 3.866).

Table 16, Panel A suggests that, as expected, the manipulation of the presentation formats (Treatment) has a significant ($p = 0.009$) main effect on the accuracy of the participants in describing the differences between years 2 and 4. A paired comparison test was conducted to determine if the participants receiving the 2-D Treatment were more effective or accurate than those participants receiving the Tabular or 3-D Treatment. Results revealed that the participants viewing the 2-D displays were significantly ($p = 0.007$) more accurate than participants viewing the Tabular display (see Table 16, Panel B). There was no significant difference in the effectiveness or accuracy between participants viewing the 2-D displays and participants viewing the 3-D perspective display. Thus, the paired comparison tests provide partial support for hypothesis H1a.

Table 16
Pilot Test Results of H1a
ANCOVA Model on Effectiveness (Accuracy) in Trend Analysis Task
'What are the differences between year 2 and 4?'

Tests of Between-Subjects Effects on Effectiveness

Panel A: Mean Score on the Task and ANCOVA Results Using Score as the Dependent Variable.

	Actual Mean	ANCOVA Adjusted Mean*
Tabular Display (n = 27)	3.407	3.380
2-D Displays (n = 30)	3.866	3.858
3-D Perspective Display (n = 29)	3.620	3.662

Source of Variation	Type III SS	DF	Mean Square	F stat	p value*
Corrected Model	7.645	11	0.695	1.964	0.044
Intercept	21.669	1	21.669	61.246	<0.001
PQ1	0.643	1	0.643	1.818	0.182
PQ2	0.341	1	0.341	0.963	0.330
AGE	0.729	1	0.729	2.059	0.155
GEN	0.133	1	0.133	0.375	0.542
SMRT	0.014	1	0.014	0.041	0.840
TSMRT	0.450	1	0.450	1.272	0.263
Treatment	2.981	2	1.490	4.212	0.009
Rank	0.044	1	0.044	0.124	0.726
Treatment*Rank	1.206	2	0.603	1.704	0.189
Error	26.181	74	0.354		
Total	1173.00	86			
Corrected Total	33.826	85			

*Adjusted Mean is for the effect of the covariate. Adjusted R-Squared = 0.111.
*Treatment p-values are one-tail, all others are two-tail. See Table 11 for definition of the variables.

Panel B: Bonferroni Pairwise Comparisons for Test H1a

(I) Treatment (J) Treatment	Mean Difference (I-J)	Std Error	p value*
2-D Displays Tabular Display	0.478	0.166	0.007
3-D Display	0.196	0.166	0.362

*p-values are one-tail.

The fourth dependent variable used to test hypothesis H1a was the sixth question in the trend analysis task, which asked participants to select from a template to indicate changes in data when going from year 1 to year 2 to year 3. Those participants viewing the 2-D displays (mean score 7.666) were on average the most accurate (had the highest score) on this trend analysis task. Those participants viewing the tabular display (mean score 7.222) were 6% less accurate than those viewing the 2-D display (mean score 7.666). Those participants viewing a 3-D perspective display (mean score 6.931) were 10% less accurate than those viewing the 2-D display (mean score 7.666).

Results of the ANCOVA analysis show that the covariates score on practice question two ($p = 0.010$), age of participants ($p = 0.015$), and gender of participants ($p=0.040$) are significantly associated with the accuracy of the participants in selecting from a template to indicate changes in data when going from year 1 to year 2 to year 3. As expected, the results also suggest that the manipulation of the presentation formats (Treatment) has a significant ($p = 0.040$) main effect on the accuracy of the participants in this trend analysis task. A paired comparison test was conducted to determine if the participants receiving the 2-D Treatment were more effective or accurate than those participants receiving the Tabular or 3-D Treatment. Results revealed that the participants viewing the 2-D displays were not significantly ($p = 0.051$) more accurate than participants viewing the 3-D perspective display. There was no significant difference in the effectiveness or accuracy between participants viewing the 2-D displays and participants viewing the tabular display.

4.2.2.2 Results of Hypothesis H1b. Hypothesis H1b has four dependent measures of efficiency − the time spent in seconds by each participant when answering what the data differences were between years 1 and 4, the time spent in seconds by each participant in describing what was occurring in the data when going from year 2 to year 3 and year 4, the time spent in seconds by each participant when selecting from a template the differences in data between years 2 and 4, and the time spent in seconds by each participant when selecting from a template changes in data when going from year 1 to year 2 to year 3.

The mean results for all dependent variables are hypothesized in H1b. In constructing the models to test H1b, all covariates mentioned in section 4.2.2 were included in the model along with the manipulated variable Treatment.

Prior to presenting ANCOVA results for the four dependent variables used to test hypothesis H1b, a MANCOVA analysis was conducted. As shown (Table 17), the overall F-statistic for the manipulated variable Treatment is significant ($p < 0.001$) using Pillai's Trace. The results are also significant ($p < 0.001$) using Hotelling's Trace. These significant results allow for analysis of the univariate results which are provided on Table 18 to 19.

The first dependent variable used to test hypothesis H1b was the time spent in seconds by each participant when answering what the data differences were between years 1 and 4. On average those participants viewing the 2-D displays (mean seconds 127) were the most efficient or used the least amount of time (in seconds) on this first trend analysis task. Those participants viewing a 3-D perspective display (mean seconds 148) used 16% more time (in seconds), than those viewing the 2-D displays (mean

seconds 127). Those participants viewing the tabular display (mean seconds 132) used 4% more time (in seconds), than those participants viewing the 2-D displays (mean seconds 127). Results of the ANCOVA analysis show that age of the participants is significantly ($p < 0.001$) associated with the time spent by each participant in describing what the data differences were between years 1 and 4. However, the results suggest that the manipulation of the presentation formats (Treatment) do not have a significant ($p = 0.118$) main effect on the time spent by each participant in this trend analysis task. Since there is not a significant main effect a paired comparison test was not conducted.

Table 17
Multivariate Tests of H1b.

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.418	12.726	<0.001
PQ1	Pillai's Trace	0.016	0.294	0.881
PQ2	Pillai's Trace	0.102	2.025	0.100
AGE	Pillai's Trace	0.183	3.979	0.006
GEN	Pillai's Trace	0.031	0.575	0.682
SMRT	Pillai's Trace	0.065	1.236	0.304
TSMRT	Pillai's Trace	0.095	1.873	0.125
Treatment	Pillai's Trace	0.431	4.941	<0.001
Rank	Pillai's Trace	0.486	16.803	<0.001
Treatment*Rank	Pillai's Trace	0.150	1.455	0.179

PQ1= the score on practice question 1; PQ2 = the score on practice question 2;
AGE = age of participants; GEN = gender of participants.
SMRT = the score on Mental Rotation Test.
TSMRT = time spent on Mental Rotations Test.
Treatment = manipulation variables.
Rank = the order of performing trend analysis task first then pattern recognition task and vise versa.
Treatment * Rank = interaction between the manipulation variables and the order of performing tasks.

The second dependent variable used to test hypothesis H1b was the time spent in seconds by each participant when describing what was occurring in the data when going from year 2 to year 3 and year 4. On average those participants viewing the 2-D display (mean seconds 129) were the most efficient (used the least amount of time in seconds) on this trend analysis task. Those participants viewing a 3-D perspective display (mean seconds 155) used 20% more time (in seconds), than those viewing the 2-D displays (mean seconds 129). Those participants viewing the tabular display (mean seconds 137) used 6% more time (in seconds), than those participants viewing the 2-D displays (mean seconds 129). Results of the ANCOVA analysis show that age of the participant is significantly ($p < 0.001$) associated with the time spent by each participant in describing what was occurring in the data when going from year 2 to year 3 and year 4. However, the results suggest that the manipulation of the presentation formats (Treatment) do not have a significant ($p = 0.254$) main effect on the time spent by each participant in this trend analysis task. Since there is not a significant main effect a paired comparison test was not conducted.

The third dependent variable used to test hypothesis H1b was the time spent in seconds by each participant when selecting from a template the differences in data between years 2 and 4. Table 18, Panel A indicates that on average those participants viewing the 2-D display (mean seconds 32) were the most efficient (used the least amount of time in seconds) on this trend analysis task. Those participants viewing a 3-D perspective display (mean seconds 50) used 56% more time (in seconds), than those viewing the 2-D displays (mean seconds 32). Those participants viewing the tabular

display (mean score 34) used 6% more time than those participants viewing the 2-D displays (mean score 32).

Table 18, Panel A suggests that score on the score on the Mental Rotation Test is significantly ($p < 0.047$) associated with the time spent by each participant when selecting from a template the differences in data between years 2 and 4. The results suggest that the manipulation of the presentation formats (Treatment) has a significant ($p <0.001$) main effect on the time spent by each participant in this trend analysis task. A paired comparison test was conducted to determine if the participants receiving the 2-D Treatment were more efficient than those participants receiving the Tabular or 3-D Treatment. Results revealed that the participants viewing the 2-D displays were significantly ($p < 0.001$) more efficient than participants viewing the 3-D perspective display (see Table 18, Panel B). There was no significant difference in the efficiency between participants viewing the tabular display and participants viewing the 2-D perspective display. Thus, the paired comparison tests provide partial support for hypothesis H1b.

Table 18
Pilot Test Results of H1b
ANCOVA Model on Efficiency (Time Spent) in Trend Analysis Task
'What are the differences between year 2 and 4?'

Tests of Between-Subjects Effects on Efficiency

Panel A: Mean Time Spent on the Task and ANCOVA Results Using Time Spent as the Dependent Variable

	Actual Mean (Seconds)	ANCOVA Adjusted Mean* (Seconds)
Tabular Display (n = 27)	34.407	33.812
2-D Displays (n = 30)	32.000	32.298
3-D Perspective Display (n = 29)	50.482	50.512

Source of Variation	Type III SS	DF	Mean Square	F stat	p value*
Corrected Model	7867.174	11	715.198	5.214	<0.001
Intercept	5777.610	1	5777.610	42.123	<0.001
PQ1	65.148	1	65.148	0.475	0.493
PQ2	33.177	1	33.177	0.242	0.624
AGE	39.769	1	39.769	0.290	0.592
GEN	154.097	1	154.097	1.123	0.293
SMRT	557.712	1	557.712	4.066	0.047
TSMRT	130.348	1	130.348	0.950	0.333
Treatment	5274.726	2	2637.363	19.228	<0.001
Rank	756.492	1	756.492	5.515	0.022
Treatment*Rank	19.044	2	9.522	0.069	0.933
Error	10149.815	74	137.160		
Total	148745.000	86			
Corrected Total	18016.988	85			

*Adjusted Mean is for the effect of the covariate. Adjusted R-Squared = 0.353.
*Treatment p-values are one-tail, all others are two-tail. See Table 11 for definition of the variables.

Panel B: Bonferroni Pairwise Comparisons for Test H1b

(I) Treatment	(J) Treatment	Mean Difference (I-J)	Std Error	p value*
2-D Displays	Tabular Display	-1.514	3.262	0.500
	3-D Display	-18.214	3.267	<0.001

*p-values are one-tail

The fourth dependent variable used to test hypothesis H1b was the time spent in seconds by each participant when selecting from a template changes in data when going from year 1 to year 2 to year 3. Table 19, Panel A indicates that on average those participants viewing the 2-D display (mean seconds 54) were the most efficient (used the least amount of time in seconds) on this trend analysis task. Those participants viewing a 3-D perspective display (mean seconds 100) used 85% more time (in seconds), than those viewing the 2-D displays (mean seconds 54). Those participants viewing the tabular display (means seconds 69) used 27% more time than those participants viewing the 2-D displays (mean seconds 54).

Table 19, Panel A suggests that the score on practice question two ($p = 0.028$) and the time spent on the Mental Rotation Test ($p = 0.015$) are significantly associated with the time spent by each participant in selecting from a template changes in data when going from year 1 to year 2 to year 3.The results suggest that the manipulation of the presentation formats (Treatment) has a significant ($p < 0.001$) main effect on the time spent by each participant in this trend analysis task. A paired comparison test was conducted. Results revealed that the participants viewing the 2-D displays were significantly ($p < 0.001$) more efficient than participants viewing the 3-D perspective display (see Table 19, Panel B). There was no significant difference in the efficiency between participants viewing the tabular display and participants viewing the 2-D perspective display. Thus, the paired comparison tests provide partial support for hypothesis H1b.

Table 19
Pilot Test Results of H1b
ANCOVA Model on Efficiency (Time Spent) in Trend Analysis Task
'What is happening as you go from year 1 to year 2 to year 3?'

Tests of Between-Subjects Effects on Efficiency

Panel A: Mean Time Spent on the Task and ANCOVA Results Using Time Spent as the Dependent Variable

	Actual Mean (Seconds)	ANCOVA Adjusted Mean* (Seconds)
Tabular Display (n = 27)	69.296	67.979
2-D Displays (n = 30)	54.466	57.824
3-D Perspective Display (n = 29)	100.206	96.166

Source of Variation	Type III SS	DF	Mean Square	F stat	p value*
Corrected Model	43916.8420	11	3992.440	5.118	<0.001
Intercept	21661.621	1	21661.621	27.768	<0.001
PQ1	49.906	1	49.906	0.064	0.801
PQ2	3943.532	1	3943.532	5.055	0.028
AGE	54.234	1	54.234	0.070	0.793
GEN	528.119	1	528.119	0.677	0.413
SMRT	994.898	1	994.898	1.275	0.262
TSMRT	4876.167	1	4876.167	6.251	0.015
Treatment	22694.485	2	11347.243	14.546	<0.001
Rank	93.118	1	93.118	0.119	0.731
Treatment*Rank	1234.886	2	617.443	0.792	0.457
Error	57726.472	74	780.087		
Total	579561.000	86			
Corrected Total	101643.314	85			

*Adjusted Mean is for the effect of the covariate. Adjusted R-Squared = 0.348.
*Treatment p-values are one-tail, all others are two-tail. See Table 11 for definition of the variables.

Panel B: Bonferroni Pairwise Comparisons for Test H1b

(I) Treatment	(J) Treatment	Mean Difference (I-J)	Std Error	p value*
2-D Displays	Tabular Display	-10.155	7.780	0.294
	3-D Display	-40.342	7.790	<0.001

*p-values are one-tail.

4.2.2.3 Results of Hypothesis H1c. Hypothesis H1c has five dependent measures of accuracy − the score on the third question, which asked the participants to estimate what the ROE would be in year 6 if each of the variables comprising ROE in year 5 doubles; and the scores on parts 'a', 'b', 'c' and 'd' of the fourth question, which asked the participants to estimate the average of turnover, profitability and leverage for the years 1, 2, 4 and 5, and use the estimated average to calculate a new ROE.

The mean results for all dependent variables are hypothesized in H1c. In constructing the models to test H1c, all covariates mentioned in section 4.2.2 were included in the model along with the manipulated variable Treatment.

A MANCOVA analysis was conducted. As shown (Table 20), the overall statistic for the manipulated variable Treatment is not significant ($p = 0.076$) using Pillai's Trace.The results are also not significant ($p = 0.075$) using Hotelling's Trace. Thus, hypothesis H1c was not supported.

Table 20
Multivariate Tests of H1c

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.202	3.550	0.006
PQ1	Pillai's Trace	0.107	1.671	0.153
PQ2	Pillai's Trace	0.045	0.653	0.660
AGE	Pillai's Trace	0.116	1.830	0.118
GEN	Pillai's Trace	0.099	1.536	0.190
SMRT	Pillai's Trace	0.124	1.979	0.092
TSMRT	Pillai's Trace	0.036	0.529	0.754
Treatment	Pillai's Trace	0.219	1.744	0.076
Rank	Pillai's Trace	0.070	1.051	0.395
Treatment*Rank	Pillai's Trace	0.124	0.938	0.500

PQ1= the score on practice question 1; PQ2 = the score on practice question 2;
AGE = age of participants; GEN = gender of participants.
SMRT = the score on Mental Rotation Test.
TSMRT = time spent on Mental Rotations Test.
Treatment = manipulation variables.
Rank = the order of performing trend analysis task first then pattern recognition task and vise versa.
Treatment * Rank = interaction between the manipulation variables and the order of performing tasks.

4.2.2.4 Result of Hypothesis H1d. Hypothesis H1d has two dependent measures of efficiency − the time spent in seconds by each participant when estimating what the ROE would be in year 6 if each of the variables comprising ROE in year 5 doubles, and the time spent in seconds by each participant when estimating the average of turnover, profitability and leverage for the years 1, 2, 4 and 5, and using the estimated average to calculate a new ROE.

The mean results for all dependent variables are hypothesized in H1d. In constructing the models to test H1d, all covariates mentioned in section 4.2.2 were included in the model along with the manipulated variable Treatment.

A MANCOVA analysis was conducted. As shown (Table 21), the overall F-statistic for the manipulated variable Treatment is not significant (p = 0.350) using

Pillai's Trace. The results are also not significant (p = 0.351) using Hotelling's Trace. Hypothesis H1d was not supported.

Table 21
Multivariate Tests of H1d.

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.059	2.285	0.109
PQ1	Pillai's Trace	0.010	0.364	0.696
PQ2	Pillai's Trace	0.021	0.771	0.466
AGE	Pillai's Trace	0.117	4.825	0.011
GEN	Pillai's Trace	0.005	0.182	0.834
SMRT	Pillai's Trace	0.007	0.241	0.766
TSMRT	Pillai's Trace	0.003	0.118	0.889
Treatment	Pillai's Trace	0.059	1.118	0.350
Rank	Pillai's Trace	0.023	0.866	0.425
Treatment*Rank	Pillai's Trace	0.073	1.411	0.233

PQ1= the score on practice question 1; PQ2 = the score on practice question 2;
AGE = age of participants; GEN = gender of participants.
SMRT = the score on Mental Rotation Test.
TSMRT = time spent on Mental Rotations Test.
Treatment = manipulation variables.
Rank = the order of performing trend analysis task first then pattern recognition task and vise versa.
Treatment * Rank = interaction between the manipulation variables and the order of performing tasks.

4.2.3. Hypotheses Testing of the Pattern Analysis Task

Results of H2a, H2b, H2c, and H2d are shown in tables 22 to 27. Multivariate MANCOVA was first used to test for the significance of the manipulation variable Treatment. If the manipulation variable Treatment was significant, ANCOVA was then conducted to analysis the univariate results. Covariates included in the analysis are the scores on practice questions one and two, gender, age, score on the Mental Rotations Test, time spent on the Mental Rotations Test, and the task order performed by each participant (for example the trend analysis task first and the pattern recognition task second, and vice versa).

The following paragraphs report significant results, supporting the hypothesis, with tables on top of narration.

4.2.3.1 Results of Hypothesis H2a. Hypothesis H2a has two dependent measures of accuracy − the score on the second question, which asked the participants to separate companies one through six into two groups based on similar financial characteristics; and the score on the fourth question which asked the participants to select one of the six companies if the goal is to have high profitability, high turnover but low leverage at the same time.

The mean results for all dependent variables are hypothesized in H2a. In constructing the models to test H2a, all covariates mentioned in section 4.2.3 were included in the model along with the manipulated variable Treatment.

Prior to presenting ANCOVA results for the two dependent variables used to test hypothesis H2a, a MANCOVA analysis was conducted. As shown (Table 22), the overall F-statistic for the manipulated variable Treatment is significant ($p < 0.001$) using

Pillai's Trace. The results are also significant ($p < 0.001$) using Hotelling's Trace. These significant results allow for analysis of the univariate results which are provided on Table 23.

Table 22
Multivariate Tests of H2a

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.419	26.318	<0.001
PQ1	Pillai's Trace	0.017	0.649	0.525
PQ2	Pillai's Trace	0.016	0.579	0.563
AGE	Pillai's Trace	0.087	3.494	0.036
GEN	Pillai's Trace	0.039	1.472	0.236
SMRT	Pillai's Trace	0.172	7.606	<0.001
TSMRT	Pillai's Trace	0.014	0.510	0.602
Treatment	Pillai's Trace	0.470	11.379	<0.001
Rank	Pillai's Trace	0.023	0.867	0.425
Treatment*Rank	Pillai's Trace	0.139	2.755	0.030

PQ1= the score on practice question 1; PQ2 = the score on practice question 2;
AGE = age of participants; GEN = gender of participants.
SMRT = the score on Mental Rotation Test.
TSMRT = time spent on Mental Rotations Test.
Treatment = manipulation variables.
Rank = the order of performing trend analysis task first then pattern recognition task and vise versa.
Treatment * Rank = interaction between the manipulation variables and the order of performing tasks.

The first dependent variable used to test hypothesis H2a was the second question in the pattern recognition task, which asked participants to separate companies one through six into two groups based on similar financial characteristics. Those participants viewing the 3-D perspective display (mean score 5.069) were on average the most accurate (had the highest score) on this pattern recognition task. Those participants viewing a tabular display (mean score 4.555) were 10% less accurate than those viewing the 3-D perspective display (mean score 5.069). Those participants viewing the 2-D

displays (mean score 4.800) were 5% less accurate than those viewing the 3-D perspective display (mean score 5.069).

Results of the ANCOVA show that the covariates score on Mental Rotations Test ($p < 0.001$) is significantly associated with the accuracy of the participants in separating companies one through six into two groups based on similar financial characteristics. Results also, contrary to expectation, show that the manipulation of the presentation formats (Treatment) does not have a significant ($p = 0.129$) main effect on the accuracy of the participants in this pattern recognition task. Since there is not a significant main effect a paired comparison test was not conducted.

The second dependent variable used to test hypothesis H2a was the fourth question in the pattern recognition task which asked participants to select one of the six companies given the goal was to have higher profitability, higher turnover but lower leverage. Table 23, Panel A suggests that those participants viewing the 3-D perspective display (mean score 0.724) were on average the most accurate (had the highest score) on this pattern recognition task. Those participants viewing a tabular display (mean score 0.000) were 100% less accurate than those viewing the 3-D perspective display (mean score 0.724). Those participants viewing the 2-D displays (mean score 0.200) were 73% less accurate than those viewing the 3-D perspective display (mean score 0.724).

Table 23, Panel A suggests that age of the participants is significantly ($p = 0.033$) associated with the accuracy of the participants in selecting the company given the goal was to have higher profitability, higher turnover but lower leverage. Panel A also suggests that the interaction between the presentation formats (Treatment) and the Task Order has a significant effect ($p = 0.035$) for the accuracy of the participants in this

pattern recognition task. A paired comparison test was conducted to determine if the participants receiving the 3-D Treatment were more accurate than those participants receiving the Tabular or 2-D Treatment. Results revealed that the participants viewing the 3-D perspective display were significantly ($p < 0.001$) more accurate than participants viewing the Tabular display and participants viewing the 2-D displays (see Table 23, Panel B).

Additional analysis on the interaction effect between the presentation format (Treatment) and the Task Order revealed that among the participants performing trend analysis task first and the pattern recognition task later, those participants viewing the 3-D perspective display were significantly ($p<0.0001$) more accurate than participants viewing the Tabular display and participants viewing the 2-D displays. However, among the participants performing pattern recognition task first and the trend analysis task later, those participants viewing the 3-D perspective display were significantly ($p = 0.001$) more accurate than participants viewing the Tabular display but not significantly more accurate than participants viewing the 2-D displays ($p = 0.082$). (Results of additional analysis not shown in tables here).

Table 23
Pilot Test Results of H2a
ANCOVA Model on Effectiveness (Accuracy) in Pattern Recognition Task
'Assuming you cannot select a company solely because of a single variable, for example highest profitability. Comparatively, if it is better to have a higher profitability, higher turnover but lower leverage at the same time, which company you will select?'

Tests of Between-Subjects Effects on Effectiveness

Panel A: Mean Score on the Task and ANCOVA Results Using Score as the Dependent Variable.

	Actual Mean	ANCOVA Adjusted Mean*
Tabular Display (n = 27)	0.000	0.004
2-D Displays (n = 30)	0.200	0.194
3-D Perspective Display (n = 29)	0.724	0.731

Source of Variation	Type III SS	DF	Mean Square	F	p value*
Corrected Model	9.482	11	0.862	7.055	<0.001
Intercept	0.145	1	0.145	1.184	0.280
PQ1	0.000	1	0.000	0.004	0.950
PQ2	0.000	1	0.000	0.003	0.954
AGE	0.574	1	0.574	4.698	0.033
GEN	0.000	1	0.000	0.003	0.957
SMRT	0.025	1	0.025	0.207	0.650
TSMRT	0.126	1	0.126	1.034	0.312
Treatment	7.363	2	3.681	30.129	<0.001
Rank	0.012	1	0.012	0.098	0.755
Treatment*Rank	0.856	2	0.428	3.505	0.035
Error	9.042	74	0.122		
Total	27.000	86			
Corrected Total	18.523	85			

*Adjusted Mean is for the effect of the covariate. Adjusted R-Squared = 0.439.
*Treatment p-values are one-tail, all others are two-tail. See Table 11 for definition of the variables.

Panel B: Bonferroni Pairwise Comparisons for Test H2a

(I) Treatment	(J) Treatment	Mean Difference (I-J)	Std Error	p value*
3-D Display	Tabular Display	0.727	0.097	<0.001
	2-D Displays	0.537	0.097	<0.001

*p-values are one-tail.

4.2.3.2 Results of Hypothesis H2b. Hypothesis H2b has two dependent measures of efficiency − the time spent in seconds by each participant in separating companies one through six into two groups based on similar financial characteristics, and the time spent in seconds by each participant in selecting one of the six companies if the goal is to have high profitability, high turnover but low leverage at the same time.

The mean results for all dependent variables are hypothesized in H2b. In constructing the models to test H2b, all covariates mentioned in section 4.2.3 were included in the model along with the manipulated variable Treatment.

A MANCOVA analysis was conducted. As shown (Table 24), the overall F-statistic for the manipulated variable Treatment is not significant (p = 0.312) using Pillai's Trace. The results are also not significant (p = 0.315) using Hotelling's Trace. Hypothesis H2b was not supported.

Table 24
Multivariate Tests of H2b

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.118	4.905	0.010
PQ1	Pillai's Trace	0.035	1.307	0.277
PQ2	Pillai's Trace	0.002	0.065	0.937
AGE	Pillai's Trace	0.024	0.884	0.417
GEN	Pillai's Trace	0.000	0.111	0.989
SMRT	Pillai's Trace	0.076	3.008	0.056
TSMRT	Pillai's Trace	0.293	15.103	<0.001
Treatment	Pillai's Trace	0.063	1.202	0.312
Rank	Pillai's Trace	0.017	0.613	0.544
Treatment*Rank	Pillai's Trace	0.113	2.222	0.069

PQ1= the score on practice question 1; PQ2 = the score on practice question 2;
AGE = age of participants; GEN = gender of participants. SMRT = the score on Mental Rotation Test.
TSMRT = time spent on Mental Rotations Test. Treatment = manipulation variables.
Rank = the order of performing trend analysis task first then pattern recognition task and vise versa.
Treatment * Rank = interaction between the manipulation variables and the order of performing tasks.

4.2.3.3 Results of Hypothesis H2c. Hypothesis H2c has four dependent measures of accuracy − the score on the first question, which asked the participants what the data differences were between companies one and six; the score on the third question, which asked the participants to describe the pattern of financial ratios they were seeing in group one compared to group two; the score on the fifth question, which asked the participants what the data differences were between companies four and six by selecting choices from a template; and the score on the sixth question, which asked the participants to describe the pattern of financial ratios they were seeing in group two compared to group one by selecting choices from a template.

The mean results for all dependent variables are hypothesized in H2c. In constructing the models to test H2c, all covariates mentioned in section 4.2.3 were included in the model along with the manipulated variable Treatment.

Prior to presenting ANCOVA results for the two dependent variables used to test hypothesis H2c, a MANCOVA analysis was conducted. As shown (Table 25), the overall F-statistic for the manipulated variable Treatment is significant ($p = 0.010$) using Pillai's Trace. The results are also significant ($p = 0.011$) using Hotelling's Trace. These significant results allow for analysis of the univariate results which are provided on Table 26.

The first dependent variable used to test hypothesis H2c was the first question in the pattern recognition task, which asked participants what the data differences were between companies one and six. Those participants viewing the tabular display (mean score 2.777) were on average the most accurate (had the highest score) on this pattern recognition task. Those participants viewing a tabular display (means score 2.777) were

3% more accurate than those viewing the 3-D perspective display (mean score 2.689). Those participants viewing the 2-D displays (mean score 2.566) were 5% less accurate than those viewing the 3-D perspective display (mean score 2.689). Results of the ANCOVA show that, contrary to expectation, the manipulation of the presentation formats (Treatment) does not have a significant ($p = 0.290$) main effect on the accuracy of the participants in describing what the data differences were between companies one and six. Since there is not a significant main effect a paired comparison test was not conducted.

Table 25
Multivariate Tests of H2c.

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.643	31.991	<0.001
PQ1	Pillai's Trace	0.053	0.998	0.414
PQ2	Pillai's Trace	0.065	1.228	0.307
AGE	Pillai's Trace	0.083	1.605	0.182
GEN	Pillai's Trace	0.097	1.903	0.119
SMRT	Pillai's Trace	0.029	0.530	0.714
TSMRT	Pillai's Trace	0.094	1.832	0.132
Treatment	Pillai's Trace	0.256	2.645	0.010
Rank	Pillai's Trace	0.014	0.249	0.910
Treatment*Rank	Pillai's Trace	0.116	1.112	0.358

PQ1= the score on practice question 1; PQ2 = the score on practice question 2;
AGE = age of participants; GEN = gender of participants.
SMRT = the score on Mental Rotation Test.
TSMRT = time spent on Mental Rotations Test.
Treatment = manipulation variables.
Rank = the order of performing trend analysis task first then pattern recognition task and vise versa.
Treatment * Rank = interaction between the manipulation variables and the order of performing tasks.

The second dependent variable used to test hypothesis H2c was the third question in the pattern recognition task which asked participants to describe the pattern of financial ratios they were seeing in group one compared to group two. Those participants viewing the 3-D perspective display (mean score 2.517) were on average the most accurate (had the highest score) on this pattern recognition task. Those participants viewing a tabular display (mean score 1.777) were 31% less accurate than those viewing the 3-D perspective display (mean score 2.517). Those participants viewing the 2-D displays (mean score 2.333) were 8% less accurate than those viewing the 3-D perspective display (mean score 2.517)

Table 26, Panel A suggests that the manipulation of the presentation formats (Treatment) has a significant ($p = 0.001$) main effect on the on the accuracy of the participants in describing the pattern of financial ratios they were seeing in group one compared to group two. A paired comparison test was conducted to determine if the participants receiving the 3-D Treatment were more accurate than those participants receiving the Tabular or 2-D Treatment. Results revealed that the participants viewing the 3-D perspective display were significantly ($p = 0.002$) more accurate than participants viewing the Tabular display. There was no significant difference in the efficiency between participants viewing the 2-D displays and participants viewing the 3-D perspective display (see Table 26 Panel B). Thus, the paired comparison tests provide partial support for hypothesis H2c.

Table 26
Pilot Test Results of H2c
ANCOVA Model on Effectiveness (Accuracy) in Pattern Recognition Task
'Group one includes companies 1, 3, 4, and 6, and group two includes companies 2 and 5. Compared to group two, what are the patterns of the financial ratios you are seeing in group one?

Tests of Between-Subjects Effects on Effectiveness

Panel A: Mean Score on the Task and ANCOVA Results Using Score as the Dependent Variable.

	Actual Mean	**ANCOVA Adjusted Mean***
Tabular Display (n = 27)	1.777	1.727
2-D Displays (n = 30)	2.333	2.387
3-D Perspective Display (n = 29)	2.517	2.501

Source of Variation	**Type III SS**	**DF**	**Mean Square**	**F**	**p value***
Corrected Model	22.149	11	2.014	2.942	0.003
Intercept	1.346	1	1.346	1.966	0.165
PQ1	1.493	1	1.493	2.182	0.144
PQ2	1.825	1	1.825	2.667	0.107
AGE	0.265	1	0.265	0.387	0.536
GEN	1.795	1	1.795	2.623	0.110
SMRT	0.832	1	0.832	1.215	0.274
TSMRT	1.920	1	1.920	2.805	0.098
Treatment	9.071	2	4.535	6.626	0.001
Rank	0.197	1	0.197	0.288	0.593
Treatment*Rank	1.645	2	0.822	1.202	0.307
Error	50.653	74	0.685		
Total	497	86			
Corrected Total	72.802	85			

*Adjusted Mean is for the effect of the covariate. Adjusted R-Squared = 0.439.
*****Treatment p-values are one-tail, all others are two-tail. See Table 11 for definition of the variables.

Panel B: Bonferroni Pairwise Comparisons for Test H2c

(I) Treatment	(J) Treatment	Mean Difference (I-J)	Std Error	p value*****
3-D Display	Tabular Display	0.775	0.230	0.002
	2-D Displays	0.114	0.231	0.500

*p-values are one-tail.

The third dependent variable used to test hypothesis H2c was the fifth question in the pattern recognition task, which asked participants to what the data differences were between companies four and six by selecting choices from a template. Those participants viewing the 2-D displays (mean score 3.000) were on average the most accurate (had the highest score) on this pattern recognition task. Those participants viewing a tabular display (2.851) were 6% more accurate than those viewing the 3-D perspective display (mean score 2.689). Those participants viewing the 2-D displays (mean score 3.000) were 11% more accurate than those viewing the 3-D perspective display (mean score 2.689).

Results of the ANCOVA show that the manipulation of the presentation formats (Treatment) has a significant ($p = 0.028$) main effect on the accuracy of the participants in what the data differences were between companies four and six by selecting choices from a template. A paired comparison test was conducted to determine if the participants receiving the 3-D Treatment were more effective or accurate than those participants receiving the Tabular or 2-D Treatment. Contrary to prediction, results revealed that the participants viewing the 2-D displays were significantly ($p = 0.026$) more accurate than participants viewing the 3-D perspective display. There was no significant difference in the efficiency between participants viewing the tabular display and participants viewing the 3-D perspective display. Thus, hypothesis H2c was partially not supported.

The fourth dependent variable used to test hypothesis H2c was the sixth question in the pattern recognition task, which asked participants to describe the pattern of financial ratios they were seeing in group two compared to group one. Those participants viewing the 2-D display (mean score 2.833) were on average the most accurate (had the highest score) on this pattern recognition task. Those participants viewing a tabular

display (mean score 2.703) were 2% more accurate than those viewing the 3-D perspective display (mean score 2.655). Those participants viewing the 2-D displays (mean score 2.833) were 7% more accurate than those viewing the 3-D perspective display (mean score 2.655). Results of the ANCOVA show that, contrary to expectation, the manipulation of the presentation formats (Treatment) does not have a significant ($p = 0.400$) main effect on the accuracy of the participants in describing the pattern of financial ratios they were seeing in group two compared to group one. Since there is not a significant main effect a paired comparison test was not conducted.

4.2.3.4 Results of Hypothesis H2d. Hypothesis H2d has four dependent measures of accuracy − the time spent in seconds by each participant in describing what the data differences were between companies one and six; the time spent by each participant in describing the pattern of financial ratios they were seeing in group one compared to group two; the time spent by each participant in describing what the data differences were between companies four and six by selecting choices from a template; and the time spent by each participant in describing the pattern of financial ratios they were seeing in group two compared to group one by selecting choices from a template

The mean results for all dependent variables are hypothesized in H2d. In constructing the models to test H2d, all covariates mentioned in section 4.2.3 were included in the model along with the manipulated variable Treatment.

Prior to presenting ANCOVA results for the two dependent variables used to test hypothesis H2d, a MANCOVA analysis was conducted. As shown (Table 27), the overall F-statistic for the manipulated variable Treatment is significant ($p < 0.001$) using Pillai's

Trace. The results are also significant ($p < 0.001$) using Hotelling's Trace. These significant results allow for analysis of the univariate results.

Table 27
Multivariate Tests of H2d

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.337	9.041	<0.001
PQ1	Pillai's Trace	0.018	0.330	0.857
PQ2	Pillai's Trace	0.099	1.950	0.112
AGE	Pillai's Trace	0.080	1.542	0.199
GEN	Pillai's Trace	0.059	1.115	0.356
SMRT	Pillai's Trace	0.033	0.602	0.663
TSMRT	Pillai's Trace	0.152	3.172	0.019
Treatment	Pillai's Trace	0.414	4.694	<0.001
Rank	Pillai's Trace	0.497	17.542	<0.001
Treatment*Rank	Pillai's Trace	0.189	1.877	0.068

PQ1= the score on practice question 1; PQ2 = the score on practice question 2;
AGE = age of participants; GEN = gender of participants.
SMRT = the score on Mental Rotation Test.
TSMRT = time spent on Mental Rotations Test.
Treatment = manipulation variables.
Rank = the order of performing trend analysis task first then pattern recognition task and vise versa.
Treatment * Rank = interaction between the manipulation variables and the order of performing tasks.

The first dependent variable used to test hypothesis H2d was the time spent by each participant in answering what the data differences were between companies one and six. On average those participants viewing the 2-D display (mean seconds 111) were the most efficient (used the least amount of time in seconds) on this pattern recognition task. Those participants viewing the 2-D displays (mean seconds 111) used 46% less time (in seconds) than those participants viewing the 3-D display (mean seconds 163). While those participants viewing the tabular display (mean seconds 112) used 45% less time (in seconds) than those participants viewing the 3-D perspective display (mean seconds163).

Results of the ANCOVA show that, the manipulation of the presentation formats (Treatment) has a significant ($p < 0.001$) main effect on the efficiency of the participants in describing what the data differences were between companies one and six. A paired comparison test was conducted to determine if the participants receiving the 3-D Treatment were more efficient or used less time (in seconds) than those participants receiving the Tabular or 2-D Treatment. Results show that those participants viewing tabular display ($p <0.001$) and those participants viewing the 2-D displays ($p = 0.001$) were significantly more efficient or used less time (in seconds) than those participant viewing the 3-D perspective display in this pattern recognition task.

The second dependent variable used to test hypothesis H2d was the time spent by each participant in describing the pattern of financial ratios they were seeing in group one compared to group two. On average those participants viewing the 2-D display (mean seconds 97) were the most efficient (used the least amount of time in seconds) on this pattern recognition task. Those participants viewing the 2-D displays (mean seconds 97) used 8% less time (in seconds), than those participants viewing the 3-D display (mean seconds 105). While those participants viewing the tabular display (mean seconds 121) used 15% more time (in seconds), than those participants viewing the 3-D perspective display (mean seconds 105). Results of the ANCOVA show that, the manipulation of the presentation formats (Treatment) does not have a significant ($p = 0.221$) main effect on the time spent by each participants in describing the pattern of financial ratios they were seeing in group one compared to group two. Since there is not a significant main effect a paired comparison test was not conducted.

The third dependent variable used to test hypothesis H2d was the time spent by each participant in answering what the data differences were between companies four and six by selecting choices from a template. On average those participants viewing the 2-D display were the most efficient (used the least amount of time in seconds) on this first pattern recognition task. Those participants viewing the 2-D displays (mean seconds 27) used 38% less time (in seconds) than those participants viewing the 3-D display (mean seconds 43). While those participants viewing the tabular display (mean seconds 30) used 31% less time (in seconds) than those participants viewing the 3-D perspective display (mean seconds 43).

Results of the ANCOVA show that, the manipulation of the presentation formats (Treatment) has a significant ($p < 0.001$) main effect on the efficiency of the participants in describing what the data differences were between companies four and six. A paired comparison test was conducted to determine if the participants receiving the 3-D Treatment were more efficient or used less time (in seconds) than those participants receiving the Tabular or 2-D Treatment. Results show that those participants viewing tabular display ($p <0.001$) and those participants viewing the 2-D displays ($p < 0.001$) were significantly more efficient or used less time (in seconds) than those participant viewing the 3-D perspective display in this pattern recognition task.

The fourth dependent variable used to test hypothesis H2d was the time spent by each participant in describing the pattern of financial ratios they were seeing in group two compared to group one by selecting choices from a template. On average those participants viewing the 2-D display (mean seconds 35) were the most efficient (used the least amount of time in seconds) on this pattern recognition task. Those participants

viewing the 2-D displays (mean seconds 35) used 37% less time (in seconds) than those participants viewing the 3-D display (mean seconds 55). While those participants viewing the tabular display (mean seconds 46) used 17% less time (in seconds) than those participants viewing the 3-D perspective display (mean seconds 55).

Results of the ANCOVA show that, the manipulation of the presentation formats (Treatment) has a significant ($p = 0.017$) main effect on the efficiency of the participants in describing the pattern of financial ratios they were seeing in group two compared to group one by selecting choices from a template. A paired comparison test was conducted to determine if the participants receiving the 3-D Treatment were more efficient or used less time (in seconds) than those participants receiving the Tabular or 2-D Treatment. Results show that those participants viewing the 2-D displays ($p = 0.015$) were significantly more efficient or used less time (in seconds) than those participant viewing the 3-D perspective display in this pattern recognition task. There is no significant difference in efficiency between those participants viewing the tabular display and those viewing the 3-D perspective display.

In conclusion, hypothesis H2d was not supported.

4.3 Lessons Learned from the Pilot Results and How to Improve

4.3.1 Rank Order Effect

As shown by Tables 23 the order in which the trend analysis and pattern recognition tasks were presented resulted in a significant interaction with the presentation formats (Treatment). The interaction effect between the Treatment condition and the task order suggests that participant responses depend on whether participants perform the trend analysis task first and the pattern recognition task later or vice versa. To address the issue of rank order effects, the main experiment employs a full factorial 3x2 between-subjects design.

4.3.2 Insufficient Training on Display Formats

Despite the fact that participants were required to perform two different treatment tasks (trend analysis and pattern recognition tasks), the pilot study employed the same training task to familiarize participants with their assigned display formats.

Participants viewing the 3-D perspective display scored, on average, only 2.966 and 3.034 (on a four point scale) on the two practice questions; these scores are lower than that of participants viewing either the tabular display (3.222 and 3.444) or 2-D displays (3.000 and 3.400). It seems that participants, especially those viewing the 3-D perspective display, need more training in terms of practice questions.

The main experiment employs separate training materials for the two treatment tasks: the trend analysis task and the pattern recognition task. The practice questions on display formats are increased from two to four questions, to better train the participants

on their assigned display format, which was particularly important for those receiving the 3-D perspective display.

4.3.3 Insufficient Training on ROE

The pilot study did not provide participants an opportunity to review the concepts of the DuPont analysis, or to practice calculating the return on equity. Even though participants had already learned the concepts of DuPont analysis in the principles of accounting class, all treatment groups performed poorly in the trend analysis task involving calculating ROE. The main experiment provides two practice questions to train participants on the concept of DuPont analysis and on the calculation of ROE, respectively.

4.3.4 Scrolling Up and Down the Screen

Participants in both the 2-D and 3-D treatment conditions had to scroll up and down the screen when answering questions, since the display and the response area did not fit on one screen. Debriefing with some participants suggested that participants felt constrained by having to scroll up and down the screen while answering the questions. It seems that the physical motion of scrolling up and down the screen interfered with performance, possibly because of the "split-attention effect" as suggested by Chandler and Sweller (1992). This split-attention effect likely had undesirable consequences on both the time spent and the accuracy of responses from participants viewing either the 2-D displays or 3-D perspective display (Chandler and Sweller 1992). Participants viewing the tabular display did not have such a problem.

The 2-D line graphs (see Figure 1) or 2-D bar charts (see Figure 3) and 3-D perspective displays (see Figure 2 and 4) cannot be further reduced in area without sacrificing the quality of perception. This prevents fitting the questions and response boxes into a single screen. However, the study does attempt to follow Tufte's (1983) guidelines for ensuring that the displays exhibit graphical excellence.

Following Tufte's guidelines of graphical excellence (1983), the study develops figures 1, 2, 3, and 4 to show the data, to induce the viewer to think about the substance and to encourage the viewer to compare different pieces of data. Similar to the 3-D perspective display (see Figures 2 and 4), this study places four separate 2-D line graphs (see Figure 1) and four separate 2-D bar graphs (see Figure 3) in a single page area. The total area of each of the Figures 1, 2, 3, and 4 are the same, and viewers can clearly view the scale and description of each figure.

Tufte (1983 p. 56) suggested that "different people see the same area differently; perceptions change with experience; and perceptions are context dependent." The main experiment has four practice questions to familiarize participants with their assigned display format. Though increased training cannot remove the split-attention effect, increased training can help to familiar participants with the need to scroll up and down the screen while performing the task.

Participants viewing the tabular display do not need to scroll up and down the screen. Compared to viewers of the 2-D display and the viewers of the 3-D perspective display, viewers of the tabular display have less mental workload. To measure the differences in mental workload among the three treatment groups four survey questions (Reid and Nygren 1988) measuring each participant's mental workload are added to the

main experiment. The four questions adopted from Reid and Nygren (1988) are: 1) Very little mental effort or concentration was required to complete tasks, 2) Tasks performed were almost automatic, requiring little or no attention, 3) Tasks performed were very complex and required total attention, and 4) Extensive mental effort and concentration were necessary in tasks. Measures of mental workload are used as one of the covariates in the main experiment.

Chapter 5: Main Experiment

5.1 Sample Size, Inter-Coder Reliability

5.1.1 Sample Size

To determine the necessary sample size, Faul et.al.'s (2007) G*Power 3 was used. According to the analysis two hundred sixteen participants are needed to detect a medium effect size of 0.25, with the standard α-level of 0.05 and a power of 0.80, if one-way ANOVA statistical analysis of six groups is performed. Cohen (1992) suggests that with a power of 0.80, there is an 80% chance of detecting an effect if that effect genuinely exists.

5.1.2 Inter-Coder Reliability

Two hundred fifty eight undergraduate business students participated in the main experiment. Two hypothesis-blind coders worked independently to code the responses of each participant. The same two coders, who had coded the responses from the pilot tests, were used as coders for the main experiments. When assigning score-points to the participants' responses on the practice questions, trend analysis questions, and the pattern recognition questions, the two coders were instructed to follow the grading schema as outlined in Table 5, Table 6, and Table 7, respectively. When assigning score-points to the participants' responses on the Mental Rotations Tests, the two coders were instructed to follow the grading schema as discussed in section 3.6.2. Other than the aforementioned assignment of score-points, which required some judgment, all other responses to

questions like the survey or demographic questions were automatically captured and reported by the survey software. Each of the two coders separately coded the responses and then reconciled their differences.

Subjectivity of the coder and differences in coding between the two coders should be low, as most of the questions asked were in the form of multiple choices. The survey software used in the study (Select Survey) allowed the experimenter to deploy open-ended questions such as asking participants to respond by writing short sentences or writing a newly calculated value, or to deploy multiple choices questions such as asking participants to respond by selecting choices from a given template.

Measures of the time spent (in seconds) by each participant when answering a question (practice questions, trend analysis questions, and pattern recognition questions) were calculated via a database query based on timestamps captured by the survey software.

Other than these opened end questions − TAQ1, TAQ2, TAQ3, TAQ4, PRQ1, PRQ3, PRQ4, PQ1, PQ2, PQ3, PQ4, PQ5 and PQ6− all other questions were in the form of multiple choices. Like the dependent variables of the time spent in seconds by each participant when answering a question, the survey software automatically captured the responses of all multiple-choice questions, which could be exported to an Excel spreadsheet.

Smith (2000) suggests that when a coding system yields scores, the agreement between the scores assigned by two coders can be tested through use of a correlation coefficient. Since the two blind coders of the study were simply assigning scores rather than actually coding subjective responses, the agreement between the scores assigned by

the two blind coders was tested though a simple correlation between their assigned scores. The correlation between each of the scores assigned by the coders was separately evaluated. For example, the scores assigned by each of the two coders to the TAQ1 were set up as column one and column two in the order of the treatment groups of tabular display, 2-D displays, and a 3-D perspective display to calculate the inter-coder correlation. A similar procedure was performed on all the scores of the Trend Analysis task and the Pattern Recognition task.

Smith et al. (1992) suggest that a satisfactory percentage of agreement between two coders will require an inter-coder correlation of 0.85 or more. Table 28 and 29 reports the results of the Pearson's Product-Moment Correlation Coefficients of the scores (before reconciliation) assigned by coder one and coder two to the trend analysis task and pattern recognition task, respectively. Except for practice question two of the trend analysis task, which has an inter-coder correlation of 0.799, all other inter-coder correlation statistics are well above the benchmark of 0.85 as suggested by Smith et al. (1992). Since practice question two of the trend analysis task is a covariate, rather than a manipulated variable, a 0.799 inter-coder correlation for the scores assigned to practice question two (before reconciliation) is not a major concern. It is concluded that the degree of inter-coder reliability is sufficiently high to proceed with data analysis. Starting from the next section onward, the data set used and reported in the rest of the study was after the reconciliation of differences between the two coders.

Table 28
Pearson's Product-Moment Correlation Coefficients of the Scores Assigned by Coder One and Coder Two to the Trend Analysis Task

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		PQ1	TSPQ1	PQ2	TSPQ2	PQ3	TSPQ3	PQ4
Coder 1	PQ1	0.919**						
Coder 1	TSPQ1		0.959**					
Coder 1	PQ2			0.799**				
Coder 1	TSPQ2				0.998**			
Coder 1	PQ3					0.993**		
Coder 1	TSPQ3						0.973**	
Coder 1	PQ4							0.982**

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		TSPQ4	PQ5	TSPQ5	PQ6	TSPQ6	TAQ1	TSTAQ1
Coder 1	TSPQ4	0.992**						
Coder 1	PQ5		0.952**					
Coder 1	TSPQ5			0.994**				
Coder 1	PQ6				1.000**			
Coder 1	TSPQ6					0.989**		
Coder 1	TAQ1						0.979**	
Coder 1	TSTAQ1							0.986**

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		TAQ2	TSTAQ2	TAQ3	TSTAQ3	TAQ4a	TAQ4b	TAQ4c
Coder 1	TAQ2	0.914**						
Coder 1	TSTAQ2		0.994**					
Coder 1	TAQ3			1.000**				
Coder 1	TSTAQ3				1.000**			
Coder 1	TAQ4a					1.000**		
Coder 1	TAQ4b						0.937**	
Coder 1	TAQ4c							0.987**

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		TAQ4d	TSTAQ4	TAQ5	TSTAQ5	TAQ6	TSTAQ6	S1
Coder 1	TAQ4d	1.000**						
Coder 1	TSTAQ4		0.999**					
Coder 1	TAQ5			0.991**				
Coder 1	TSTAQ5				0.997**			
Coder 1	TAQ6					0.993**		
Coder 1	TSTAQ6						0.999**	
Coder 1	S1							0.995**

See Table 11 for definition of the variables

Table 28 (Continued)
Pearson's Product-Moment Correlation Coefficients of the Scores Assigned by Coder One and Coder Two to the Trend Analysis Task

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		S2	S3	S4	S5	S6	S7	S8
Coder 1	S2	0.998**						
Coder 1	S3		1.000**					
Coder 1	S4			1.000**				
Coder 1	S5				1.000**			
Coder 1	S6					0.999**		
Coder 1	S7						1.000**	
Coder 1	S8							1.000**

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		S9	S10	M1	M2	M3	AGE	GEN
Coder 1	S9	0.974**						
Coder 1	S10		1.000**					
Coder 1	M1			1.000**				
Coder 1	M2				1.000**			
Coder 1	M3					0.979**		
Coder 1	AGE						0.975**	
Coder 1	GEN							1.000**

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		SS	GPA	SAT	FTWE	PTH	ARWE	HE
Coder 1	SS	1.000**						
Coder 1	GPA		1.000**					
Coder 1	SAT			1.000**				
Coder 1	FTWE				1.000**			
Coder 1	PTH					1.000**		
Coder 1	ARWE						1.000**	
Coder 1	HE							1.000**

		Coder 2	Coder 2
		SMRT	TSMRT
Coder 1	SMRT	0.997**	
Coder 1	TSMRT		0.990**

See Table 11 for definition of the variables

Table 29
Pearson's Product-Moment Correlation Coefficients of the Scores Assigned by Coder One and Coder Two to the Pattern Recognition Task

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		PQ1	TSPQ1	PQ2	TSPQ2	PQ3	TSPQ3	PQ4
Coder 1	PQ1	0.970**						
Coder 1	TSPQ1		1.000**					
Coder 1	PQ2			0.924**				
Coder 1	TSPQ2				1.000**			
Coder 1	PQ3					0.982**		
Coder 1	TSPQ3						1.000**	
Coder 1	PQ4							0.975**

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		TSPQ4	PQ5	TSPQ5	PQ6	TSPQ6	PRQ1	TSPRQ1
Coder 1	TSPQ4	1.000**						
Coder 1	PQ5		0.910**					
Coder 1	TSPQ5			1.000**				
Coder 1	PQ6				1.000**			
Coder 1	TSPQ6					1.000**		
Coder 1	PRQ1						0.956**	
Coder 1	TSPRQ1							1.000**

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		PRQ2	TSPRQ2	PRQ3	TSPRQ3	PRQ4	TSPRQ4	PRQ5
Coder 1	PRQ2	0.995**						
Coder 1	TSPRQ2		1.000**					
Coder 1	PRQ3			0.993**				
Coder 1	TSPRQ3				1.000**			
Coder 1	PRQ4					1.000**		
Coder 1	TSPRQ4						1.000**	
Coder 1	PRQ5							0.985**

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		TSPRQ5	PRQ6	TSPRQ6	S1	S2	S3	S4
Coder 1	TSPRQ5	1.000**						
Coder 1	PRQ6		1.000**					
Coder 1	TSPRQ6			0.975**				
Coder 1	S1				0.979**			
Coder 1	S2					1.000**		
Coder 1	S3						0.974**	
Coder 1	S4							0.999**

See Table 11 for definition of the variables

Table 29 (Continued)
Pearson's Product-Moment Correlation Coefficients of the Scores Assigned by Coder One and Coder Two to the Pattern Recognition Task

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		S5	S6	S7	S8	S9	S10	M1
Coder 1	S5	1.000**						
Coder 1	S6		0.895**					
Coder 1	S7			0.995**				
Coder 1	S8				0.978**			
Coder 1	S9					1.000**		
Coder 1	S10						1.000**	
Coder 1	M1							1.000**

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2	Coder 2
		M2	M3	AGE	GEN	SS	GPA	SAT
Coder 1	M2	1.000**						
Coder 1	M3		1.000**					
Coder 1	AGE			0.873**				
Coder 1	GEN				1.000**			
Coder 1	SS					1.000**		
Coder 1	GPA						1.000**	
Coder 1	SAT							0.959**

		Coder 2	Coder 2	Coder 2	Coder2	Coder 2	Coder 2
		FTWE	PTH	ARWE	HE	SMRT	TSMRT
Coder 1	FTWE	0.975**					
Coder 1	PTH		0.998**				
Coder 1	ARWE			1.000**			
Coder 1	HE				1.000**		
Coder 1	SMRT					0.993**	
Coder 1	TSMRT						1.000**

See Table 11 for definition of the variables

5.2 Testing of Raw Data

The study employs a 3 x 2 between-subjects (display format x task) design. The independent variables are display type and task type. Display type was manipulated at three levels: no graphical display (table only), 2-D displays, and 3-D perspective display (see Tables 1-2, and Figures 1-3 for examples of display formats used). Task type was manipulated at two levels: trend analysis and pattern recognition task. The remainder of this section discusses the tests of statistical assumptions.

5.2.1 Testing of Outliers and Influential Observations

Field (2005) suggests that both outliers and influential observations have to be considered simultaneously when determining whether to drop an observation that meets the criteria of being either an outlier or an influential observation. For example, an observation can have a small standardized residual (not an outlier), but can be very influential (Field 2005). Field (2005) suggests that if an observation's standardized residual has an absolute value greater than 3, the observation is an outlier. Cook and Weisberg (1982) suggest that if an observation has a Cook's distance greater than 1, the observation is an influential observation. Tests for influential observations and outliers, at one standard deviation, were performed on each dependent variable (see Tables 3 and 4). Results suggest that one observation for the dependent variable TAQ3, "Based on the ROE of year 5 what would be year 6 ROE if each of the variables of ROE in year 5 had doubled," has a Cook's distance of 1.243, and a standardized residual of -10.93413. This observation, within the 3-D perspective treatment group, is dropped from the statistical analysis. Hypothesis H1c has five dependent measures of accuracy − the score on the

third question (TAQ3), the scores on parts a, b, c and d of the fourth question (TAQ4a, b, c, d). The aforementioned influential outlier was one of the observations for the dependent variable TAQ3 within the 3-D perspective treatment group, and therefore affects hypothesis H1c. Two separate MANCOVA models were run with and without the influential outlier. Comparison between the multivariate test results of the two different MANCOVA models showed that the exclusion of the influential outlier improved the significance of the manipulation of the presentation formats, as revealed in the significance level of the Pillai's Trace going from 0.025 to 0.005. By dropping the outlier hypothesis H1c had four significant (alpha<0.10) differences (instead of two) in terms of accuracy between the participants viewing the 2-D displays and participants viewing the tabular display or 3-D perspective display. The results of hypothesis H1c are reported without the outlier (see section 5.3.5). Other than the dropped observation, there is no other observation which is either an outlier or an influential observation.

5.2.2 Testing of Assumptions

Kolmogorov-Smirnov and Shapiro-Wilk tests were performed on each dependent variable (see Table 3 and 4) to test the assumption of normality. Field (2005) suggests that if the Kolmogrov-Smirnov test or the Shapiro-Wilk test is significant ($p < 0.05$), then the distribution of the sample data is significantly different from a normal distribution. Table 30 shows that all the dependent variables are significantly different from a normal distribution.

Assumptions of the univariate ANOVA analysis include: 1) dependent variable is normally distributed within each group, 2) the variances in each group are roughly equal,

3) each observation should be independent and 4) each dependent variable should be on an interval scale (Field, 2005). Table 30 shows that the dependent variables are non-normal. Levene's tests were performed on each of the dependent variables (see Table 3 and 4) to test the assumption of homogeneity of variances. Field (2005) suggests that if Levene's test is significant at $p <= 0.05$, the assumption of homogeneity of variance has been violated. Table 30 shows that about half of the dependent variables violated the assumptions of homogeneity of variance. Field (2005) suggests that ANOVA is fairly robust in terms of violations of the assumption of normality and homogeneity of variance when the cell sizes are relatively equal. This study has relatively equal cell sizes for both the Trend Analysis Task (the treatment group of tabular display has 42 observations, 2-D displays has 40 observations, and 3-D perspective display has 42 observations), and the Pattern Recognition Task (the treatment group of tabular display has 43 observations, 2-D displays has 41 observations, and 3-D perspective display has 43 observations). Therefore, no transformation of data is necessary. In regard to the third assumption of the univariate ANOVA analysis, each observation of the study is statistically independent. In regard to the fourth assumption of the univariate ANOVA analysis, all dependent variables of the study are on an interval scale (see Tables 6 and Table 7).

Table 30
Results of Testing For Normality, and Testing For Homogeneity of Variance

Dependent Variables	Results of Testing For Normality	Kolmogorov -Smirnov Test	Shapiro-Wilk Test	Result of Testing For Homogeneity of Variances	Levene's Tesr Based on Mean
TAQ1	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P < 0.001
TSTQ1	Non-Normal Distribution	P < 0.001	P < 0.001	Homogeneity of Variance	P = 0.273
TAQ2	Non-Normal Distribution	P < 0.001	P < 0.001	Homogeneity of Variance	P = 0.945
TSTAQ2	Non-Normal Distribution	P < 0.001	P < 0.001	Homogeneity of Variance	P = 0.124
TAQ3	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P = 0.033
TSTAQ3	Non-Normal Distribution	P < 0.001	P < 0.001	Homogeneity of Variance	P = 0.946
TAQ4a	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P = 0.006
TAQ4b	Non-Normal Distribution	P < 0.001	P < 0.001	Homogeneity of Variance	P = 0.157
TAQ4c	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P < 0.001
TAQ4d	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P < 0.001
TSTAQ4	Non-Normal Distribution	P = 0.001	P < 0.001	Homogeneity of Variance	P = 0.317
TAQ5	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P = 0.050
TSTAQ5	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P = 0.007

See Table 11 for definition of the variables

Table 30 (Continued)
Results of Testing For Normality, and Testing For Homogeneity of Variance

Dependent Variables	Results of Testing For Normality	Kolmogorov -Smirnov Test	Shapiro-Wilk Test	Result of Testing For Homogeneity of Variances	Levene's Tesr Based on Mean
TAQ6	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P < 0.001
TSTAQ6	Non-Normal Distribution	P < 0.001	P < 0.001	Homogeneity of Variance	P = 0.239
PRQ1	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P = 0.007
TSPRQ1	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P = 0.011
PRQ2	Non-Normal Distribution	P < 0.001	P < 0.001	Homogeneity of Variance	P = 0.051
TSPRQ2	Non-Normal Distribution	P < 0.001	P < 0.001	Homogeneity of Variance	P = 0.062
PRQ3	Non-Normal Distribution	P < 0.001	P < 0.001	Homogeneity of Variance	P = 0.057
TSPRQ3	Non-Normal Distribution	P = 0.016	P < 0.001	Heterogeneity of Variance	P = 0.035
PRQ4	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P < 0.001
TSPRQ4	Non-Normal Distribution	P < 0.001	P < 0.001	Homogeneity of Variance	P = 0.635
PRQ5	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P = 0.003
TSPRQ5	Non-Normal Distribution	P < 0.001	P < 0.001	Heterogeneity of Variance	P < 0.001
PRQ6	Non-Normal Distribution	P < 0.001	P < 0.001	Homogeneity of Variance	P = 0.154
TSPRQ6	Non-Normal Distribution	P < 0.001	P < 0.001	Homogeneity of Variance	P = 0.218

See Table 11 for definition of the variables

MANOVA has similar assumptions to ANOVA but MANOVA further assumes that dependent variables have multivariate normality within groups and that the correlation between any two dependent variables is the same in all groups (homogeneity of covariance matrix). Field (2005) suggests that instead of checking the assumption of multivariate normality, it is more practical and useful to check the assumption of univariate normality for each dependent variable. Field (2005) stipulates that univariate normality is a necessary condition for multivariate normality, though univariate normality does not guarantee multivariate normality. Table30 shows that all the dependent variables are non-normally distributed within each group. Therefore, multivariate normality does not exist in this study's data set. However, test statistics commonly used in MANOVA analysis−the Pillai-Bartlett trace, the Hotelling-Lawley trace, and the Wilks's lambda−are robust to violations of multivariate normality (Field 2005).

Although Box's test can be used to compare the variance-covariance matrices between treatment groups, Box's test is very unstable especially when the assumption of multivariate normality is not tenable. Bray and Maxwell (1985) found that when cell sizes are equal, the Pillai-Bartlett trace is the most robust to potential violations of the homogeneity assumption. Field (2005) suggests that the Pillai-Bartlett trace and Hotelling-Lawley trace are robust to violations of homogeneity when sample sizes are equal. This study has relatively equal cell sizes for both the Trend Analysis Task, and the Pattern Recognition Task. Therefore, no transformation of data is necessary since the Pillai-Bartlett trace and Hotelling-Lawley trace are used in this study.

5.2.3 Plan of Statistical Analysis

This study employs MANCOVA analysis to test hypotheses. By including all dependent variables in a single analysis, MANCOVA takes into account the relationships among dependent variables. If the MANCOVA model is significant, then univariate ANCOVA will be used to separately test each dependent variable of a hypothesis. Rather than just putting the theoretical or expected covariates (see section 3.60) into the MANCOVA analysis, a separate regression will be run for each dependent variable with all the theoretical covariates and demographic variables (see section 3.65) to test for significance of all possible covariates to be included in subsequent MANCOVA analyses. Insignificant covariates in MANCOVA or ANCOVA models are dropped from the final analysis.

5.2.4 Cronbach's Alpha

This study used the Reid and Nygren (1988) scale to measure a participant's perceived mental workload. The Reid and Nygren (1988) scale comprises the following four statements: (1) very little mental effort or concentration was required to complete tasks, (2) tasks performed were almost automatic, requiring little or no attention, (3) tasks were very complex and required total attention, and (4) extensive mental effort and concentration was necessary in tasks. Participants were asked to select a scale number, from one (strongly disagree) to seven (strongly agree), indicating the extent to which they agreed with each of the four statements. After reverse coding the first two statements, a reliability analysis was performed to confirm whether the four-statement scale is a dependable measure of the mental workload construct. The reliability test shows a

Cronbach's Alpha of 0.849, which is higher than the threshold of 0.70 for reliability per Nunnally (1978). With a high Cronbach's Alpha of 0.849, a single measure of a participant's mental workload was calculated by averaging a participant's responses to the four statements (after reverse coding the first two statements). This average reflects a participant's mental workload and is used as a covariate measure in the multivariate and univariate models.

5.3 Manipulation Check

As discussed in section 3.90 different manipulation questions were designed for the treatment conditions of display format and task. Another manipulation question was designed to test participants' post experiment knowledge of DuPont analysis. Participants were asked to answer three manipulation questions by selecting the choices of true or false. The pilot study used the same manipulation questions as described in following paragraphs.

To test the manipulation of display formats, participants viewing the tabular display were asked whether: "The tables that you see in this experiment have data points of zero (answer is false)."Participants viewing the 2-D line displays were asked whether: "The line graphs that you see in this experiment have their axis started from zero (answer is false)." If participants viewed the 2-D bar displays they were asked whether: "The bar charts that you see in this experiment have their axis started from zero (answer is true)." Finally participants viewing the 3-D display were asked whether: "The graphs that you see in this experiment have their axis started from zero (answer is false)" (See Table 10).

To test the manipulation of tasks, participants performing the trend analysis task were asked to respond to the statement: "Within the context of this experiment companies can have negative ROE (answer is true)." Participants performing the pattern analysis task were asked to respond to the statement: "Within the context of this experiment, companies with the same ROE also have the same turnover, profitability, and leverage ratios (answer is false)" (See Table 10). To test participants' post experiment knowledge of DuPont analysis, all participants were asked whether: "Return on equity is the sum of turnover, profitability, and leverage ratios (answer is false)" (See Table 10).

5.3.1 Results of the Manipulation Check Questions

Table 31 shows that participants in the trend analysis task did not perform well when answering the manipulation check questions. Compared to those participants viewing the 2-D displays and the 3-D perspective display, those participants viewing the tabular display had higher accuracy when answering the manipulation check questions. Out of 42 participants viewing the tabular display, seven participants wrongly answered one manipulation check question. Out of 40 participants viewing the 2-D displays, 19 participants wrongly answered one manipulation check question, seven participants wrongly answered two manipulation check questions, and one participant wrongly answered all three manipulation questions. Out of the 42 participants viewing the 3-D perspective display, 16 participants wrongly answered one manipulation check question, seven participants wrongly answered two manipulation check questions, and one participant wrongly answered all three manipulation questions.

Table 32 shows that participants in the pattern recognition task also did not perform well when answering the manipulation check questions. Compared to those participants viewing the 2-D displays and the 3-D perspective display, those participants viewing the tabular display had higher accuracy when answering the manipulation check questions. Out of 43 participants viewing the tabular displays, seven participants wrongly answered one manipulation check question, and one participant wrongly answered two manipulation questions. Out of 41 participants viewing the 2-D display, 13 participants wrongly answered one manipulation check question, four participants wrongly answered two manipulation check questions, and one participant wrongly answered all three manipulation questions. Out of the 43 participants viewing the 3-D perspective display, 15 participants wrongly answered one manipulation check question, and two participants wrongly answered two manipulation check questions.

It should be noted that the pilot study had a lower error rate in responses to the manipulation check questions. However, the pilot study had a smaller sample size.

The reason for such a high error rate in responses to the manipulation check questions was due to the fact that participants had to recall from memory the display when answering the manipulation check. Participants could not look at the tables or graphical displays when answering the manipulation check questions. Another problem is that the manipulation check questions were written to assess participants' level of understanding of the characteristics of the display format they were presented, rather than merely assessing their recognition of the display format they were given. As a result, the manipulation check questions did not operationalize as intended. However, the study had significant results partially supporting the hypotheses.

For the trend analysis task, one participant viewing the 2-D displays and one participant viewing the 3-D perspective display answered all three manipulation check questions incorrectly. The results of hypothesis H1 a-d did not change after dropping these two participants. For the pattern recognition task, one participant viewing the 2-D displays answered all three manipulation check questions incorrectly. The results of hypothesis 2a-d did not change after dropping this participant. Nevertheless, the study will report the high error rate of the response to the manipulation check questions as a limitation.

Table 31

Results of the Manipulation Check Questions of the Trend Analysis Task

Trend analysis Task	**Tabular Display**	**2-D Displays**	**3-D Display**
No of Participants	(n=42)	(n=40)	(n=42)
Answered one manipulation question wrongly	n=7	n=19	n=16
Answered two manipulation questions wrongly	n=0	n=7	n=7
Answered three manipulation questions wrongly	n=0	n=1	n=1
M1 answered wrongly	n=1	n=14	n=11
M2 answered wrongly	n=2	n=7	n=6
M3 answered wrongly	n=4	n=15	n=8

M1 = The tables (graphs) you see in the experiment, have data points of zero
M2 = Return on Equity is the SUM of turnover, profitability, and leverage ratios
M3 = Within the context of this experiment, companies can have negative ROE

Table 32

Results of the Manipulation Check Questions of the Pattern Recognition Task

Trend analysis Task	**Tabular Display**	**2-D Displays**	**3-D Display**
No of Participants	(n=43)	(n=41)	(n=43)
Answered one manipulation question wrongly	n=7	n=13	n=15
Answered two manipulation questions wrongly	n=1	n=4	n=2
Answered three manipulation questions wrongly	n=0	n=1	n=0
M1 answered wrongly	n=4	n=13	n=13
M2 answered wrongly	n=0	n=4	n=4
M3 answered Wrongly	n=5	n=7	n=2

M1 = The tables (graphs) you see in the experiment, have data points of zero
M2 = Return on Equity is the SUM of turnover, profitability, and leverage ratios
M3 = Within the context of this experiment, companies with the same ROE also have the same turnover, profitability, and leverage ratios

5.4 Results of the Trend Analysis Task

The purpose of this section is to report whether there is support for each of the hypotheses H1a, H1b, H1c and H1d. Hypotheses H1a-H1d posit that 2-D displays are more effective and efficient than tabular and 3-D displays for a trend analysis task. This section is organized in the following way: descriptive statistics of the variables, the Pearson's Product-Moment Correlation of the variables, results of the regression analyses of possible covariates, and the MANCOVA and ANCOVA results of each of the hypotheses.

5.4.1 Descriptive Statistics

Table 33 shows the descriptive statistics for the trend analysis task. Based on the results of one-way ANOVA the following paragraphs briefly describe those covariates and demographic variables that had significant differences in their means among treatment groups. Descriptive information for the dependent variables will be discussed in the results section for each hypothesis.

The first covariate evaluated was performance on the practice questions. For the score on practice question one (PQ1) (What are the values of the factors of the apartment rented in year 2?), participants viewing the 2-D displays on average had the highest score. Those participants viewing the 2-D displays (mean score 3.375) were significantly more accurate ($p = 0.018$) than those participants viewing the 3-D perspective display (mean score 2.619). There was no significant difference in accuracy between those participants viewing the tabular display (mean score 3.238) and those participants viewing the 2-D

displays (mean score 3.375) or those participants viewing the 3-D perspective display (mean score 2.619).

For the score on practice question two (PQ2) (What are the differences between the apartments rented in year 5 and 6?), participants viewing the 2-D displays on average had the highest score. Those participants viewing the 2-D displays (mean score 2.825) were significantly more accurate ($p = 0.001$) than those participants viewing the 3-D perspective display (mean score 2.214). Those participants viewing the tabular display (mean score 2.714) were significantly more accurate ($p = 0.005$) than those participants viewing the 3-D perspective display (mean score 2.214). There was no significant difference in accuracy between those participants viewing the tabular display (mean score 2.714) and those participants viewing the 2-D display (mean score 2.825).

For the score on practice question four (PQ4) (What are the differences between the apartments rented in year 2 and 4?), participants viewing the 2-D displays on average had the highest score. Those participants viewing the 2-D displays (mean score 3.825) were significantly more accurate ($p = 0.001$) than those participants viewing the 3-D perspective display (mean score 3.023). Those participants viewing the tabular display (mean score 3.619) were significantly more accurate ($p = 0.007$) than those participants viewing the 3-D perspective display (mean score 3.023). There was no significant difference in accuracy between those participants viewing the tabular display (mean score 3.619) and those participants viewing the 2-D displays (mean score 3.825).

The second covariate evaluated was the time spent on the practice questions. For the time spent by each participant in answering practice question two (TSPRQ2), those participants viewing the tabular display were the most efficient. Those participants

viewing the tabular display (mean seconds 101) were significantly more efficient (p=0.006) or used less time (in seconds) than those participants viewing the 3-D perspective display (mean seconds 152). Those participants viewing the 2-D displays (mean seconds 112) were significantly more efficient (p=0.052) or used less time (in seconds) than those participants viewing the 3-D perspective display (mean seconds 152). There was no significant difference in efficiency between those participants viewing the tabular display (mean seconds 101) and those participants viewing the 2-D displays (mean seconds 112).

For the time spent by each participant in answering practice question three (TSPRQ3), those participants viewing the tabular display were the most efficient. Those participants viewing the tabular display (mean seconds 60) were significantly more efficient (p = 0.006) or used less time (in seconds) than those participants viewing the 3-D perspective display (mean seconds 83). There was no significant difference in efficiency between those participants viewing the 2-D displays (mean seconds 76) and those participants viewing the 3-D perspective display (mean seconds 83). There was also no significant difference in efficiency between those participants viewing the tabular display (mean seconds 60) and those participants viewing the 2-D displays (mean seconds 76).

The next covariate evaluated was time spent on the Mental Rotations Test (MRT). On average those participants viewing the 3-D perspective display were the most efficient. Those participants viewing the 3-D perspective display (mean seconds 515) were significantly more efficient (p = 0.022) or used less time (in seconds) than those participants viewing the tabular display (mean seconds 679). There was no significant

difference in efficiency between those participants viewing the 2-D displays (mean seconds 621) and those participants viewing the 3-D perspective display (mean seconds 515). There was also no significant difference in efficiency between those participants viewing the tabular display (mean seconds 679) and those participants viewing the 2-D displays (mean seconds 621).

The mean MRT scores between tabular, 2-D and 3-D treatment groups were 19, 18, and 17 points, respectively. T-tests indicate no significant difference in spatial ability among treatment groups. The next covariate evaluated was perceived mental workload (MW). As predicted in section 3.6.4, participants viewing the 3-D perspective display had the highest perceived mental workload (MW) (mean 4.803) when compared to participants viewing the 2-D displays (mean 4.343) or participants viewing the tabular display (mean 4.428). There is no significant difference in mental workload among treatment groups.

Table 33
Descriptive Statistics for the Trend Analysis Task

Panel A: Mean (Standard Deviation) and Range of Practice Questions.

	Tabular Display (n=42)	**2-D Displays (n=40)**	**3-D Display (n=42)**
Practice Question One: What are the values of the factors of apartment rented in year 2?	3.238 (1.303) 0.000 to 4.000	3.375 (1.212) 0.000 to 4.000	2.619 (1.146) 0.000 to 4.000
Time (seconds) Spent on Practice Question one?	85 (50) 31 to 263	100 (35) 35 to 215	105 (41) 35 to 217
Practice Question Two: What are the differences between the apartments rented in years 5 and 6?	2.714 (0.596) 1.000 to 3.000	2.825 (0.594) 0.000 to 3.000	2.214 (0.898) 0.000 to 3.000
Time (seconds) Spent on Practice Question Two	101 (37) 50 to 197	112 (46) 29 to 204	152 (112) 29 to 704
Practice Question Three What are the values of the factors of apartment rented in year 4?	3.238 (1.284) 0.000 to 4.000	3.400 (1.215) 0.000 to 4.000	2.833 (1.286) 0.000 to 4.000
Time (seconds) Spent on Practice Question Three	60 (29) 21 to 164	76 (29) 19 to 141	83 (37) 10 to 176
Practice Question Four: What are the differences between the apartments rented in year 2 and 4?	3.619 (0.935) 0.000 to 4.000	3.825 (0.446) 2.000 to 4.000	3.023 (1.092) 0.000 to 4.000
Time (seconds) Spent on Practice Question Four	98 (48) 29 to 301	95 (45) 19 to 222	112 (63) 10 to 353

See Table 11 for definition of the variables.

Table 33 (Continued)
Descriptive Statistics for the Trend Analysis Task

Panel A: Mean (Standard Deviation) and Range of Practice Questions

	Tabular Display (n=42)	**2-D Displays (n=40)**	**3-D Display (n=42)**
Practice Question Five What is a company's ROE if its turnover ratio is 2, profitability ratio is 5%, and leverage ratio is 1.1?	0.500 (0.506) 0.000 to 1.000	0.500 (0.506) 0.000 to 1.000	0.476 (0.505) 0.000 to 1.000
Time (seconds) Spent on Practice Question Five	142 (61) 52 to 340	124 (57) 54 to 389	151 (79) 13 to 405
Practice Question Six ROE is the sum (multiplicative) of turnover, profitability, and leverage?	0.976 (0.154) 0.000 to 1.000	0.925 (0.266) 0.000 to 1.000	0.928 (0.260) 0.000 to 1.000
Time (seconds) Spent on Practice Question Six	30 (15) 10 to 80	28 (12) 8 to 59	30 (12) 5 to 62

Panel B: Mean (Standard Deviation) and Range of Dependent Variables

	Tabular Display (n=42)	**2-D Displays (n=40)**	**3-D Display (n=42)**
Question One What are the differences between years 1 and 4?	3.452 (1.130) 0.000 to 4.000	3.650 (0.735) 1.000 to 4.000	2.595 (1.269) 0.000 to 4.000
Time (seconds) Spent on Question One	92 (46) 26 to 236	96 (43) 13 to 211	121 (58) 8 to 331
Question Two What is happening as you go from year 2 to year 3 to year 4?	5.476 (2.370) 2.000 to 8.000	5.925 (2.758) 0.000 to 8.000	4.690 (2.503) 0.000 to 8.000
Time (seconds) Spent on Question Two	136 (114) 40 to 762	127 (56) 32 to 247	152 (90) 7 to 342

See Table 11 for definition of the variables.

Table 33 (Continued)
Descriptive Statistics for the Trend Analysis Task

Panel B: Mean (Standard Deviation) and Range of Dependent Variables

	Tabular Display (n=42)	**2-D Displays (n=40)**	**3-D Display (n=42)**
Question Three Based on the ROE of year 5, what would be year 6 ROE if each of the variables of ROE in year 5 had doubled?	71.242 (53.015) -143.660 to 127.140	87.110 (35.917) -6.660 to 128.340	63.876 (56.813) -173.660 to 120.340 (n=41)
Time (seconds) Spent on Question Three	109 (80) 27 to 397	98 (69) 25 to 303	94 (80) 23 to 453
Question Four Part a Estimate the average of turnover for the years 1, 2, 4, and 5	-5.001 (22.339) -104.455 to 0.045	-1.503 (4.393) -20.295 to 1.045	-0.852 (3.508) -19.955 to 0.295
Question Four Part b Estimate the average of profitability for the years 1, 2, 4, and 5	0.796 (1.334) -1.045 to 3.955	0.626 (4.563) -20.045 to 13.955	-1.494 (8.845) -56.245 to 3.255
Question Four Part c Estimate the average of leverage for the years 1, 2, 4, and 5	-0.032 (0.860) -5.280 to 0.720	-1.898 (5.954) -35.280 to 2.720	-3.092 (1.196) -6.280 to 0.620
Question Four Part d Estimate the average of turnover, profitability and leverage for the years 1, 2, 4, and 5, and use them to calculate a new ROE	0.359 (12.605) -75.159 to 11.908	-16.392 (57.060) -325.759 to 18.241	-0.970 (6.957) -32.629 to 9.441
Time (seconds) Spent on Question Four	221 (135) 32 to 581	218 (136) 21 to 551	210 (107) 15 to 530

See Table 11 for definition of the variables.

Table 33 (Continued)
Descriptive Statistics for the Trend Analysis Task

Panel B: Mean (Standard Deviation) and Range of Dependent Variables

	Tabular Display (n=42)	**2-D Displays (n=40)**	**3-D Display (n=42)**
Question Five What are the differences between years 2 and 4? Please answer as accurately and as fast as possible.	3.761 (0.691) 1.000 to 4.000	3.850 (0.483) 2.000 to 4.000	3.642 (0.692) 1.000 to 4.000
Time (seconds) Spent on Question Five	35 (10) 8 to 59	35 (7) 23 to 55	48 (17) 8 to 113
Question Six What is happening as you go from year 1 to year 2 to year 3? Please answer as accurately and as fast as possible.	7.547 (1.040) 3.000 to 8.000	7.625 (1.147) 3.000 to 8.000	6.357 (1.818) 3.000 to 8.000
Time (seconds) Spent on Question Six	65 (40) 12 to 280	58 (29) 13 to 174	81 (43) 7 to 274

Panel C: Mean (Standard Deviation) and Range of Mental Rotations Test and Mental Workload

	Tabular Display (n=42)	**2-D Displays (n=40)**	**3-D Display (n=42)**
Score on Mental Rotations Test	19.880 (11.536) 2.000 to 40.000	18.375 (10.902) 2.000 to 40.000	17.666 (10.309) 2.000 to 38.000
Time (seconds) Spent on Mental Rotation Test	679 (329) 125 to 1645	621 (282) 199 to 1682	515 (201) 40 to 1027
Mental Workload	4.428 (1.3584) 1.750 to 7.000	4.343 (1.3571) 2.000 to 7.000	4.803 (1.3853) 2.000 to 7.000

See Table 11 for definition of the variables

Table 33 (Continued)
Descriptive Statistics for the Trend Analysis Task

Panel D: Demographic Mean and Range

	Tabular Display (n=42)	**2-D Displays (n=40)**	**3-D Display (n=42)**
AGE 18-22	n=32	n=29	n=30
23-27	n=7	n=7	n=10
29-32	n=3	n=2	n=1
33-37		n=1	-
38-42		n=1	-
43-47			n=1
48-50			
Male/Female	n=24/n=18	n=17/n =23	n=15/n=27
Student Status			
Freshman	-	-	-
Sophomore	n=14	n=8	n=12
Junior	n=20	n=19	n=17
Senior	n=8	n=13	n=13
GPA	3.110 (0.682) 2.300-3.960	3.108 (0.431) 2.000-3.880	3.214 (0.379) 2.300-3.800
SAT	1156 (123) (n=33) 950-1410	1135 (151) (n=32) 790-1430	1014 (402) (n=37) 795-1575
Years of Full Time Working Experience	2.142 (3.227) 0.000-18.000	1.912 (3.123) 0.000-14.000	1.428 (3.742) 0.000-23.000
Part Time Working Hours Per Week	12.938 (13.044) 0.000-36.000	15.750 (13.110) 0.000-45.000	19.142 (13.377) 0.000-45.000
Accounting Related Working Experience			
Full Time	n=35	n=28	n=28
Part Time	n=6	n=7	n=10
Both Full/Part Time	n=1	n=3	n=3
None	-	n=2	n=1
Highest Education			
High School	n=41	n=26	n=37
Bachelor's degree	n=1	n=3	n=5
Master degree	-	n=1	-

5.4.2 Pearson's Product-Moment Correlation

5.4.2.1 Correlation between dependent variables. Table 34 shows the Pearson's product-moment correlation coefficients and the significances for the variables of the trend analysis task. Results of two-tailed testing of the significance of the correlation between variables at alpha levels of 0.05 and 0.01 are also shown in Table 34. The following paragraphs discuss in detail the most important correlation results, in terms of whether the dependent variables for each hypothesis are significantly correlated with one another. When reporting significant correlation between two dependent variables, the following paragraphs use the symbols of the dependent variables (see Table 11) to describe each of them and report their correlation coefficient through the notation of " r ".

To test for the trend analysis task (H1a, H1b, H1c, and H1d), six questions are developed (see Table 3) and each participant answers these six questions in the same order. The first, second, fifth, and sixth questions are used to test hypotheses H1a and H1b (see Table3).

Hypothesis H1a has four dependent measures of accuracy –TAQ1, TAQ2, TAQ5, and TAQ6 (see Table 6). TAQ1 is positively related to TAQ2 (r = 0.547), TAQ5 (r = 0.277), and TAQ6 (r = 0.335). TAQ2 is positively related to TAQ5 (r = 0.190) and TAQ6 (r = 0.340). TAQ5 is positively related to TAQ6 (r = 0.467). The preceding discussion highlights the fact that the four dependent measures for hypothesis H1a are inter-correlated and MANCOVA is the appropriate statistical method to test hypothesis H1a.

Hypothesis H1b has four dependent measures of efficiency – TSTAQ1, TSTAQ2, TSTAQ5, and TSTAQ6 (see Table 6). TSTAQ1 is positively related to TSTAQ2 (r = 0.450), TSTAQ5 (r = 0.362), and TSTAQ6 (r = 0.481). TSTAQ2 is positively related to

TSTAQ6 (r = 0.256). However, there is no significant correlation between TSTAQ2 and TSTAQ5. TSTAQ5 is positively related to the time spent, in seconds, by each participant when answering the sixth question TSTAQ6 (r = 0.579). The preceding discussion highlights the fact that the four dependent measures of hypothesis H1b are inter-correlated and MANCOVA is the appropriate statistical method to test hypothesis H1b.

The third and fourth questions are used to test hypotheses H1c and H1d (see Table 3). Hypothesis H1c has five dependent measures of accuracy – TAQ3, TAQ4a, TAQ4b, TAQ4c, and TAQ4d (see Table 6). TAQ3 is positively related to TAQ4a (r = 0.256), while TAQ4b is positively related to TAQ4c (r = 0.328) (see Table 34). It seems that the five dependent measures of hypothesis H1c were not highly inter-correlated; therefore MANCOVA is not necessary to test hypothesis H1c. However, for consistency and to avoid any potential for over stating the results of hypothesis H1c the study first employed MANCOVA to test hypothesis H1c (see discussion in results section).

Hypothesis H1d has two dependent measures of efficiency – TSTAQ3 and TSTAQ4 (see Table 6). TSTAQ3 is positively related to TSTAQ4 (r = 0.316), see Table 34. Since the two dependent measures of hypothesis H1d were correlated, MANCOVA is the appropriate statistical method to test hypothesis H1d.

5.4.2.2 Multicollinearity. According to Field (2005) when multicollinearity exists, strong correlation between two predictors, in a regression model it is difficult to assess the individual importance of a predictor. Field (2005) further suggests that one way of identifying multicollinearity is to scan a correlation matrix of all the predictor variables and see if any correlate above 0.80. Table 34 shows that age of the participants (AGE) and full time working experience of the participants (FTWE) are highly correlated at 0.825. However, these two variables, AGE and FTWE were not used simultaneously in any of the final MANCOVA analyses reported in the results section. The problem of multicollinearity is not significant in this study.

5.4.2.3. Correlation between treatment or covariates and dependent variables. Table 34 also shows information about the correlations between the treatment variables and the dependent variables, and between the covariates and demographic variables and the dependent variables. As before, significance is shown at alpha levels of 0.05 and 0.01. All significance levels are based on two-tailed tests. A review of the table indicates that there are some significant correlations between the treatment variable and dependent variables. Additionally, several of the dependent measures are correlated with covariates and demographic variables.

Field (2005) comments that caution must be taken when interpreting correlation coefficients because the correlation coefficient says nothing about which variable causes the other to change. However, the correlation matrix does provide preliminary evidence that the treatment may be associated with some of the dependent variables. There is also preliminary evidence that some covariate and demographic variables may be important controls in the MANOVA and ANOVA models. Therefore, the study used regression

analysis for each dependent variable to test the significance of the association between the dependent variables and all of the covariates and demographic variables.

Table 34
Pearson's Product-Moment Correlation Coefficients of the Variables of the Trend Analysis Task.

	TREATMENT	PQ1	TSPQ1	PQ2	TSPQ2
TREATMENT	1.000	-0.203*	0.191*	-0.273*	0.274**
PQ1	-0.203*	1.000	-0.072	0.460**	0.089
TSPQ1	0.191*	-0.072	1.000	0.019	0.327**
PQ2	-0.273**	0.460**	0.019	1.000	0.163
TSPQ2	0.274**	0.089	0.327**	0.163	1.000
PQ3	-1.310	0.688**	-0.016	0.439**	0.150
TSPQ3	0.277**	0.042	0.664**	0.214	0.389**
PQ4	-0.264**	0.518**	0.041	0.648**	0.045
TSPQ4	0.112	0.056	0.307**	0.220*	0.754**
PQ5	-0.020	0.162	0.098	0.205*	0.133
TSPQ5	0.053	-0.064	0.350**	0.104	0.472**
PQ6	-0.085	0.154	0.013	0.142	0.170
TSPQ6	0.006	0.089	0.260 *	0.228	0.372**
TAQ1	-0.305**	0.412**	0.100	0.591**	0.052
TSTAQ1	0.234**	0.085	0.486**	0.128	0.618**
TAQ2	-0.126	0.489**	0.120	0.382**	0.098
TSTAQ2	0.073	0.221*	0.282**	0.167	0.332**
TAQ3	-0.059	-0.062	-0.300	-0.045	-0.106
TSTAQ3	-0.079	-0.057	0.197*	0.189*	0.195*
TAQ4a	0.128	-0.177*	0.052	0.072	0.072
TAQ4b	-0.161	0.148	-0.200	0.245**	0.074
TAQ4c	0.032	-0.086	0.130	0.165	0.150
TAQ4d	-0.016	0.086	0.004	-0.013	0.063
TSTAQ4	-0.037	0.156	0.189*	0.293**	0.251**
TAQ5	-0.078	-0.305**	0.138	0.221*	0.161
TSTAQ5	0.395**	0.371**	0.264**	-0.239**	0.441**
TAQ6	-0.330**	-0.089	0.025	0.374**	0.159
TSTAQ6	0.167	-0.020	0.374**	0.037	0.562**
AGE	0.049	0.040	-0.004	0.029	0.349**
GEN	0.177*	-0.007	0.008	-0.032	0.183*
SS	0.066	-0.055	0.165	0.051	0.218*
GPA	0.027	0.188	0.061	0.112	0.177
SAT	0.063	0.012	-0.229*	0.051	-0.197*
FTWE	-0.088	-0.105	-0.136	0.112	0.297*
PTH	0.193	0.024	-0.111	-0.036	0.020
ARWE	0.154	0.046	-0.118	0.096	-0.012
HE	0.126	0.196*	0.046	-0.013	0.204
SMRT	-0.084	0.104	-0.609	0.161	-0.102
TSMRT	-0.240**	0.104	0.051	0.173	-0.050
MW	0.113	-0.229*	0.108	-0.191*	0.172

Figures shown in the table are Pearson Correlation Coefficients.
*Correlation is significant at the 0.05 level (2 tailed).
** Correlation is significant at the 0.01 level (2 tailed)
See Table 11 for definition of the variables.

Table 34 (Continued)
Pearson's Product-Moment Correlation Coefficients of the Variables of the Trend Analysis Task.

	PQ3	TSPQ3	PQ4	TSPQ4	PQ5
TREATMENT	0.131	0.277**	-0.264*	0.112	-0.020
PQ1	0.688**	0.042	0.518**	0.056	0.162
TSPQ1	-0.166	0.664**	0.041	0.307**	0.098
PQ2	0.439**	0.214*	0.648**	0.220*	0.205*
TSPQ2	0.150	0.389**	0.045	0.754**	0.133
PQ3	1.000	0.171	0.552**	0.189*	0.237**
TSPQ3	0.171	1.000	0.224*	0.394**	0.204**
PQ4	0.552**	0.224**	1.000	0.218*	0.147
TSPQ4	0.189	0.394**	0.218*	1.000	0.086
PQ5	0.237**	0.204*	0.147	0.086	1.000
TSPQ5	0.067	0.448**	0.111	0.458**	0.090
PQ6	0.140	0.105	0.316**	0.175	0.101
TSPQ6	0.095	0.369**	0.081	0.461**	0.162
TAQ1	0.361**	0.093	0.664*	0.105	0.101
TSTAQ1	0.109	0.447**	0.190*	0.723**	0.094
TAQ2	0.439	0.186	0.513**	0.159	0.235**
TSTAQ2	0.235	0.327*	0.222*	0.480**	0.088
TAQ3	-0.142	-0.062	-0.058	-0.011	-0.230*
TSTAQ3	0.210*	-0.143	0.189	0.091	0.367**
TAQ4a	-0.039	0.077	-0.018	0.077	0.006
TAQ4b	0.117	-0.049	0.072	0.074	0.085
TAQ4c	0.211*	0.139	0.051	0.168	0.098
TAQ4d	-0.090	0.029	-0.071	0.062	0.139
TSTAQ4	0.206	0.306**	0.228*	0.240**	0.289**
TAQ5	0.400**	0.205	0.262*	0.206*	0.237**
TSTAQ5	-0.124	0.293**	-0.209*	0.401**	-0.045
TAQ6	0.495**	0.133	0.397**	0.115	0.192*
TSTAQ6	-0.006	0.402**	-0.075	0.479**	-0.049
AGE	0.013	0.041	-0.013	0.180*	0.008
GEN	0.160	0.073	0.054	0.072	0.177*
SS	0.071	0.169	0.211*	0.227	-0.109
GPA	-0.058	0.150	-0.114	0.081	0.029
SAT	0.029	-0.019	-0.023	-0.450**	0.249*
FTWE	0.065	-0.067	-0.018	0.151	0.046
PTH	-0.145	0.023	-0.082	0.071	0.041
ARWE	0.054	-0.119	0.071	-0.089	-0.153
HE	0.006	0.051	-0.009	0.015	0.030
SMRT	0.24 9**	-0.036	0.012	-0.091	0.294**
TSMRT	0.229*	0.059	0.161	-0.039	0.292**
MW	-0.238	-0.057	-0.335**	0.064	-0.168

Figures shown in the table are Pearson Correlation Coefficients.
*Correlation is significant at the 0.05 level (2 tailed).
** Correlation is significant at the 0.01 level (2 tailed).
See Table 11 for definition of the variables.

Table 34 (Continued)
Pearson's Product-Moment Correlation Coefficients of the Variables of the Trend Analysis Task.

	TSPQ5	PQ6	TSPQ6	TAQ1	TSTAQ1
TREATMENT	0.053	-0.085	0.006	-0.305**	-0.234**
PQ1	-0.064	0.154	-0.089	0.412**	0.085
TSPQ1	0.350**	0.013	0.260**	0.100	0.486**
PQ2	0.104	0.142	0.228*	0.591**	0.128
TSPQ2	0.472**	0.170	0.372**	0.052	0.618**
PQ3	0.067	0.140	0.095	0.361**	0.109
TSPQ3	0.448**	0.105	0.369**	0.093	0.447**
PQ4	0.111	0.310*	0.081	0.664**	0.190*
TSPQ4	0.458**	0.175	0.461**	0.105	0.723**
PQ5	0.090	0.101	0.162	0.101	0.094
TSPQ5	1.000	0.148	0.526**	-0.003	0.439**
PQ6	0.148	1.000	0.123	0.199*	0.124
TSPQ6	0.526**	0.123	1.000	-0.012	0.394**
TAQ1	-0.003	0.199*	-0.012	1.000	0.146
TSTAQ1	0.439**	0.124	0.394**	0.146	1.000
TAQ2	-0.067	0.143	0.036**	0.547**	0.203*
TSTAQ2	0.390**	0.135	0.137	0.184*	0.450**
TAQ3	-0.107	-0.110	0.216*	0.021	-0.098
TSTAQ3	0.279**	0.161	-0.003	0.154	0.187*
TAQ4a	0.035	0.037	0.083	0.009	0.031
TAQ4b	-0.006	-0.028	0.195	0.192*	0.049
TAQ4c	0.098	-0.052	0.051	0.022	-0.181*
TAQ4d	0.050	0.050	0.414**	0.020	0.044
TSTAQ4	0.510**	0.126	0.215*	0.085	0.235**
TAQ5	0.138	0.069	0.206*	0.277**	0.212*
TSTAQ5	0.304**	-0.009	0.206*	-0.261**	0.362**
TAQ6	0.187*	0.193*	0.171	0.335**	0.034
TSTAQ6	0.413**	0.050	0.409**	-0.099	0.481**
AGE	0.290	-0.013	0.249	-0.016	0.172
GEN	0.061	0.059	-0.075	-0.047	0.002
SS	0.142	-0.005	0.126	0.174	0.230**
GPA	0.038	0.032	0.124	-0.115	-0.011
SAT	-0.253**	0.051	-0.286**	0.014	-0.204*
FTWE	0.190*	-0.002	0.024*	-0.092	0.116
PTH	0.090	-0.085	-0.050	-0.182*	0.030
ARWE	-0.054	0.080	0.106	0.076	-0.169
HE	0.027	-0.043	0.058	0.012	0.124
SMRT	0.068	-0.076	0.051	0.175	-0.026
TSMRT	0.154	0.006	0.222	0.171	0.000
MW	0,086	-0.085	0.028	-0.252	0.056

Figures shown in the table are Pearson Correlation Coefficients.
*Correlation is significant at the 0.05 level (2 tailed).
** Correlation is significant at the 0.01 level (2 tailed).
See Table 11 for definition of the variables.

Table 34 (Continued)
Pearson's Product-Moment Correlation Coefficients of the Variables of the Trend Analysis Task.

	TAQ2	TSTAQ2	TAQ3	TSTAQ3	TAQ4a
TREATMENT	-0.126	-0.073	-0.059	-0.079	0.128
PQ1	0.489**	0.221*	-0.062	0.197*	-0.057
TSPQ1	0.120	0.282**	-0.030	0.197*	0.052
PQ2	0.382**	0.167	-0.045	0.189**	0.072
TSPQ2	0.098	0.332**	-0.106	0.195*	0.072
PQ3	0.439**	0.235**	-0.142	0.210*	-0.039
TSPQ3	0.186*	0.327**	-0.062	0.143	0.077
PQ4	0.513**	0.222*	-0.058	0.189*	-0.018
TSPQ4	0.159	0.480**	-0.011	0.091	0.077
PQ5	0.235**	0.088	-0.230*	0.367**	0.006
TSPQ5	-0.067	0.390**	-0.107	0.279**	0.035
PQ6	0.143	0.135	-0.110	0.161	0.037
TSPQ6	0.036	0.336**	-0.137	0.216*	-0.003
TAQ1	0.547**	0.184*	0.021	0.154	0.009
TSTAQ1	0.203*	0.450**	-0.098	0.187*	0.031
TAQ2	1.000	0.427**	-0.022	0.121	-0.065
TSTAQ2	0.427*	1.000	-0.134	0.204*	-0.121
TAQ3	-0.022	-0.134	1.000	-0.408**	0.256**
TSTAQ3	0.121	0.204	-0.408**	1.000	-0.072
TAQ4a	-0.065	-0.121	0.256**	-0.072	1.000
TAQ4b	0.023	0.086	-0.055	0.114	-0.013
TAQ4c	0.068	0.107	-0.061	0.077	0.102
TAQ4d	-0.094	-0.043	-0.033	0.024	0.101
TSTAQ4	0.107	0.232*	-0.071	0.316**	0.043
TAQ5	0.190*	0.267**	-0.175	0.182*	-0.026
TSTAQ5	-0.235**	0.115	-0.083	-0.077	0.070
TAQ6	0.340**	0.239**	-0.132	0.203*	-0.036
TSTAQ6	-0.050	0.256**	-0.052	0.085	0.054
AGE	-0.208*	-0.086	-0.151	0.065	-0.004
GEN	0.024	0.060	0.148	-0.120	-0.016
SS	0.007	0.102	0.035	0.013	-0.089
GPA	0.037	0.116	-0.085	-0.058	-0.007
SAT	0.120	-0.419**	-0.093	-0.049	0.021
FTWE	-0.219*	-0.149	-0.096	0.077	0.075
PTH	-0.116	0.054	0.122	-0.003	0.041
ARWE	0.030	-0.090	0.026	-0.053	0.005
HE	-0.130	-0.132	-0.079	-0.062	0.008
SMRT	0.235**	0.132	-0.236**	0.337**	-0.084
TSMRT	0.186*	0.159	-0.110	0.235**	-0.023
MW	-0.353**	-0.098	0.064	-0.367	0.071

Figures shown in the table are Pearson Correlation Coefficients.
*Correlation is significant at the 0.05 level (2 tailed).
** Correlation is significant at the 0.01 level (2 tailed).
See Table 11 for definition of the variables.

Table 34 (Continued)
Pearson's Product-Moment Correlation Coefficients of the Variables of the Trend Analysis Task.

	TAQ4b	TAQ4c	TAQ4d	TSTAQ4	TAQ5
TREATMENT	-0.161	-0.032	-0.016	-0.037	-0.078
PQ1	0.177	0.148	-0.086	0.086	0.156
TSPQ1	-0.020	0.130	0.004	0.189*	0.138
PQ2	0.245**	0.165	-0.013	0.293*	0.221*
TSPQ2	0.074	0.150	0.063	0.251**	0.161
PQ3	0.117	0.211*	-0.090	0.206*	0.400*
TSPQ3	-0.049	0.139	0.029	0.306*	0.205*
PQ4	0.072	0.051	-0.071	0.228	0.262**
TSPQ4	0.074	0.168	0.062	0.240**	0.206*
PQ5	0.085	0.088	-0.139	0.289**	0.237*
TSPQ5	-0.006	0.098	0.050	0.560**	0.138
PQ6	-0.028	-0.052	0.050	0.126	0.069
TSPQ6	0.083	0.195*	0.051	0.414**	0.215*
TAQ1	0.192	0.022	0.020	0.085	0.277**
TSTAQ1	0.049	0.181*	0.044	0.235*	0.212*
TAQ2	0.023	0.068	-0.094	0.107	0.190*
TSTAQ2	0.086	0.107	-0.043	0.232**	0.267**
TAQ3	-0.055	-0.061	-0.033	-0.071	-0.175
TSTAQ3	0.114	0.077	0.024	0.316**	0.182*
TAQ4a	-0.013	0.102	0.101	0.043	-0.026
TAQ4b	1.000	0.328**	0.002	-0.033	0.306**
TAQ4c	0.328**	1.000	0.154	0.083	0.311**
TAQ4d	0.002	0.154	1.000	-0.250**	-0.090
TSTAQ4	-0.033	0.083*	-0.250**	1.000	0.258**
TAQ5	0.306*	0.311**	-0.090	0.258*	1.000
TSTAQ5	-0.171	-0.017	0.067	0.069	0.083
TAQ6	0.129	0.349**	-0.076	0.274**	0.467**
TSTAQ6	-0.066	0.144	0.053	0.187*	0.171
AGE	0.052	0.080	-0.011	0.036	0.080
GEN	-0.061	-0.111	-0.138	0.004	0.129
SS	-0.041	-0.055	0.075	0.021	0.095
GPA	-0.006	-0.069	0.011	0.061	-0.029
SAT	-0.007	-0.175	-0.096	-0.019	-0.061
FTWE	0.059	0.085	-0.011	0.045	0.041
PTH	-0.044	-0.053	-0.019	0.052	-0.196*
ARWE	-0.107	-0.302**	-0.052	-0.066	-0.173
HE	0.045	0.008	0.025	-0.067	-0.010
SMRT	0.117	0.056	-0.024	0.182*	0.190*
TSMRT	0.103	0.096	-0.116	0.421*	0.309**
MW	0.004	-0.154	-0.068	-0.050	-0.214*

Figures shown in the table are Pearson Correlation Coefficients.
*Correlation is significant at the 0.05 level (2 tailed).
** Correlation is significant at the 0.01 level (2 tailed).
See Table 11 for definition of the variables.

Table 34 (Continued)
Pearson's Product-Moment Correlation Coefficients of the Variables of the Trend Analysis Task.

	TSTAQ5	TAQ6	TSTAQ6	AGE	GEN
TREATMENT	0.395**	-0.330**	0.167	0.049	0.177*
PQ1	-0.305**	0.371**	-0.089	-0.020	0.040
TSPQ1	0.264**	0.025	0.374**	-0.004	0.008
PQ2	-0.239**	0.374**	0.037	0.029	-0.032
TSPQ2	0.441**	0.159	0.562**	0.349**	0.183*
PQ3	-0.124	0.495**	-0.006	0.013	0.160
TSPQ3	0.293**	0.133	0.402**	0.041	0.073
PQ4	-0.209	0.397**	-0.075	-0.013	0.054
TSPQ4	0.401**	0.115	0.479**	0.180**	0.072
PQ5	-0.045	0.192*	-0.049	0.008	-0.177*
TSPQ5	0.304**	0.181*	0.413**	0.209*	0.061
PQ6	-0.009	0.193*	0.050	-0.013	0.059
TSPQ6	0.206*	0.171	0.049**	0.249**	-0.075
TAQ1	-0.261	0.335*	-0.099	-0.016	-0.047
TSTAQ1	0.362**	0.034	0.481**	0.172	0.002
TAQ2	-0.235**	0.340**	-0.500	-0.208*	0.024
TSTAQ2	0.115	0.239**	0.256**	-0.086	0.060
TAQ3	-0.0083	-0.132	-0.052	-0.151	0.148
TSTAQ3	-0.077	0.203*	0.085	0.065	-0.120
TAQ4a	0.070	-0.036	0.054	-0.004	-0.016
TAQ4b	-0.171	0.129	-0.066	0.052	-0.061
TAQ4c	-0.017	0.349**	0.144	0.080	-0.111
TAQ4d	0.067	-0.076	0.053	-0.011	-0.138
TSTAQ4	0.069	0.274**	0.187*	0.036	0.004
TAQ5	0.083	0.467**	0.171	0.080	0.129
TSTAQ5	1.000	-0.204*	0.579**	0.251**	0.163
TAQ6	-0.204**	1.000	0.122	0.128	0.005
TSTAQ6	0.579**	0.122	1.000	0.345**	0.204*
AGE	0.251**	0.128	0.345**	1.000	0.014
GEN	0.163	0.005	0.204*	0.014	1.000
SS	0.221*	0.017	0.184*	0.473**	0.155
GPA	0.148	0.023	0.171	-0.056	0.077
SAT	0.343**	0.052	-0.225*	-0.001	0.025
FTWE	0.195*	0.114	0.325*	0.825**	-0.003
PTH	-0.005	-0.137	-0.026	0.005	0.207*
ARWE	-0.119	-0.053	-0.155	0.089	0.110
HE	0.241**	-0.067	0.082	0.477**	0.102
SMRT	-0.228*	0.147	-0.218	-0.056	-0.311**
TSMRT	-0.053	0.339*	0.050	-0.051	-0.077
MW	0.202*	-0.270**	0.178*	0.187*	0.072

Figures shown in the table are Pearson Correlation Coefficients.
*Correlation is significant at the 0.05 level (2 tailed).
** Correlation is significant at the 0.01 level (2 tailed).
See Table 11 for definition of the variables.

Table 34 (Continued)
Pearson's Product-Moment Correlation Coefficients of the Variables of the Trend Analysis Task.

	SS	GPA	SAT	FTWE	PTH
TREATMENT	0.066	0.027	0.063	-0.088	-0.193*
PQ1	-0.007	-0.055	0.188	0.012	-0.105
TSPQ1	0.165	0.061	-0.229**	-0.136	-0.111
PQ2	0.089	-0.032	0.051	0.112	-0.036
TSPQ2	0.218*	0.177	-0.197*	0.297**	0.020
PQ3	0.071	-0.058	0.029	0.065	-0.145
TSPQ3	0.169	0.150	-0.109	-0.067	0.023
PQ4	0.211*	-0.114	-0.023	-0.018	-0.082
TSPQ4	0.227*	0.081	-0.450**	0.151	0.071
PQ5	-0.109	0.029	0.249*	0.046	0.041
TSPQ5	0.142	0.038	-0.253*	0.190*	0.090
PQ6	-0.005	0.032	0.051	-0.002	-0.085
TSPQ6	0.126	0.124	-0.286**	0.224*	-0.050
TAQ1	0.174	-0.115	0.014	-0.092	-0.182*
TSTAQ1	0.236**	-0.011	-0.204**	0.116	0.030
TAQ2	0.007	0.037	0.120	-0.219*	-0.116
TSTAQ2	0.102	0.116	-0.419**	0.149	0.054
TAQ3	0.035	-0.085	-0.093	-0.096	0.122
TSTAQ3	0.013	-0.058	-0.049	0.077	-0.003
TAQ4a	-0.089	-0.007	0.021	0.075	0.041
TAQ4b	-0.041	-0.006	-0.007	0.059	-0.044
TAQ4c	-0.055	-0.069	-0.175	0.085	-0.053
TAQ4d	0.075	0.011	-0.096	-0.011**	-0.019
TSTAQ4	0.021	0.061	-0.019	0.045	0.052
TAQ5	0.095	-0.029	-0.061	0.041	-0.196*
TSTAQ5	0.221*	0.148	-0.343**	0.195*	-0.005
TAQ6	0.017	0.023	0.052	0.114	-0.137
TSTAQ6	0.184*	0.171	0.225*	0.325**	-0.026
AGE	0.473**	-0.056	-0.001	**0.825****	0.005
GEN	0.155	0.077	0.025	-0.003	0.207*
SS	1.000	-0.082	-0.181	0.308**	0.027
GPA	-0.082	1.000	0.060	-0.145	-0.165
SAT	-0.181	0.060	1.000	-0.038	-0.060
FTWE	0.308**	-0.145	-0.038	1.000	0.038
PTH	0.027	-0.165	-0.060	0.038	1.000
ARWE	0.260**	-0.015	0.068	0.099	0.029
HE	0.391**	-0.019	0.058	0.347*	-0.153
SMRT	-0.061	-0.080	0.159	-0.043	-0.163
TSMRT	0.074	0.099	0.053	-0.059	-0.158
MW	0.068	0.036	-0.169	0.176	-0.007

Figures shown in the table are Pearson Correlation Coefficients.
*Correlation is significant at the 0.05 level (2 tailed).
** Correlation is significant at the 0.01 level (2 tailed).
See Table 11 for definition of the variables.
The number of observations for the variable of SAT is 103.

Table 34 (Continued)
Pearson's Product-Moment Correlation Coefficients of the Variables of the Trend Analysis Task.

	ARWE	HE	SMRT	TSMRT	MW
TREATMENT	0.154	0.126	-0.084	-0.204	-0.113
PQ1	0.024	0.046	0.196*	0.104	-0.229*
TSPQ1	-0.118	0.046	-0.069	0.051	0.108
PQ2	0.096	-0.013	0.161	0.173	0.191*
TSPQ2	-0.012	0.204*	-0.102	-0.050	0.172
PQ3	0.054	0.006	0.249**	0.229*	-0.238**
TSPQ3	-0.119	0.051	-0.306	0.059	-0.057
PQ4	0.071	-0.009	0.120	0.161	-0.335*
TSPQ4	-0.089	0.015	-0.091	-0.039	0.064
PQ5	-0.153	0.030	0.294**	0.292**	-0.168
TSPQ5	-0.054	0.027	0.068	0.154	0.086
PQ6	0.080	-0.043	-0.076	0.006	-0.085
TSPQ6	-0.106	0.058	0.051	0.222*	0.028
TAQ1	0.076	0.012	0.175	0.171	-0.252**
TSTAQ1	-0.169	0.124	-0.026	0.000	0.056
TAQ2	0.030	-0.130	0.235**	0.186*	-0.353
TSTAQ2	-0.090	-0.132	0.132	0.159	-0.098
TAQ3	0.026	-0.079	-0.236**	-0.110	0.064
TSTAQ3	-0.053	-0.062	0.337**	0.235**	-0.037
TAQ4a	0.005	0.008	-0.084	-0.023	0.071
TAQ4b	-0.107	0.045	0.117	0.103	0.004
TAQ4c	-0.302**	0.008	0.056	0.096	-0.154
TAQ4d	-0.052	0.025	-0.024	-0.116	-0.068
TSTAQ4	-0.066	-0.067	0.182*	0.421**	-0.050
TAQ5	-0.173	-0.010	0.190*	0.309**	-0.214*
TSTAQ5	-0.119	0.241**	-0.228**	-0.053	0.202*
TAQ6	-0.053	-0.067	0.147	0.339**	-0.270**
TSTAQ6	-0.155	0.082	-0.218*	0.050	0.178*
AGE	0.089	0.477**	-0.056	-0.051	0.187*
GEN	0.110	0.102	-0.311**	-0.077	0.072
SS	0.260**	0.391**	-0.061	-0.074	0.068
GPA	-0.015	-0.019	-0.080	0.099	0.036
SAT	0.068	0.058	0.159	0.053	-0.169
FTWE	0.099	0.347**	-0.043	-0.059	0.176
PTH	0.029	-0.153	-1.630	-0.158	-0.007
ARWE	1.000	0.219*	-0.018	-0.123	0.010
HE	0.219*	1.000	-0.093	-0.032	0.098
SMRT	-0.018	-0.093	1.000	0.592**	-0.113
TSMRT	-0.123	-0.032	0.592*	1.000	-0.213*
MW	0.090	0.098	-0.113	-0.213*	1.000

Figures shown in the table are Pearson Correlation Coefficients.
*Correlation is significant at the 0.05 level (2 tailed).
** Correlation is significant at the 0.01 level (2 tailed).
See Table 11 for definition of the variables.

5.4.3 Regression Analysis of Possible Covariates

Regression models allow researchers to estimate the individual contribution of each predictor to the model. If the t-statistic associated with a predictor has a significant alpha value, then the predictor is making a significant contribution to the model (Field, 2005). This study developed separate regression models for each dependent measure using twenty-four possible predictors. Only those predictors that were common covariates across the models used to test a single hypothesis were retained in the models.

A complete discussion of the study's covariates is provided in section 3.60. The covariate measures are: the score on each of the six practice questions (PQ1, PQ2, PQ3, PQ4, PQ5 and PQ6) (see Table 5), the time spent in seconds by each participant when answering each of the six practice questions (TSPQ1, TSPQ2, TSPQ3, TSPQ4, TSPQ5, TSPQ6) (see Table 5), the score on the Mental Rotations Test (SMRT), the time spent in seconds by each participant when answering the Mental Rotations Test (TSMRT), mental workload (MW), gender of the participants (GEN), and age of the participants (AGE). The demographic data of the participants that were considered are: student status (SS), undergraduate overall GPA (GPA), SAT score (SAT), amount of full time working experience (FTWE), part time working hours (PTH), accounting related working experience (ARWE), and highest level of education (HE). Hypothesis H1a has four dependent measures of accuracy − the score on the first question (TAQ1), the score on the second question (TAQ2), the score on the fifth question (TAQ5), and the score on the sixth question (TAQ6). HypothesisH1b uses the time spent on each of the tasks used to test hypothesis H1a to derive four dependent measures of efficiency − the time spent in

seconds answering TAQ1 (TSTAQ1), TAQ2 (TSTAQ2), TAQ5 (TSTAQ5), and TAQ6 (TSTAQ6)

Hypothesis H1c has five dependent measures of accuracy − the score on the third question (TAQ3), the score on part 'a' of the fourth question (TAQ4a), the score on part 'b' of the fourth question (TAQ4b), the score on part 'c' of the fourth question (TAQ4c), the score on part 'd' of the fourth question (TAQ4d). Two measures of efficiency are used to test hypothesis H1d − the time spent in seconds by each participant when answering TAQ3 (TSTAQ3), and when answering all four parts of the fourth question; that is, TAQ4a-TAQ4d (TSTAQ4).

In total there are fifteen dependent measures for hypotheses H1a, H1b, H1c and H1d (see Table 3 and 17). A separate regression model was run for each of the fifteen dependent measures using the twenty-four possible predictors mentioned above in each regression model.

$$\text{Dependent Measures (TAQ1-6 and TSTAQ1-6)} = b_0 + b_1\text{Treatment} + b_2\,\text{PQ1} + b_3\text{TSPQ1} + b_4\text{PQ2} + b_5\text{TSPQ2} + b_6\text{PQ3} + b_7\text{TSPQ3} + b_8\text{PQ4} + b_9\text{TSPQ4} + b_{10}\text{PQ5} + b_{11}\text{TSPQ5} + b_{12}\text{PQ6} + b_{13}\text{TSPQ6} + b_{14}\text{AGE} + b_{15}\text{GEN} + b_{16}\text{SS} + b_{17}\text{GPA} + b_{18}\text{SAT} + b_{19}\text{FTWE} + b_{20}\text{PTH} + b_{21}\text{ARWE} + b_{22}\text{HE} + b_{23}\text{SMT} + b_{24}\text{TSMRT} + b_{25}\text{MW} + e.$$

The results of the analysis indicated that for H1a the following covariates were significant at p-values < 0.05 in one or more of the models, and were therefore retained for testing of H1a: 1) practice question one (PQ1), 2) practice question three (PQ3), 3)

time spent on practice question five (TSPQ5), 4) age (AGE), 5) part time working hours (PTH), 6) highest level of education (HE), and 7) mental workload (MW). For hypothesis H1b the following covariates were significant at p-values < 0.05 in one or more of the models, and were therefore retained for testing of H1b: 1) practice question one (PQ1), 2) time spent on practice question four (TSPQ4), 3) SAT score (SAT), 4) accounting related working experience (ARWE), 5) highest level of education (HE), 6) the score on Mental Rotations Test (SMRT). For hypothesis H1c the following covariates were significant at p-values < 0.05 in one or more of the models, and were therefore retained for testing of H1c: 1) highest level of education (HE), 2) time spent on mental rotations test (TSMRT). Finally, for hypothesis H1d the following covariates were significant at p-values < 0.05 in one or more of the models, and were therefore retained for testing of H1d: 1) practice question five (PQ5), 2) time spent on practice question 5 (TSPQ5), 3) score on the Mental Rotations Test (SMRT), 4) and time spent on Mental Rotations Test (TSMRT).

The following sections discuss the results of testing hypotheses H1a, H1b, H1c, and H1d.

5.4.4 Results of H1a

Four dependent variables are used to test the trend analysis task hypothesis H1a − the score on the first question (TAQ1), the score on the second question (TAQ2), the score on the fifth question (TAQ5), and the score on the sixth question (TAQ6). In constructing the models to test hypothesis H1a, seven covariates − practice question one (PQ1), practice question three (PQ3), time spent on practice question five (TSPQ5), age (AGE), part time working hours (PTH), highest level of education (HE), and mental

workload (MW) − were included in the model (see section 5.4.3) along with the manipulated variable Treatment.

The first dependent variable used to test hypothesis H1a was TAQ1, which asked what the data differences were between years 1 and 4 (see Table 3). Table 35, Panel A indicates that on average those participants viewing the 2-D displays were the most accurate (had the highest score) on this trend analysis task. Those participants viewing a 3-D display (mean score 2.595) were 30% less accurate than those viewing the 2-D display (mean score 3.650) and those using the tabular display (mean score 3.452) were 5% less accurate than those viewing the 2-D display (mean score 3.650).

The second dependent variable (TAQ2) used to test hypothesis H1a asked what participants perceived to be occurring in the data when going from year 2 to year 3 and year 4 (see Table 3). Table 35, Panel A indicates that on average those participants viewing the 2-D displays were the most accurate (had the highest score) on this trend analysis task. Those participants viewing a 3-D perspective display (mean score 4.690) were 20% less accurate than those viewing the 2-D display (mean score 5.925) and those using the tabular display (mean score 5.476) were 7% less accurate than those viewing the 2-D displays (mean score 5.925).

The third dependent variable (TAQ5) used to test hypothesis H1a asked participants to select from a template indicating the differences in data between years 2 and 4 (see Table 3). On average those participants viewing the 2-D displays were again the most accurate. Those participants viewing a 3-D perspective display (mean score 3.642) were 5% less accurate than those viewing the 2-D displays (mean score 3.850)

while those viewing the tabular display (3.761) were 2% less accurate than those viewing the 2-D displays (mean score 3.850) (Table 35, Panel A).

Finally, the last dependent variable (TAQ6) used to test H1a asked participants to select from a template indicating changes in data as you go from year 1 to year 2 to year 3 (see Table 3). As indicated (Table 35, Panel A), on average participants viewing the 2-D displays were the most accurate. Those participants viewing a 3-D perspective display (mean score 6.357) were 16% less accurate than those viewing the 2-D displays (mean score 7.625) while those viewing the tabular display (7.547) were 1% less accurate than those viewing the 2-D displays (mean score 7.625) (Table 35, Panel A). As discussed in this and the prior three paragraphs, the mean results for all dependent variables are as hypothesized in H1a, which predicted that participants viewing a set of 2-D displays will be the most effective in generating hypotheses for what caused the changes in the trend of accounting data when compared to participants using a single 3-D perspective display or participants using a table.

Prior to presenting ANCOVA results for the four dependent variables used to test hypothesis H1a, a MANCOVA analysis was conducted. As shown (Table 35, Panel B), the overall F-statistic for the manipulated variable Treatment is significant ($p = 0.001$) using Pillai's Trace. The results are also significant ($p < 0.001$) using Hotelling's Trace. These significant results allow for analysis of the univariate results which are provided on Panel C of Table 35.

Panel C indicates that the only covariate significantly ($p = 0.040$) associated with TAQ1 is PQ1. Results indicate that manipulation of the presentation formats (Treatment) is significantly ($p = 0.001$) associated with the accuracy of the participants in describing

the differences between years 1 and 4 (TAQ1). A paired comparison test (Table 35, Panel D) shows that the participants viewing the 2-D displays were significantly (p=0.001) more effective or accurate than those participants viewing the 3-D perspective display in this trend analysis task (TAQ1). There was no significant difference in the effectiveness, or accuracy between participants viewing the 2-D displays and participants viewing the tabular display. Thus, the paired comparison tests provide partial support for H1a, which predicted that participants viewing a set of 2-D displays will be the most effective in generating hypotheses for what caused the changes in the trend of accounting data when compared to participants using a single 3-D perspective display or participants using a table.

Panel C indicates that the covariates, PQ1 ($p = 0.003$), and MW ($p = 0.011$) are all significantly associated with TAQ2. Contrary to expectation, the results suggest that the manipulation of the presentation formats (Treatment) does not have an significant ($p = 0.310$) effect on the accuracy of the participants in describing what is happening as you go from year 2 to year 3 to year 4 (TAQ2). Since there is no significant main effect a paired comparison test was not conducted.

Panel C indicates that the covariates, PQ1 ($p = 0.049$), PQ3 ($p < 0.001$), PTH ($p = 0.047$), and MW ($p = 0.050$) are all significantly associated with TAQ5. Again it was found that, contrary to expectations, the results suggest that the manipulation of the presentation formats (Treatment) does not have a significant ($p = 0.296$) effect on the accuracy of the participants in describing the data differences between years 2 and 4 (TAQ5). Thus, a paired comparison test was not conducted.

Finally, Panel C, indicates that the covariates PQ3 ($p < 0.001$), TSPQ5 ($p = 0.016$), AGE ($p = 0.027$), and MW ($p = 0.023$) are significantly associated with the accuracy of the participants' responses to TAQ6. Results also indicate that manipulation of the presentation formats (Treatment) has a significant ($p < 0.001$) effect on the accuracy of the participants in describing what is happening to the data as you go from year 1 to year 2 to year 3 (TAQ6). A paired comparison test was conducted to determine if the participants viewing the 2-D displays were more effective or accurate than those participants reviewing the tabular or 3-D perspective display. Results revealed that the participants viewing the 2-D display were significantly ($p < 0.001$) more accurate than participants viewing the 3-D perspective display (see Table 35, Panel D). There was no significant difference in effectiveness or accuracy between participants viewing the 2-D displays and participants viewing the tabular display. Thus, the paired comparison tests provide partial support for H1a, which predicted that participants viewing a set of 2-D displays will be the most effective in generating hypotheses for what caused the changes in the trend of accounting data when compared to participants using a single 3-D perspective display or participants using a table.

Table 35
Test Results of H1a
(Participants using a set of 2-D displays will be the most effective in generating hypotheses for what caused the changes in the trend of accounting data)
MANCOVA Model on Effectiveness (Accuracy) in Trend Analysis Task
'What are the differences between years 1 and 4?' (TAQ1)
'What is happening as you go from year 2 to year 3 to year 4?' (TAQ2)
'What are the differences between years 2 and 4?' (TAQ5)
'What is happening as you go from year 1 to year 2 to year 3?' (TAQ6)

Tests of Between-Subjects Effects on Effectiveness

Panel A: Mean Scores on the Tasks.

Dependent Variable	Treatment	Actual Mean	MANCOVA Adjusted Mean*
TAQ1	Tabular Display (n=42)	3.452	3.368
	2-D Displays (n=40)	3.650	3.558
	3-D Perspective Display (n=42)	2.595	2.767
TAQ2	Tabular Display (n=42)	5.476	5.180
	2-D Displays (n=40)	5.925	5.632
	3-D Perspective Display (n=42)	4.690	5.266
TAQ5	Tabular Display (n=42)	3.761	3.723
	2-D Displays (n=40)	3.850	3.828
	3-D Perspective Display (n=42)	3.642	3.703
TAQ6	Tabular Display (n=42)	7.547	7.435
	2-D Displays (n=40)	7.625	7.543
	3-D Perspective Display (n=42)	6.357	6.548

*Adjusted Mean is for the effects of the covariates.

Panel B: Multivariate Tests

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.615	44.256	< 0.001
PQ1	Pillai's Trace	0.117	3.684	0.007
PQ3	Pillai's Trace	0.174	5.487	< 0.001
TSPQ5	Pillai's Trace	0.059	1.741	0.146
AGE	Pillai's Trace	0.071	2.130	0.082
PTH	Pillai's Trace	0.046	1.344	0.262
HE	Pillai's Trace	0.032	0.931	0.449
MW	Pillai's Trace	0.101	3.112	0.018
Treatment	Pillai's Trace	0.224	3.527	0.001

PQ1= score on practice question 1; PQ2 = score on practice question 2;
PQ3 = score on practice question 3. TSPQ5 = time spent on practice question five.
PTH = part time working hours. HE = highest level of education. MW = mental workload

Table 35: Test Results of H1a (Continued)

Panel C: ANCOVA Results Using Scores as the Dependent Variables

Dependent Variables	Source of Variation	Type III SS	DF	Mean Square	F Stat	p value*
TAQ1	Corrected Model	49.628	9	5.514	5,417	<0.001
	Intercept	22.581	1	22.581	22.182	<0.001
	PQ1	4.406	1	4.406	4.328	0.040
	PQ3	0.703	1	0.703	0.690	0.408
	TSPQ5	0.731	1	0.731	0.718	0.399
	AGE	0.003	1	0.003	0.003	0.956
	PTH	1.683	1	1.683	1.653	0.201
	HE	0.011	1	0.011	0.100	0.919
	MW	2.910	1	2.910	2.859	0.094
	Treatment	12.488	2	6.244	6.134	0.001
	Error	116.049	114	1.018		
	Total	1456.000	124			
	Corrected Total	165.677	123			
TAQ2	Corrected Model	287.861	9	31.985	6.899	<0.001
	Intercept	91.963	1	91.963	19.836	<0.001
	PQ1	41.917	1	41.917	9.041	0.003
	PQ3	9.017	1	9.017	1.945	0.166
	TSPQ5	0.059	1	0.059	0.013	0.910
	AGE	10.057	1	10.057	2.169	0.144
	PTH	4.340	1	4.340	0.936	0.335
	HE	3.972	1	3.972	0.857	0.357
	MW	30.812	1	30.812	6.646	0.011
	Treatment	4.436	2	2.218	0.478	0.310
	Error	528.526	114	4.636		
	Total	4372.000	124			
	Corrected Total	816.387	123			
TAQ5	Corrected Model	12.716	9	1.413	4.409	<0.001
	Intercept	40.464	1	40.464	126.265	<0.001
	PQ1	1.265	1	1.265	3.948	0.049
	PQ3	5.386	1	5.386	16.807	<0.001
	TSPQ5	0.579	1	0.579	1.807	0.182
	AGE	0.410	1	0.410	1.279	0.260
	PTH	1.296	1	1.296	4.044	0.047
	HE	0.186	1	0.186	0.580	0.448
	MW	1.255	1	1.255	3.916	0.050
	Treatment	0.336	2	0.168	0.525	0.296
	Error	36.534	114	0.320		
	Total	1793.000	124			
	Corrected Total	49.250	123			

Table 35: Test Results of H1a (Continued)

Panel C: ANCOVA Results Using Scores as the Dependent Variables

Dependent Variables	Source of Variation	Type III SS	DF	Mean Square	F Stat	p value*
TAQ6	Corrected Model	116.879	9	12.987	9.456	<0.001
	Intercept	120.368	1	120.368	87.644	<0.001
	PQ1	0.044	1	0.044	0.032	0.858
	PQ3	17.586	1	17.586	12.805	<0.001
	TSPQ5	8.244	1	8.244	6.003	0.016
	AGE	6.910	1	6.910	5.031	0.027
	PTH	1.373	1	1.373	1.000	0.319
	HE	4.014	1	4.014	2.923	0.090
	MW	7.297	1	7.297	5.313	0.023
	Treatment	21.835	2	10.917	7.949	<0.001
	Error	156.564	114	1.373		
	Total	6647.000	124			
	Corrected Total	273.444	123			

TAQ1 Adjusted R Squared = 0.244.
TAQ2 Adjusted R Squared = 0.301.
TAQ5 Adjusted R Squared = 0.200.
TAQ6 Adjusted R Squared = 0.382.
PQ1= practice question 1; PQ2 = practice question 2; PQ3 = practice question 3,
TSPQ5 = time spent on practice question five. PTH = part time working hours.
HE = highest level of education. MW = mental workload
*Treatment p-values are one-tail, all others are two-tail.

Panel D: Bonferroni Pairwise Comparisons for Test H1a

Dependent Variables	(I) Treatment (J) Treatment	Mean Difference (I-J)	Std Error	p value*
TAQ1	2-D Displays Tabular Display	0.190	0.229	0.500
	3-D Display	0.791	0.235	0.001
TAQ6	2-D Displays Tabular Display	0.108	0.266	0.500
	3-D Display	0.995	0.273	<0.001

*p-values are one-tail.

5.4.5 Results of H1b

Four dependent measures are used to test the trend analysis task hypothesis H1b − the time spent in seconds by each participant when answering the first question (TSTAQ1), the time spent in seconds by each participant when answering the second question (TSTAQ2), the time spent in seconds by each participant when answering the fifth question (TSTAQ5), and the time spent in seconds by each participant when answering the sixth question (TSTAQ6). In constructing the models to test hypothesis H1b, six covariates - practice question one (PQ1), time spent on practice question four (TSPQ4), accounting related working experience (ARWE), highest level of education (HE) and score on Mental Rotations Test (SMRT) were included in the model (see section 5.4.3) along with the manipulated variable Treatment

Section 5.3.1 reports that out of the 42 participants viewing the tabular display, only 33 participants had taken the SAT. Out of the 40 participants viewing the 2-D displays, only 32 participants had taken the SAT. Out of the 42 participants viewing the 3-D perspective display, only 37 participants had taken the SAT. By including the SAT score as a covariate in the MANCOVA analysis, the cell sizes not only reduced to 33, 32, and 37 for the tabular display, 2-D displays and 3-D perspective display treatment group, respectively, but the cell sizes also become unequal. With unequal cell sizes among the treatment groups, multivariate tests are no longer robust to the violation of the assumptions of MANCOVA. A separate MANCOVA was run including the SAT variable and the TSTAQ5 results are unchanged and the TSTAQ2 results become significant. However, the Levenw's test indicates heteroscedasticity, therefore, the SAT is excluded from the reported results.

The first dependent variable used to test H1b was TSTAQ1 the time spent in seconds by each participant when answering what the data differences were between years one and four (see Table 3). Table 36, Panel A indicates that on average those participants viewing the tabular display were the most efficient (used the least time in seconds) on this trend analysis task. Those participants viewing a 3-D perspective display (mean seconds 121) used 26% more time (in seconds), than those viewing the 2-D displays (mean seconds 96). But, those participants viewing the 2-D displays (mean seconds 96) used 4% more time (in seconds) than those viewing the tabular display (mean seconds 92).

The second dependent variable TSTAQ2 used to test H1b was the time spent in seconds by each participant when answering what he or she perceived to be occurring in the data when going from year 2 to year 3 and year 4 (see Table 3). Table 36, Panel A indicates that on average those participants viewing the 2-D displays were the most efficient (used the least time in seconds) on this trend analysis task. Those participants viewing a 3-D display (mean seconds 152) used 20% more time (in seconds), than those viewing the 2-D displays (mean seconds 127). Those participants viewing a tabular display (mean seconds 136) used 7% more time (in seconds), than those viewing the 2-D displays (mean seconds 127).

The third dependent variable TSTAQ5 used to test H1b was the time spent in seconds by each participant when selecting choices from a template to indicate the differences in data between years 2 and 4 (see Table 3). Table 36, Panel A indicates on average those participants viewing the 2-D displays and those participants viewing the tabular spent the same amount of time in seconds (mean seconds 35). Those participants

viewing a 3-D perspective display (mean seconds 48) used 37% more time (in seconds), than those viewing the 2-D displays (mean seconds 35).

The fourth dependent variable TSTAQ6 used to test H1b was the time spent in seconds by each participant when selecting choices from a template to indicate the changes in data as you go from year 1 to year 2 to year 3 (see Table 3). Table 36, Panel A indicates on average those participants viewing the 2-D displays were the most efficient (used the least time in seconds) on this trend analysis task. Those participants viewing a 3-D display (mean seconds 81) used 39% more time than those viewing the 2-D displays (mean seconds 58). Those participants viewing the tabular display (mean seconds 65) used 12% more time (in seconds) than those viewing the 2-D display (mean seconds 58). The mean results for the four dependent variables from this and the preceding paragraph provide mixed evidence of support for H1b, which predicted that participants viewing a set of 2-D displays will be the most efficient in generating hypotheses for what caused the changes in the trend of accounting data when compared to participants using a single 3-D perspective display or participants viewing a table.

Prior to presenting ANCOVA results for the four dependent variables used to test hypothesis H1b, a MANCOVA analysis was conducted. As shown (Table 36, Panel B), the overall F-statistic for the manipulated variable Treatment is significant ($p < 0.001$) using Pillai's Trace. The results are also significant ($p < 0.001$) using Hotelling's Trace. These significant results allow for analysis of the univariate results which are provided on Panel C of Table 32.

Panel C indicates that the covariates TSPQ4 ($p < 0.001$), ARWE ($p = 0.010$), and HE ($p = 0.039$) are significantly associated with the time spent in seconds by each

participant (TSTAQ1) when answering the first question in the trend analysis task. Results indicate that manipulation of the presentation formats (Treatment) is significantly ($p = 0.005$) associated with the time spent in seconds by each participant (TSTAQ1) when describing what the data differences were between years 1 and 4. A paired comparison test (Table 36, Panel D) shows that there was no significant difference in the efficiency (time used in seconds) between participants viewing the tabular display, viewing the 2-D displays, and participants viewing the 3-D perspective display when answering what the data differences were between years 1 and 4.

Panel C, indicates that the covariates PQ1 ($p = 0.013$), and TSPQ4 ($p <0.001$) are significantly associated with the time spent in seconds by each participant (TSTAQ2) when answering the second question in the trend analysis task. Contrary to expectations, the results suggest that manipulation of the presentation formats (Treatment) does not have a significant ($p = 0.182$) effect on the time spent in seconds by each participant (TSTAQ2) when describing what he or she perceived to be occurring in the data when going from year 2 to year 3 and year 4. Since there is not significant main effect a paired comparison test was not conducted.

Panel C, indicates that the covariates PQ1 ($p = 0.003$), TSPQ4 ($p < 0.001$), ARWE ($p = 0.019$), and HE ($p = 0.001$) are associated with the time spent in seconds by each participant (TSTAQ5) when answering the fifth question in the trend analysis task. Results suggest that the manipulation of the presentation formats (Treatment) is significantly ($p <0.001$) associated with the time spent by participants (TSTAQ5) when describing the data differences between years 2 and 4. A paired comparison test (Table 36, Panel D) shows that the participants viewing the 2-D displays were significantly ($p <$

0.001) more efficient or used less time in seconds than participants viewing the 3-D perspective display when describing the data differences between years 2 and 4. There was no significant difference in the efficiency (time used in seconds) between participants viewing the 2-D displays and participants viewing the tabular display when answering the fifth question in the trend analysis task. Thus, the paired comparison tests provide partial support for hypothesis H1b, which predicted that participants viewing a set of 2-D displays will be the most efficient in generating hypothesis for what caused the changes in the trend of accounting data when compared to participants using a single 3-D perspective display or participants viewing a table.

Panel C, indicates that the covariates TSPQ4 ($p < 0.001$), and SMRT ($p = 0.049$) are associated with the time spent in seconds by each participant when answering the sixth question in the trend analysis task TSTAQ6. Contrary to expectations, results suggest that manipulation of the presentation formats (Treatment) does not have a significant ($p = 0.057$) effect on the time spent by the participants in describing what is happening as you go from year 1 to year 2 to year 3 (TSTAQ6). Since, there is no significant main effect a paired comparison test was not conducted.

Table 36
Test Results of H1b
(Participants viewing a set of 2-D displays will be the most efficient in generating hypothesis for what caused the changes in the trend of accounting data)
MANCOVA Model on Efficiency (Less Time) in Trend Analysis Task
'What are the differences between years 1 and 4?' (TSTAQ1)
'What is happening as you go from year 2 to year 3 to year 4?' (TSTAQ2)
'What are the differences between years 2 and 4?' (TSTAQ5)
'What is happening as you go from year 1 to year 2 to year 3?' (TSTAQ6)

Tests of Between-Subjects Effects on Efficiency

Panel A: Mean Time Spent on the Tasks.

Dependent Variable	Treatment	Actual Mean (Seconds)	MANCOVA Adjusted Mean* (Seconds)
TSTAQ1	Tabular Display (n=42)	92.785	98.230
	2-D Displays (n=40)	96.975	100.900
	3-D Perspective Display (n=42)	121.904	117.100
TSTAQ2	Tabular Display (n=42)	136.452	132.500
	2-D Displays (n=40)	127.525	130.400
	3-D Perspective Display (n=42)	152.357	153.500
TSTAQ5	Tabular Display (n=42)	35.309	36.341
	2-D Displays (n=40)	35.400	36.632
	3-D Perspective Display (n=42)	48.452	46.247
TSTAQ6	Tabular Display (n=42)	65.523	66.999
	2-D Displays (n=40)	58.500	61.243
	3-D Perspective Display (n=42)	81.333	77.246

*Adjusted Mean is for the effects of the covariates.

Panel B: Multivariae Tests

Variables	Multivariate Test	Value	F Stat	|p value|
Intercept	Pillai's Trace	0.302	12.220	< 0.001
PQ1	Pillai's Trace	0.125	4.030	0.004
TSPQ4	Pillai's Trace	0.594	41.274	< 0.001
ARWE	Pillai's Trace	0.100	3.143	0.017
HE	Pillai's Trace	0.158	5.290	0.001
SMRT	Pillai's Trace	0.067	2.042	0.093
Treatment	Pillai's Trace	0.235	3.791	<0.001

PQ1 = score on practice question 1. HE = highest level of education.
TSPQ4 = time spent in seconds by each participant when answering practice question 4.
ARWE = accounting related working experience. SMRT = the score on Mental Rotations Test

Table 36: Test Results of H1b (Continued)

Panel C: ANCOVA Results Using Time Spent as the Dependent Variables.

Dependent Variables	Source of Variation	Type III SS	DF	Mean Square	F stat	p value*
TSTAQ1	Corrected Model	191337.073	7	27333.868	23.887	<0.001
	Intercept	1188.392	1	1188.392	1.039	0.310
	PQ1	1889.213	1	1889.213	1.651	0.201
	TSPQ4	143034.257	1	143034.257	124.995	<0.001
	ARWE	7832.995	1	7832.995	6.845	0.010
	HE	5009.286	1	5009.286	4.378	0.039
	SMRT	610.019	1	610.019	0.533	0.467
	Treatment	10698.636	2	5349.318	4.675	0.005
	Error	132740.927	116	1144.318		
	Total	1665262.000	124			
	Corrected Total	324078.000	123			
TSTAQ2	Corrected Model	320879.012	7	45839.859	7.740	<0.001
	Intercept	7939.585	1	7939.585	1.341	0.249
	PQ1	37955.049	1	37955.049	6.409	0.013
	TSPQ4	205254.519	1	205254.519	34.657	<0.001
	ARWE	819.495	1	819.495	0.138	0.711
	HE	17593.043	1	17593.043	2.971	0.087
	SMRT	15309.892	1	15309.382	2.585	0.111
	Treatment	12083.532	2	6041.766	1.020	0.182
	Error	687007.787	116	5922.481		
	Total	3402301.000	124			
	Corrected Total	1007886.798	123			
TSTAQ5	Corrected Model	10731.338	7	1533.048	14.160	<0.001
	Intercept	5175.406	1	5175.406	47.801	<0.001
	PQ1	1004.753	1	1004.753	9.280	0.003
	TSPQ4	2516.580	1	2516.580	23.244	<0.001
	ARWE	614.737	1	614.737	5.678	0.019
	HE	1325.330	1	1325.330	12.241	0.001
	SMRT	272.048	1	272.048	2.513	0.116
	Treatment	2368.839	2	1184.419	10.940	<0.001
	Error	12559.210	116	108.269		
	Total	219616.000	124			
	Corrected Total	23290.548	123			

Table 36: Test Results of H1b (Continued)

Panel C: ANCOVA Results Using Time Spent as the Dependent Variables.

Dependent Variables	Source of Variation	Type III SS	DF	Mean Square	F stat	p value*
TSTAQ6	Corrected Model	59297.801	7	8471.114	7.593	<0.001
	Intercept	11041.823	1	11041.823	9.897	0.002
	PQ1	245.762	1	245.762	0.220	0.640
	TSPQ4	32975.793	1	32975.793	29.558	<0.001
	ARWE	3476.405	1	3476.405	3.116	0.080
	HE	1445.381	1	1445.381	1.296	0.257
	SMRT	4434.606	1	4434.606	3.975	0.049
	Treatment	4909.006	2	2454.503	2.200	0.057
	Error	129413.619	116	1115.635		
	Total	772470.000	124			
	Corrected Total	188711.419	123			

TSTAQ1 Adjusted R Squared = 0.566. TSTAQ2 Adjusted R Squared = 0.277
TSTAQ5 Adjusted R Squared = 0.428. TSTAQ6 Adjusted R Squared = 0.273
PQ1 = score on practice question 1.
TSPQ4 = time spent in seconds by each participant when answering practice question 4.
ARWE = accounting related working experience.
HE = highest Level of Education.
SMRT = the score on Mental Rotations Test
*Treatment p-values are one-tail, all others are two-tail

Panel D: Bonferroni Pairwise Comparisons for Test H1b

Dependent Variables	(I) Treatment (J) Treatment	Mean Difference (I-J)	Std Error	p value*
TSTAQ1	2-D Displays Tabular Display	7.087	7.633	0.500
	3-D Display	-16.188	7.808	0.060
TSTAQ5	2-D Displays Tabular Display	0.291	2.348	0.500
	3-D Display	-9.614	2.402	<0.001

*p-values are one-tail.

5.4.6 Results of H1c

Five dependent variables are used to test the trend analysis task hypothesis H1c− the score on the third question (TAQ3), the scores on parts 'a' 'b' 'c' and 'd' of the fourth question (TAQ4a, b, c, d). In constructing the models to test hypothesis H1c, two covariates − highest level of education (HE), and time spent on the mental rotations test (TSMRT) − were included in the models (see section 5.4.3) along with the manipulated variable Treatment.

The first dependent variable used to test hypothesis H1c was TAQ3, which asked the participants to estimate what the ROE would be in year 6 (TAQ3) if each of the variables comprising ROE in year 5 doubles. The correct answer to the third question is 128.34% (see Table 6). The difference between each participant's response and the correct answer is the dependent measure (accuracy) used to test H1c (TAQ3) (see Table 3). Lower scores on this measure indicate greater accuracy. Table 37, Panel A, shows that all participants had a mean estimation of year 6 ROE (TAQ3) that was less than the correct answer by 71.242 (tabular), 87.110 (2-D), and 63.876 (3-D).

The other dependent variable used to test hypothesis H1c was question four, which asked the participants to estimate the average of turnover (TAQ4a), profitability (TAQ4b) and leverage (TAQ4c) for the years 1, 2, 4 and 5, and use the estimated average to calculate a new ROE (TAQ4d). The correct answers to the fourth question are: (a) average turnover = 1.045, (b) average profitability = 3.955%, (c) average leverage = 2.72 and (d) the new ROE = 11.241% (see Table 6). Differences between each participant's responses and the correct answers are the dependent measures (accuracy) used to test H1c (see Table 3). Again, lower scores on this measure indicate greater accuracy.

Table 37, Panel A provides mean results for TAQ4a-d. For TAQ4a participant responses are greater than the correct answer by 5.001 (tabular), 1.503 (2-D), and 0.874 (3-D). The participant's mean estimation of profitability (TAQ4b) was less than the correct answer for tabular display (0.796), and 2-D displays (0.626), while the mean answer for the 3-D perspective display was greater than the correct answer by 1.541. For TAQ4c, the mean estimation of leverage, participant's answers were greater than the correct answer by 0.032 (tabular), 1.898 (2-D), and 0.322 (3-D). Finally, for TAQ4d, mean estimation of ROE, participant's answers were greater than the correct answer for 2-D (16.392) and 3-D (1.055), while answers were less than the correct answer for the tabular display (0.359)

Differences between each participant's responses and the correct answers are the dependent measures (accuracy) used to test hypothesis H1c. The differences between each participant's responses to question three and four can either be a positive or negative value.

The mean results for the dependent variables from the preceding paragraphs provide little support for H1c, which predicted that participants viewing a set of 2-D displays will be the most effective in an accounting judgment involving estimation of values when compared to participants using a single 3-D perspective display or participants using a table.

Prior to presenting ANCOVA results for the four dependent variables used to test hypothesis H1c, a MANCOVA analysis was conducted. As shown (Table 37, Panel B), the overall F-statistic for the manipulated variable Treatment is significant ($p = 0.005$) using Pillai's Trace. The results are also significant ($p = 0.004$) using Hotelling's Trace.

These significant results allow for analysis of the univariate results, which are provided on Panel B of Table 37.

Panel C indicates that manipulation of the presentation formats (Treatment) is significantly ($p = 0.032$) associated with the accuracy of the participants in estimating what the ROE would be in year 6 (TAQ3) if each of the variables comprising ROE in year 5 doubles. A paired comparison test (Table 37, Panel D) shows that the participants viewing the 3-D perspective display were more ($p = 0.031$) effective or accurate than those participants viewing the 2-D displays in this trend analysis task (TAQ3). Thus, hypothesis H1c predicting that participants viewing a set of 2-D displays will be the most effective in an accounting judgment involving estimation of values, when compared to participants using a single 3-D perspective display or participants using a table, was not supported.

Contrary to expectation, Panel C indicates that the manipulation of the presentation formats (Treatment) does not have a significant ($p = 0.170$) effect on the accuracy of the participants in estimating the average of turnover for the years 1, 2, 4 and 5 (TAQ4a). Since there is no significant main effect a paired comparison test was not conducted. Thus, hypothesis H1c predicting that participants viewing a set of 2-D displays will be the most effective in an accounting judgment involving estimation of values, when compared to participants using a single 3-D perspective display or participants using a table, was not supported.

Panel C indicates that the manipulation of the presentation formats (Treatment) does not have a significant ($p = 0.094$) effect on the accuracy of the participants in

estimating the average of profitability for the years 1, 2, 4 and 5 (TAQ4b). Since there is no significant main effect a paired comparison test was not conducted.

Panel C indicates that manipulation of the presentation formats (Treatment) is significantly ($p = 0.018$) associated with the accuracy of the participants in estimating the average of leverage for the years 1, 2, 4 and 5 (TAQ4c). A paired comparison test (Table 37, Panel D) shows that the participants viewing the tabular display were more ($p = 0.032$) effective or accurate than those participants viewing the 2-D displays in this trend analysis task (TAQ4c). A paired comparison test (Table 37, Panel D) also shows that the participants viewing the 3-D perspective display were more ($p = 0.047$) effective or accurate than those participants viewing the 2-D displays in this trend analysis task (TAQ4c). Thus, H1c was not supported.

Panel C indicates that manipulation of the presentation formats (Treatment) is significantly ($p = 0.022$) associated with the accuracy of the participants in estimating the average of turnover, profitability and leverage for the years 1, 2, 4 and 5, and using the estimated average to calculate a new ROE (TAQ4d). A paired comparison test (Table 37, Panel D) shows that the participants viewing the tabular display were more ($p = 0.026$) effective or accurate than those participants viewing the 2-D displays in this trend analysis task (TAQ4d). Thus, hypothesis H1c was not supported.

In summary, H1c is not supported. While the MANCOVA result showed a significant treatment effect, the significant pairwise comparisons at the ANCOVA level show that TAQ3, TAQ4c and TAQ4d, are opposite what was predicted, indicating that use of 2-D displays did not result in greater accuracy by participants.

Table 37
Test Results of H1c
(Participants viewing a set of 2-D displays will be the most effective in an accounting judgment involving estimation of values),
MANOVA Model on Effectiveness (Accuracy) in Trend Analysis Task
'Based on the ROE of year 5 (Bubble 5), what would be year 6 ROE if each of the variables of ROE in year 5 had doubled?' (TAQ3)
'Estimate the average of turnover, profitability and leverage for the years 1, 2, 4, and 5, and use them to calculate a new ROE' (TAQ4a), (TAQ4b), (TAQ4c), (TAQ4d)

Tests of Between-Subjects Effects on Effectiveness

Panel A: Mean Scores on the Tasks.

Dependent Variable	Treatment	Actual Mean	MANCOVA Adjusted Mean*
TAQ3	Tabular Display (n=42)	71.242	72.184
	2-D Displays (n=40)	87.110	88.030
	3-D Perspective Display (n=41)	63.876	62.015
TAQ4a	Tabular Display (n=42)	-5.001	-5.054
	2-D Displays (n=40)	-1.503	-1.490
	3-D Perspective Display (n=41)	-0.874	-0.834
TAQ4b	Tabular Display (n=42)	0.796	0.777
	2-D Displays (n=40)	0.626	0.563
	3-D Perspective Display (n=41)	-1.541	-1.457
TAQ4c	Tabular Display (n=42)	-0.032	-0.105
	2-D Displays (n=40)	-1.898	-1.933
	3-D Perspective Display (n=41)	-0.322	-0.213
TAQ4d	Tabular Display (n=42)	0.359	1.756
	2-D Displays (n=40)	-16.392	-16.329
	3-D Perspective Display (n=41)	-1.055	--2.547

The number of observations of 3-D Perspective Display is 41 instead of 42 as an outlier was dropped
*Adjusted Mean is for the effects of the covariates.

Panel B: Multivariate Tests

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.229	6.789	< 0.001
HE	Pillai's Trace	0.013	0.306	0.908
TSMRT	Pillai's Trace	0.048	1.148	0.339
Treatment	Pillai's Trace	0.205	2.627	0.005

HE = highest level of education.
TSMRT = time spent on mental rotations test.

Table 37: Test Results of H1c (Continued)

Panel C: ANOVA Results Using Scores as the Dependent Variables.

Dependent Variables	Source of Variation	Type III SS	DF	Mean Square	F stat	p value*
TAQ3	Corrected Model	19410.708	4	4852.677	1.998	0.099
	Intercept	73061.681	1	73061.681	30.077	0.000
	HE	2331.893	1	2331.892	0.960	0.329
	TSMRT	5704.231	1	5704.231	2.348	0.128
	Treatment	13641.724	2	68200.862	2.808	0.032
	Error	286643.074	118	2429.179		
	Total	978643.709	123			
	Corrected Total	306053.782	122			
TAQ4a	Corrected Model	414.336	4	103.584	0.563	0.690
	Intercept	28.689	1	28.689	0.156	0.694
	HE	2.950	1	2.950	0.016	0.899
	TSMRT	0.622	1	0.622	0.003	0.954
	Treatment	400.509	2	200.255	1.088	0.170
	Error	21714.821	118	184.024		
	Total	22890.690	123			
	Corrected Total	22129.157	122			
TAQ4b	Corrected Model	171.343	4	42.836	1.246	0.295
	Intercept	29.571	1	29.571	0.860	0.356
	HE	15.855	1	15.855	0.461	0.498
	TSMRT	15.943	1	15.943	0.464	0.497
	Treatment	116.794	2	58.397	1.698	0.094
	Error	4057.884	118	34.389		
	Total	4229.407	123			
	Corrected Total	4229.227	122			
TAQ4c	Corrected Model	98.817	4	24.704	2.004	0.098
	Intercept	24.965	1	24.965	2.025	0.157
	HE	1.571	1	1.571	0.127	0.722
	TSMRT	15.390	1	15.390	1.248	0.266
	Treatment	84.540	2	42.470	3.429	0.018
	Error	1454.674	118	12.328		
	Total	1620.097	123			
	Corrected	1553..492	122			

Table 37: Test Results of H1c (Continued)

Panel C: ANOVA Results Using Scores as the Dependent Variables.

Dependent Variables	Source of Variation	Type III SS	DF	Mean Square	F stat	p value*
TAQ4d	Corrected Model	9167.503	4	2291.876	2.029	0.095
	Intercept	41.362	1	41.362	0.037	0.849
	HE	290.027	1	290.027	0.257	0.613
	TSMRT	1875.519	1	1875.519	1.660	0.200
	Treatment	7175.704	2	3587.852	3.176	0.022
	Error	133298.654	118	1129.650		
	Total	146268.140	123			
	Corrected Total	142466.157	122			

TAQ3 Adjusted R Squared = 0.032.
TAQ4a Adjusted R Squared = -0.015
TAQ4b Adjusted R Squared = 0.008.
TAQ4c Adjusted R Squared = 0.037
TAQ4d Adjusted R Squared = 0.033
HE = highest level of education.
TSMRT = time spent on mental rotations test.
*Treatment p-values are one-tail, all others are two-tail.

Panel D: Bonferroni Pairwise Comparisons for Test H1c

Dependent Variables	(I) Treatment	(J) Treatment	Mean Difference (I-J)	Std Error	p value*
TAQ3	2-D Displays	Tabular Display	15.846	11.026	0.230
		3-D Display	26.015	11.098	0.031
TAQ4c	2-D Displays	Tabular Display	-1.828	0.785	0.032
		3-D Display	-1.720	0.791	0.047
TAQ4d	2-D Displays	Tabular Display	-18.085	7.519	0.026
		3-D Display	-13.782	7.568	0.106

*p-values are one-tail.

5.4.7. Results of H1d

Two dependent variables were used to test the trend analysis task for hypothesis H1d− the time spent in seconds by each participant when answering the third question (TSTAQ3), and the time spent in seconds by each participant when answering all four parts of the fourth question (TSTAQ4). In constructing the models to test hypothesis H1d, four covariates − practice question five (PQ5), time spent on practice question five (TSPQ5), score on the mental rotations test (SMRT), and time spent on the mental rotations test (TSMRT) − were included in the model (see section 5.4.3) along with the manipulated variable Treatment.

The first dependent variable used to test hypothesis H1d was the time spent in seconds by each participant when answering the third question (TSTAQ3) which asked what the ROE in year six would be if each of the variables of ROE in year 5 doubled (see Table 3). Table 38, Panel A indicates that those participants viewing a tabular display (mean seconds 109) used 11% more time (in seconds) than those viewing the 2-D displays (mean seconds 98) .Those participants viewing the 3-D perspective display (mean seconds 94) were as efficient as those participants viewing the 2-D displays when answering the third question.

The second dependent variable used to test hypothesis H1d was the time spent in seconds by each participant when answering the fourth question in the trend analysis task (TSTAQ4), which asked the participants to estimate the average of turnover, profitability, and leverage for the years 1, 2, 4, and 5, and use them to calculate a new ROE (see Table 3). Table 38, Panel A indicates on average participants- viewing the tabular display (mean seconds 221), participants viewing the 2-D displays (mean seconds 218), and

participants viewing a 3-D perspective display (mean second 210) – of different treatment groups spent roughly the same amount of time in seconds when answering the fourth question of the trend analysis task (TSTAQ4).

The mean results for the dependent variables from the preceding paragraphs provide little support for H1d, which predicted that participants viewing a set of 2-D displays will be the most efficient in an accounting judgment involving estimation of values when compared to participants viewing a single 3-D perspective display or subjects viewing a table.

A MANCOVA analysis was conducted. As shown (Table 38, Panel B), the overall F-statistic for the manipulated variable Treatment is not significant ($p = 0.751$) using Pillai's Trace. The results are also not significant ($p = 0.757$) using Hotelling's Trace. Analysis of the univariate results is not necessary. Thus, hypothesis H1d was not supported.

Table 38
Test Results of H1d
(Participants viewing a set of 2-D displays will be the most efficient in an accounting judgment involving estimation of values**)**

MANCOVA Model on Efficiency (Less Time) in Trend Analysis Task
'Based on the ROE of year 5 (Bubble 5), what would be year 6 ROE if each of the variables of ROE in year 5 had doubled?' (TSTAQ3)
'Estimate the average of turnover, profitability and leverage for the years 1, 2, 4, and 5, and use them to calculate a new ROE' (TSTAQ4)

Tests of Between-Subjects Effects on Efficiency

Panel A: Mean Time Spent on the Tasks.

Dependent Variable	Treatment	Actual Mean (Seconds)	MANCOVA Adjusted Mean* (Seconds)
TSTAQ3	Tabular Display (n=42)	109.500	107.300
	2-D Displays (n=40)	98.925	103.800
	3-D Perspective Display (n=42)	94.809	92.387
TSTAQ4	Tabular Display (n=42)	221.333	207.700
	2-D Displays (n=40)	218.325	227.900
	3-D Perspective Display (n=42)	210.095	214.600

*Adjusted Mean is for the effects of the covariates.

Panel B: Multivariate Tests

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.006	0.329	0.720
PQ5	Pillai's Trace	0.113	7.369	0.001
TSPQ5	Pillai's Trace	0.274	21.855	< 0.001
SMRT	Pillai's Trace	0.078	4.921	0.009
TSMRT	Pillai's Trace	0.130	8.696	<0.001
Treatment	Pillai's Trace	0.016	0.479	0.751

PQ1 =score on practice question 5.
TSPQ5 = time spent in seconds by each participant when answering practice question 5.
SMRT = Score on Mental Rotations Test.
TSMRT = Time Spent on Mental Rotations Test.

5.5 Results of the Pattern Recognition Task

The purpose of this section is to test whether there is support for each of the hypothesis H2a, H2b, H2c and H2d. For the pattern recognition task, H2a-d posit that a 3-D display will result in greater effectiveness and efficiency than tabular or 2-D displays. This section is organized in the following way: descriptive statistics of the variables, the Pearson's Product-Moment Correlation of the variables, results of the regression analyses of possible covariates, and the MANCOVA results of each of the hypotheses.

5.5.1 Descriptive Statistics

Table 39 shows the descriptive statistics for the pattern recognition task. Based on the results of one-way ANOVA, the following paragraphs briefly describe those covariates and demographic variables that had significant differences between their means among treatment groups. Descriptive information concerning the dependent variables will be reported separately in the results section of each of the hypotheses.

The first covariate evaluated was performance on the practice questions. For the score on practice question two (PQ2) (What are the differences between the apartments 5 and 6?), participants viewing the tabular display on average had the highest score. Those participants viewing the tabular display (mean score 2.860) were significantly more accurate ($p = 0.014$) than those participants viewing the 3-D perspective display (mean score 2.441). There was no significant difference in accuracy between those participants viewing the tabular display (mean score 2.860) and those participants viewing the 2-D displays (mean score 2.682). There was also no significant differences in accuracy

between those participants viewing the 2-D displays (mean score 2.682) and those participants viewing the 3-D perspective display (mean score 2.441).

For practice question three (PQ3) (What are the values of the factors of apartment 4?), participants viewing the tabular display on average had the highest score. Those participants viewing the tabular display (mean score 3.488) were significantly more accurate ($p< 0.001$) than those participants viewing the 2-D displays (mean score 1.926). Those participants viewing the 3-D perspective display (mean score 3.139) were significantly more accurate ($p< 0.001$) than those participants viewing the 2-D displays (mean score 1.926). There was no significant difference in accuracy between those participants viewing the tabular display (mean score 3.488) and those participants viewing the 2-D displays (mean score 3.139).

For practice question four (PQ4) (What are the differences between the apartments 2 and 4?), participants viewing the tabular display on average had the highest score. Those participants viewing the tabular displays (mean score 3.488) were significantly more accurate ($p = 0.039$) than those participants viewing the 3-D perspective display (mean score 2.976).There was no significant difference in accuracy between those participants viewing the tabular display (mean score 3.488) and those participants viewing the 2-D displays (mean score 3.414). There was no significant difference in accuracy between those participants viewing the 2-D display (mean score 3.414) and those participants viewing the 3-D perspective display (mean score 2.976).

The second covariate evaluated was the time spent on the practice questions. For the time spent by participants on practice question one (TSPQ1), participants viewing the tabular display were the most efficient. Participant viewing the tabular display (mean

seconds 68) were significantly more efficient ($p < 0.001$) or used less time (in seconds) than those participants viewing the 2-D displays (mean seconds 119) and those participants viewing the 3-D perspective display ($p = 102$). There was no significant difference in terms of efficiency between participants viewing 2-D displays (mean seconds 119) and those participants viewing the 3-D perspective display (mean seconds 102).

For the time spent by participants on practice question two (TSPQ2), participants viewing the tabular display were the most efficient. Participants viewing the tabular display (mean seconds 95) were significantly more efficient ($p < 0.001$) or used less time (in seconds) than those participants viewing the 3-D displays (mean seconds 137). Those participants viewing the 2-D displays (mean seconds 105) were significantly ($p = 0.001$) more efficient than those participants viewing the 3-D perspective display (mean seconds 137). There was no significant difference in terms of efficiency between participants viewing tabular display (mean seconds 95) and those participants viewing the 2-D displays (mean seconds 105).

For the time spent by participants on practice question three (TSPQ3), participants viewing the tabular display were the most efficient. Participant viewing the tabular display (mean seconds 58) were significantly more efficient ($p < 0.001$) or used less time (in seconds) than those participants viewing the 2-D displays (mean seconds 101). There was no significant difference in terms of efficiency between participants viewing tabular display (mean seconds 58) and those participants viewing the 3-D perspective display (mean seconds 85). There was no significant difference in terms of efficiency between

participants viewing 2-D displays (mean seconds 101) and those participants viewing the 3-D perspective display (mean seconds 85).

As expected (section 3.6.4), participants viewing the 3-D perspective display had the highest perceived mental workload (MW) (mean 4.447) when compared to participants viewing the 2-D displays (mean 4.164) or participants viewing the tabular display (mean 3.720). Comparison of mean mental workload (MW) using one-way ANOVA shows that there is a significant difference on mental workload between participants viewing the tabular display and participants viewing the 3-D perspective display ($p = 0.053$).

The scores on the Mental Rotations Test (SMRT) between tabular, 2-D and 3-D treatment groups were 24, 20, and 21 points, respectively. T-tests indicate no significant difference in spatial ability among treatment groups. On the Mental Rotations Test, participants viewing the 2-D displays (mean seconds 632) used less time than participants viewing the tabular display (mean seconds 697) or participants viewing the 3-D perspective display (mean seconds 645). Comparison of mean seconds (TSMRT) using one-way ANOVA shows that there is no significant difference on the time spent by participants on the Mental Rotations Test among the treatment groups.

Table 39
Descriptive Statistics for the Pattern Recognition Task

Panel A: Mean (Standard Deviation) and Range of Practice Questions.

	Tabular Display (n=43)	**2-D Displays (n=41)**	**3-D Display (n=43)**
Practice Question One: What are the values of the factors of apartment 2?	3.255 (1.236) 0.00 to 4.00	2.658 (1.590) 0.00 to 4.00	2.790 (1.186) 0.00 to 4.00
Time (seconds) Spent on Practice Question one?	68 (27) 33 to 176	119 (44) 47 to 225	102 (35) 47 to 218
Practice Question Two: What are the differences between the apartments 5 and 6?	2.860 (0.412) 1.00 to 3.00	2.682 (0.819) 0.00 to 3.00	2.441 (0.733) 1.00 to 3.00
Time (seconds) Spent on Practice Question Two	95 (26) 44 to 168	105 (43) 22 to 255	137 (47) 40 to 274
Practice Question Three What are the values of the factors of apartment 4?	3.488 (0.909) 0.00 to 4.00	1.926 (1.081) 0.00 to 4.00	3.139 (0.804) 1.00 to 4.00
Time (seconds) Spent on Practice Question Three	58 (24) 20 to 119	101 (79) 16 to 416	85 (39) 26 to 290
Practice Question Four: What are the differences between the apartments 2 and 4?	3.488 (0.855) 0.00 to 4.00	3.414 (1.048) 0.00 to 4.00	2.976 (0.912) 1.00 to 4.00
Time (seconds) Spent on Practice Question Four	90 (37) 37 to 193	95 (38) 18 to 211	104 (43) 25 to 251
Practice Question Five What is a company's ROE if its turnover ratio is 2, profitability ratio is 5%, and leverage ratio is 1.1?	0.488 (0.505) 0.00 to 1.00	0.463 (0.504) 0.00 to 1.00	0.488 (0.505) 0.00 to 1.00
Time (seconds) Spent on Practice Question Five	127 (51) 67 to 315	151 (101) 43 to 617	132 (60) 41 to 258

Table 39 (Continued)
Descriptive Statistics for the Pattern Recognition Task

Panel A: Mean (Standard Deviation) and Range of Practice Questions.

	Tabular Display (n=43)	**2-D Displays (n=41)**	**3-D Display (n=43)**
Practice Question Six ROE is the sum (multiplicative) of turnover, profitability, and leverage?	0.860 (0.350) 0.00 to 1.00	0.853 (0.357) 0.00 to 1.00	0.953 (0.213) 0.00 to 1.00
Time (seconds) Spent on Practice Question Six	28 (11) 8 to 55	28 (13) 8 to 78	34 (15) 11 to 79

Panel B: Mean (Standard Deviation) and Range of Dependent Variables

	Tabular Display (n=43)	**2-D Displays (n=41)**	**3-D Display (n=43)**
Question One What are the differences between companies 1 and 6?	2.906 (0.478) 0.00 to 3.00	2.731 (0.671) 0.00 to 3.00	2.697 (0.637) 0.00 to 3.00
Time (seconds) Spent on Question One	83 (32) 34 to 192	89 (35) 23 to 223	115 (61) 22 to 323
Question Two Please separate companies 1 through 6 into 2 groups based on similar characteristics?	4.534 (0.797) 3.00 to 6.00	4.560 (0.975) 3.00 to 6.00	5.255 (0.726) 4.00 to 6.00
Time (seconds) Spent on Question Two	76 (37) 30 to 218	96 (52) 18 to 230	82 (52) 31 to 263
Question Three Compared to group two what are the patterns of the financial ratios you are seeing in group one?	1.953 (1.153) 0.00 to 3.00	2.195 (1.054) 0.00 to 3.00	2.116 (0.878) 0.00 to 3.00
Time (seconds) Spent on Question Three	121 (58) 36 to 301	119 (51) 29 to 296	88 (34) 20 to 156
Question Four Comparatively if it is better to have a higher profitability, higher turnover but lower leverage at the same time, which company you will select?	0.255 (0.441) 0.00 to 1.00	0.219 (0.419) 0.00 to 1.00	0.488 (0.505) 0.00 to 1.00

Table 39 (Continued)
Descriptive Statistics for the Pattern Recognition Task

Panel B: Mean (Standard Deviation) and Range of Dependent Variables

	Tabular Display (n=43)	**2-D Displays (n=41)**	**3-D Display (n=43)**
Time (seconds) Spent on Question Four	59 (31) 4 to 158	69 (32) 13 to 151	57 (28) 17 to 142
Question Five What are the differences between companies 4 and 6?	2.906 (0.293) 2.00 to 3.00	2.780 (0.612) 1.00 to 3.00	2.720 (0.590) 1.00 to 3.00
Time (seconds) Spent on Question Five	31 (8) 19 to 56	34 (9) 13 to 59	45 (17) 21 to 89
Question Six Compared to group one, what are the patterns of the financial ratios you are seeing in group two?	2.720 (0.629) 0.00 to 3.00	2.780 (0.652) 0.00 to 3.00	2.627 (0.787) 0.00 to 3.00
Time (seconds) Spent on Question Six	44 (26) 16 to 185	41 (13) 14 to 79	43 (17) 19 to 90

Panel C: Mean (Standard Deviation) and Range of Mental Rotations Test and Mental Workload

	Tabular Display (n=43)	**2-D Displays (n=41)**	**3-D Display (n=43)**
Score on Mental Rotations Test	24.302 (9.460) 4.00 to 40.00	20.561 (11.760) 4.00 to 40.00	21.023 (10.439) 4.00 to 40.00
Time (seconds) Spent on Mental Rotation Test	697 (294) 242 to 2059	632 (308) 64 to 1253	645 (218) 213 to 1141
Mental Workload	3.720 (1.193) 1.50 to 6.00	4.164 (1.490) 1.00 to 7.00	4.447 (1.504) 1.00 to 6.75

Table 39 (Continued)
Descriptive Statistics for the Pattern Recognition Task

Panel D: Demographic Mean and Range.

	Tabular Display (n=43)	**2-D Displays (n=41)**	**3-D Display (n=43)**
AGE			
18-22	n=36	**n=34**	n=36
23-27	n=3	**n=5**	n=3
29-32	n=3	**n=2**	-
33-37	n=1	-	n=3
38-42	-	-	n=1
43-47	-	-	-
48-50	-	-	-
Male/Female	n=22/n=21	n=25/n=16	n=11/n=32
Student Status			
Freshman	-	n=2	n=1
Sophomore	n=9	n=14	n=15
Junior	n=21	n=15	n=14
Senior	n=13	n=8	n=13
Graduate Student	-	n=2	-
GPA	3.252 (0.355) 2.50-3.80	3.22 (0.450) 2.00-4.00	3.375 (0.340) 2.80-4.00
SAT	1147 (156) (n=40) 600-1380	1142 (152) (n=38) 800-1600	1163 (137) (n=42) 900-1500
Years of Full Time Working Experience	1.914 (3.055) 0.00-13.00	2.356 (3.874) 0.00-16.00	2.488 (4.817) 0.00-23.00
Part Time Working Hours Per Week	13.441 (14.042) 0.00-40.00	14.658 (14.689) 0.00-45.00	17.325 (13.987) 0.00-50.00
Accounting Related Working Experience			
Full Time	n=36	n=26	n=33
Part Time	n=5	n=8	n=5
Both Full/Part Time	n=2	n=4	n=3
None	-	n=3	n=2
Highest Education			
High School	n=42	n=37	n=39
Bachelor's degree	n=1	n=4	n=4
Master degree	-	-	-

5.5.2 Pearson's Product-Moment Correlation

5.5.2.1 Correlation between dependent variables. Table 40 shows the Pearson's product-moment correlation coefficients and the significances for the variables of the pattern recognition task. Results of two-tailed testing of the significance of the correlation between variables at 0.05 and 0.01 levels are also shown in Table 40. The following paragraphs discuss in detail the most important correlation results, in terms of whether the dependent variables for each hypothesis are significantly correlated with one another. When reporting significant correlation between two dependent variables, the following paragraphs use the symbols of the dependent variables (see Table 11) to describe each of them and report their correlation coefficient through the notation of "r".

To test for the pattern recognition task (H2a, H2b, H2c, and H2d), six questions are developed (see Table 4) and each participant answers these six questions in the same order. The second and fourth questions are used to test hypotheses H2a and H2b (see Table 4).

Hypothesis H2a has two dependent measures of accuracy –PRQ2 and PRQ4 (see Table 7). PRQ2 is positively related to PRQ4 (r = 0.240) (see Table 40). Since the two dependent measures for hypothesis H2a are positively correlated and MANCOVA is the appropriate statistical method to test hypothesis H2a.

Hypothesis H2b has two dependent measures of efficiency – TSPRQ2 and TSPRQ4 (see Table 7). TSPRQ2 is positively related to TSPRQ4 (r = 0.441),therefore MANCOVA is the appropriate statistical method to test hypothesis H2b (see Table 40).

The first, third, fifth and sixth questions are used to test hypotheses H2c and H2d (see Table 3). Hypothesis H2c has four dependent measures of accuracy – PRQ1, PRQ3,

PRQ5 and PRQ6 (see Table 6). PRQ1 is positively related to PRQ5 (r = 0.367) and PRQ6 (r = 0.187) while PRQ3 is positively related to PRQ6 (r = 0.203), and PRQ5 is positively related to PRQ6 (0.281) (see Table 40). It seems that the four dependent measures of hypothesis H2c are inter-correlated; MANCOVA is the appropriate statistical method to test hypothesis H2c.

Hypothesis H2d has four dependent measures of accuracy – TSPRQ1, TSPRQ3, TSPRQ5 and TSPRQ6 (see Table 6). TSPRQ1 is positively related to TSPRQ3 (r = 0.268) and TSPRQ5 (r = 0.363) while TSPRQ3 is positively related to TSPRQ6 (r = 0.342), and TSPRQ5 is positively related to TSPRQ6 (0.279) (see Table 40). Since the four dependent measures of hypothesis H2d are inter-correlated MANCOVA is the appropriate statistical method to test hypothesis H2d.

5.5.2.2 Multicollinearity. According to Field (2005) when multicollinearity exists, strong correlation between two predictors, in a regression model it is difficult to assess the individual importance of a predictor. Field (2005) further suggests that one way of identifying multicollinearity is to scan a correlation matrix of all the predictor variables and see if any correlate above 0.80. Table 40 shows then non-existence of correlations above 0.80 between any pair of predictor variables. It is concluded that the problem of multicollinearity does not exist and is not a concern here.

5.5.2.3 Correlation between covariates or demographic variables and dependent variables. Table 40 also shows information about the correlations between the treatment variables and the dependent variables, and between the covariates and demographic variables and the dependent variables. As before, significance is shown at alpha levels of 0.05 and 0.01. All significance levels are based on two-tailed tests. A review of the table

indicates that there are some significant correlations between the treatment variables and dependent variables. Additionally, several of the dependent measures are correlated with covariates and demographic variables.

Field (2005) comments that caution must be taken when interpreting correlation coefficients because the correlation coefficient says nothing about which variable causes the other to change. However, the correlation matrix does provide preliminary evidence that the treatment may be associated with some of the dependent variables. There is also preliminary evidence that some covariate and demographic variables may be important controls in the MANOVA and ANOVA models. Therefore, the study used regression analysis for each dependent variable to test the significance of the association between the dependent variables and all of the covariates and demographic variables.

Table 40

Pearson's Product-Moment Correlation Coefficients of the Variables of the Pattern Recognition Task.

	TREATMENT	PQ1	TSPQ1	PQ2	TSPQ2	PQ3
TREATMENT	1.000	-0.141	0.332**	-0.250**	0.395**	-0.126
PQ1	-0.141	1.000	-0.023	0.227*	0.166	0.543**
TSPQ1	0.332**	-0.023	1.000	-0.079	0.403**	-0.218*
PQ2	-0.250**	0.227*	-0.079	1.000	0.108	0.233**
TSPQ2	0.395**	0.166	0.403**	0.108	1.000	0.194*
PQ3	-0.126	0.543**	-0.218*	0.233**	0.194*	1.000
TSPQ3	0.201*	0.086	0.647**	0.059	0.438**	-0.063
PQ4	-0.220*	0.307**	0.044	0.436**	0.061	0.231**
TSPQ4	0.141	0.220*	0.407**	0.205*	0.617**	0.213*
PQ5	0.000	0.172	-0.010	0.129	0.134	0.099
TSPQ5	0.031	0.129	0.418**	0.022	0.297**	0.058
PQ6	0.122	0.310**	-0.063	0.192*	0.009	0.290**
TSPQ6	0.158	0.050	0.248**	-0.007	0.356**	0.089
PRQ1	-0.143	0.216*	0.045	0.504**	0.150	0.176*
TSPRQ1	0.280**	0.089	0.360**	0.088	0.497**	0.111
PRQ2	0.332**	-0.108	0.168	-0.168	0.077	-0.043
TSPRQ2	0.052	0.201*	0.269**	0.268**	0.373**	0.091
PRQ3	0.065	-0.079	0.054	0.208*	0.170	0.017
TSPRQ3	-0.268**	0.188*	0.128	0.180*	0.217*	0.105
PRQ4	0.205*	0.036	0.086	-0.003	0.066	0.066
TSPRQ4	-0.027	0.155	0.286**	0.048	0.284**	0.056
PRQ5	-0.148	0.378**	0.149	0.254**	0.191*	0.302**
TSPRQ5	0.417**	0.082	0.227*	-0.044	0.392**	0.076
PRQ6	-0.056	0.148	0.031	0.256**	0.080	0.121
TSPRQ6	-0.027	0.086	0.133	0.006	0.101	0.068
AGE	0.051	-0.004	0.134	-0.007	0.275**	0.009
GEN	0.211*	0.018	0.111	-0.083	0.224*	0.017
SS	-0.090	-0.031	0.007	-0.121	-0.049	-0.063
GPA	0.132	0.082	-0.118	0.020	-0.066	0.119
SAT	0.096	0.040	-0.109	0.117	-0.152	-0.015
FTWE	0.057	-0.084	0.108	0.019	0.081	-0.050
PTH	0.113	-0.026	-0.057	-0.084	0.101	0.073
ARWE	0.064	-0.081	0.237**	0.065	0.070	-0.171
HE	0.112	0.019	0.173	-0.087	0.187*	-0.021
SMRT	0.127	0.210	0.029	0.143	0.121	0.237**
TSMRT	-0.078	0.257**	0.090	0.216*	0.212*	0.206*
MW	0.211*	-0.262**	0.161	-0.205*	0.123	-0.081

Figures shown in the table are Pearson Correlation Coefficients. See Table 11 for definition of the variables.
*Correlation is significant at the 0.05 level (2 tailed).** Correlation is significant at the 0.01 level (2 tailed).

Table 40 (Continued)
Pearson's Product-Moment Correlation Coefficients of the Variables of the Pattern Recognition Task

	TSPQ3	PQ4	TSPQ4	PQ5	TSPQ5	PQ6
TREATMENT	0.201*	-0220*	0.141	0.000	0.031	0.122
PQ1	0.086	0.307**	0.220*	0.172	0.129	0.310**
TSPQ1	0.647**	0.044	0.407**	-0.010	0.418**	-0.063
PQ2	0.059	0.436**	0.205*	0.129	0.022	0.192*
TSPQ2	0.438**	0.061	0.617**	0.134	0.297**	0.009
PQ3	-0.063	0.231**	0.213*	0.099	0.058	0.290**
TSPQ3	1.000	0.131	0.486**	0.036	0.376**	-0.195*
PQ4	0.131	1.000	0.168	-0.013	0.202*	-0.024
TSPQ4	0.486**	0.168	1.000	0.061	0.289**	0.029
PQ5	0.036	-0.013	0.061	1.000	0.140	0.087
TSPQ5	0.376**	0.202*	0.289**	0.140	1.000	0.024
PQ6	-0.195*	-0.024	0.029	0.087	0.024	1.000
TSPQ6	0.090	0.156	0.395**	0.080	0.294**	0.084
PRQ1	0.108	0.317**	0.191*	-0.014	0.119	0.248**
TSPRQ1	0.327**	0.068	0.570**	0.062	0.295**	0.139
PRQ2	0.026	-0.213*	-0.037	-0.106	0.008	0.001
TSPRQ2	0.304**	0.222*	0.409**	0.048	0.478**	0.147
PRQ3	0.098	0.086	0.123	0.026	0.044	-0.019
TSPRQ3	0.241**	0.280**	0.348**	0.153	0.231**	0.098
PRQ4	0.014	-0.034	-0.030	-0.124	-0.100	-0.026
TSPRQ4	0.134	0.122	0.278**	0.098	0.372**	0.191*
PRQ5	0.178*	0.291**	0.302**	0.061	0.132	0.157
TSPRQ5	0.161	-0.110	0.437**	0.066	0.119	0.090
PRQ6	0.065	0.153	0.165	0.041	-0.153	0.070
TSPRQ6	0.140	-0.007	0.178*	0.029	0.116	0.123
AGE	0.217*	-0.009	0.260**	0.005	0.065	-0.064
GEN	0.174	0.114	0.153	-0.036	0.126	-0.020
SS	-0.072	-0.109	0.003	0.072	0.032	0.010
GPA	-0.145	-0.060	-0.117	-0.050	-0.038	0.270**
SAT	-0.091	0.037	-0.154	0.063	-0.030	-0.002
FTWE	0.203*	0.055	0.082	0.000	0.125	-0.152
PTH	0.053	-0.039	0.093	0.103	-0.024	0.030
ARWE	0.183*	0.033	0.003	0.083	0.128	-0.124
HE	0.160	-0.116	0.090	0.042	0.235**	0.000
SMRT	0.088	0.122	0.219*	0.213*	0.165	0.059
TSMRT	0.144	0.270**	0.262**	0.069	0.367**	0.131
MW	0.137	-0.125	0.071	-0.116	0.121	-0.075

Figures shown in the table are Pearson Correlation Coefficients. See Table 11 for definition.
*Correlation is significant at the 0.05 level (2 tailed). ** Correlation is significant at the 0.01 level (2 tailed).

Table 40 (Continued)
Pearson's Product-Moment Correlation Coefficients of the Variables of the Pattern Recognition Task

	TSPQ6	PRQ1	TSPRQ1	PRQ2	TSPRQ2	PRQ3
TREATMENT	0.158	-0.143	0.280**	0.332**	0.052	0.065
PQ1	0.050	0.216*	0.089	-0.108	0.201*	-0.079
TSPQ1	0.248**	0.045	0.360**	0.168	0.269**	0.054
PQ2	-0.007	0.504**	0.088	-0.168	0.268**	0.208*
TSPQ2	0.356**	0.150	0.497**	0.077	0.373**	0.170
PQ3	0.089	0.176*	0.111	-0.043	0.091	0.017
TSPQ3	0.090	0.108	0.327**	0.026	0.304**	0.098
PQ4	0.156	0.317**	0.068	-0.213*	0.222*	0.086
TSPQ4	0.395**	0.191*	0.570**	-0.037	0.409**	0.123
PQ5	0.080	-0.014	0.062	-0.106	0.048	0.026
TSPQ5	0.294**	0.119	0.295**	0.008	0.478**	0.044
PQ6	0.084	0.248**	0.139	0.001	0.147	-0.019
TSPQ6	1.000	0.015	0.290**	0.002	0.256**	0.123
PRQ1	0.015	1.000	0.064	-0.014	0.274**	0.146
TSPRQ1	0.290**	0.064	1.000	0.072	0.416**	0.081
PRQ2	0.002	-0.014	0.072	1.000	-0.045	0.140
TSPRQ2	0.256**	0.274**	0.416**	-0.045	1.000	0.216*
PRQ3	0.123	0.146	0.081	0.140	0.216*	1.000
TSPRQ3	0.256**	0.102	0.268**	-0.290**	0.444**	0.149
PRQ4	0.065	0.169	0.054	0.240**	0.088	0.155
TSPRQ4	0.340**	0.134	0.334**	0.022	0.441**	0.151
PRQ5	0.220*	0.367**	0.255**	-0.159	0.212*	0.032
TSPRQ5	0.292**	0.146	0.363**	0.085	0.259**	0.059
PRQ6	-0.094	0.187*	0.097	-0.037	0.169	0.203*
TSPRQ6	0.192*	0.163	0.173	0.031	0.297**	-0.081
AGE	0.261**	0.055	0.352**	0.069	0.383**	0.028
GEN	0.034	0.111	0.190*	0.047	0.171	-0.030
SS	0.248**	-0.051	-0.017	-0.024	0.008	0.050
GPA	-0.194*	0.054	-0.020	0.045	0.076	-0.120
SAT	-0.119	0.213*	-0.284**	0.123	-0.156	0.159
FTWE	0.125	0.014	0.104	0.018	0.245**	0.040
PTH	0.034	0.132	0.056	-0.057	-0.245**	0.100
ARWE	-0.005	0.045	-0.027	-0.009	0.138	-0.073
HE	0.113	-0.052	0.265**	0.134	0.344**	0.066
SMRT	0.168	0.024	0.135	0.100	0.098	0.163
TSMRT	0.207*	0.185*	0.185*	0.057	0.373**	0.179*
MW	0.031	-0.020	0.096	0.057	0.021	-0.123

Figures shown in the table are Pearson Correlation Coefficients. See Table 11 for definition.
*Correlation is significant at the 0.05 level (2 tailed) ** Correlation is significant at the 0.01 level (2 tailed).

Table 40 (Continued)
Pearson's Product-Moment Correlation Coefficients of the Variables of the Pattern Recognition Task.

	TSPRQ3	PRQ4	TSPRQ4	PRQ5	TSPRQ5	PRQ6
TREATMENT	-0.268**	0.205*	-0.027	-0.148	0.417**	-0.056
PQ1	0.188*	0.036	0.155	0.378**	0.082	0.148
TSPQ1	0.128	0.086	0.286**	0.149	0.227*	0.031
PQ2	0.180*	-0.003	0.048	0.254**	-0.044	0.256**
TSPQ2	0.217*	0.066	0.284**	0.191*	0.392**	0.080
PQ3	0.105	0.066	0.056	0.302**	0.076	0.121
TSPQ3	0.241**	0.014	0.134	0.178*	0.161	0.065
PQ4	0.280**	-0.034	0.122	0.291**	-0.110	0.153
TSPQ4	0.348**	-0.030	0.278**	0.302**	0.437**	0.165
PQ5	0.153	-0.124	0.098	0.061	0.066	0.041
TSPQ5	0.231**	-0.100	0.372**	0.132	0.119	-0.153
PQ6	0.098	-0.026	0.191*	0.157	0.090	0.070
TSPQ6	0.256**	0.065	0.340**	0.220*	0.292**	-0.094
PRQ1	0.102	0.169	0.134	0.367**	0.146	0.187*
TSPRQ1	0.268**	0.054	0.334**	0.255**	0.363**	0.097
PRQ2	-0.290**	0.240**	0.022	-0.159	0.085	-0.037
TSPRQ2	0.444**	0.088	0.441**	0.212*	0.259**	0.169
PRQ3	0.149	0.155	0.151	0.032	0.059	0.203*
TSPRQ3	1.000	0.012	0.355**	0.284**	0.097	0.198*
PRQ4	0.012	1.000	0.107	0.165	0.164	0.072
TSPRQ4	0.355**	0.107	1.000	0.307**	0.249**	0.195*
PRQ5	0.284**	0.165	0.307**	1.000	0.173	0.281**
TSPRQ5	0.097	0.164	0.249**	0.173	1.000	0.179*
PRQ6	0.198*	0.072	0.195*	0.281**	0.179*	1.000
TSPRQ6	0.342**	0.125	0.305**	0.171	0.279**	0.093
AGE	0.099	-0.088	0.125	0.086	0.341**	0.103
GEN	0.065	0.160	0.034	0.170	0.248**	-0.067
SS	-0.031	0.098	-0.141	-0.093	-0.031	-0.104
GPA	-0.202*	0.021	-0.030	-0.029	0.057	0.009
SAT	-0.197*	0.019	-0.094	-0.068	-0.153	-0.081
FTWE	0.053	-0.092	-0.110	-0.060	0.102	-0.052
PTH	-0.055	-0.028	0.010	0.036	-0.004	0.082
ARWE	0.022	-0.078	-0.059	0.049	0.021	0.000
HE	0.058	0.006	0.105	-0.014	0.313**	-0.017
SMRT	0.058	0.111	0.163	0.229**	-0.076	0.150
TSMRT	0.209*	0.196*	.462**	0.294**	0.131	0.145
MW	-0.028	-0.083	-0.042	-0.244**	0.059	-0.181*

Figures shown in the table are Pearson Correlation Coefficients. See Table 11 for definition.
*Correlation is significant at the 0.05 level (2 tailed).** Correlation is significant at the 0.01 level (2 tailed).

Table 40 (Continued)

Pearson's Product-Moment Correlation Coefficients of the Variables of the Pattern Recognition Task.

	TSPRQ6	AGE	GEN	SS	GPA	SAT
TREATMENT	-0.027	0.051	0.211*	-0.090	0.132	0.096
PQ1	0.086	-0.004	0.018	-0.031	0.082	0.040
TSPQ1	0.133	0.134	0.111	0.007	-0.118	-0.109
PQ2	0.006	-0.007	-0.083	-0.121	0.020	0.117
TSPQ2	0.101	0.275**	0.224*	-0.049	-0.066	-0.152
PQ3	0.068	0.009	0.017	-0.063	0.119	-0.015
TSPQ3	0.140	0.217*	0.174	-0.072	-0.145	-0.091
PQ4	-0.007	-0.009	0.114	-0.109	-0.060	0.037
TSPQ4	0.178*	0.260**	0.153	0.003	-0.117	-0.154
PQ5	0.029	0.005	-0.036	0.072	-0.050	0.063
TSPQ5	0.116	0.065	0.126	0.032	-0.038	-0.030
PQ6	0.123	-0.064	-0.020	0.010	0.270**	-0.002
TSPQ6	0.192*	0.261**	0.034	0.248**	-0.194*	-0.119
PRQ1	0.163	0.055	0.111	-0.051	0.054	0.213*
TSPRQ1	0.173	0.352**	0.190*	-0.017	-0.020	-0.284**
PRQ2	0.031	0.069	0.047	-0.024	0.045	0.123
TSPRQ2	0.297**	0.383**	0.171	0.008	0.076	-0.156
PRQ3	-0.081	0.028	-0.030	0.050	-0.120	0.159
TSPRQ3	0.342**	0.099	0.065	-0.031	-0.202*	-0.197*
PRQ4	0.125	-0.088	0.160	0.098	0.021	0.019
TSPRQ4	0.305**	0.125	0.034	-0.141	-0.030	-0.094
PRQ5	0.171	0.086	0.170	-0.093	-0.029	-0.068
TSPRQ5	0.279**	0.341**	0.248**	-0.031	0.057	-0.153
PRQ6	0.093	0.103	-0.067	-0.104	0.009	-0.081
TSPRQ6	1.000	0.078	0.039	-0.050	-0.069	-0.067
AGE	0.078	1.000	0.103	0.157	-0.008	-0.326**
GEN	0.039	0.103	1.000	-0.032	0.180*	0.015
SS	-0.050	0.157	-0.032	1.000	-0.235**	-0.057
GPA	-0.069	-0.008	0.180*	-0.235**	1.000	-0.010
SAT	-0.067	-0.326**	0.015	-0.057	-0.010	1.000
FTWE	0.011	0.695**	0.113	0.255**	-0.138	-0.105
PTH	-0.083	-0.045	0.053	0.042	-0.193*	0.192*
ARWE	-0.085	0.236**	0.019	0.188*	0.004	0.074
HE	0.024	0.544**	0.130	0.232**	0.099	-0.217*
SMRT	0.053	-0.032	-0.203*	-0.019	-0.037	0.009
TSMRT	0.242**	0.041	0.075	-0.063	0.010	0.054
MW	0.126	0.053	0.144	-0.079	0.041	-0.087

Figures shown in the table are Pearson Correlation Coefficients. See Table 11 for definition of variables.
*Correlation is significant at the 0.05 level (2 tailed).** Correlation is significant at the 0.01 level (2 tailed).

Table 40 (Continued)

Pearson's Product-Moment Correlation Coefficients of the Variables of the Pattern Recognition Task

	FTWE	PTH	ARWE	HE	SMRT	TSMRT	MW
TREATMENT	0.057	0.113	0.064	0.112	-0.127	-0.078	0.211*
PQ1	-0.084	-0.026	-0.081	0.019	0.210*	0.257**	-0.262**
TSPQ1	0.108	-0.057	0.237**	0.173	0.029	0.090	0.161
PQ2	0.019	-0.084	0.065	-0.087	0.143	0.216*	-0.205*
TSPQ2	0.081	0.101	0.070	0.187*	0.121	0.212*	0.123
PQ3	-0.050	0.073	-0.171	-0.021	0.237**	0.206*	-0.081
TSPQ3	0.203*	0.053	0.183*	0.160	0.088	0.144	0.137
PQ4	0.055	-0.039	0.033	-0.116	0.122	0.270**	-0.125
TSPQ4	0.082	0.093	0.003	0.090	0.219*	0.262**	0.071
PQ5	0.000	0.103	0.083	0.042	0.213*	0.069	-0.116
TSPQ5	0.125	-0.024	0.128	0.235**	0.165	0.367**	0.121
PQ6	-0.152	0.030	-0.124	0.000	0.059	0.131	-0.075
TSPQ6	0.125	0.034	-0.005	0.113	0.168	0.207*	0.031
PRQ1	0.014	0.132	0.045	-0.052	0.024	0.185*	-0.020
TSPRQ1	0.104	0.056	-0.027	0.265**	0.135	0.185*	0.096
PRQ2	0.018	-0.057	-0.009	0.134	-0.100	0.057	0.057
TSPRQ2	0.245**	-.245**	0.138	0.344**	0.098	0.373**	0.021
PRQ3	0.040	0.100	-0.073	0.066	0.163	0.179*	-0.123
TSPRQ3	0.053	-0.055	0.022	0.058	0.120	0.209*	-0.028
PRQ4	-0.092	-0.028	-0.078	0.006	0.111	0.196*	-0.083
TSPRQ4	-0.110	0.010	-0.059	0.105	0.163	0.462**	-0.042
PRQ5	-0.060	0.036	0.049	-0.014	0.229**	0.294**	-0.244**
TSPRQ5	0.102	-0.004	0.021	0.313**	-0.076	0.131	0.059
PRQ6	-0.052	0.082	0.000	-0.017	0.150	0.145	-0.181*
TSPRQ6	0.011	-0.083	-0.085	0.024	0.053	0.242**	0.126
AGE	0.695**	-0.045	0.236**	0.544**	-0.032	0.041	0.053
GEN	0.113	0.053	0.019	0.130	-0.203*	0.075	0.144
SS	0.255**	0.042	0.188*	0.232**	-0.019	-0.063	-0.079
GPA	-0.138	-0.193*	0.004	0.099	-0.037	0.010	0.041
SAT	-0.105	0.192*	0.074	-0.217*	0.009	0.054	-0.087
FTWE	1.000	0.042	0.418**	0.416**	-0.136	-0.067	0.123
PTH	0.042	1.000	-0.141	-0.090	0.033	-0.172	0.095
ARWE	0.418**	-0.141	1.000	0.392**	-0.098	0.020	-0.019
HE	0.416**	-0.090	0.392**	1.000	-0.023	0.065	0.000
SMRT	-0.136	0.033	-0.098	-0.023	1.000	0.531**	-0.222
TSMRT	-0.067	-0.172	0.020	0.065	0.531**	1.000	-0.058
MW	0.123	0.095	-0.019	0.000	-0.222*	-0.058	1.000

Figures shown in the table are Pearson Correlation Coefficients. See Table 11 for definition.

*Correlation is significant at the 0.05 level (2 tailed).** Correlation is significant at the 0.01 level (2 tailed).

5.5.3 Regression Analysis of Possible Covariates

This study developed separate regression models for each dependent measure used to test hypotheses H2a-H2d, using twenty-four possible predictors. Only those predictors that were common covariates across the models used to test a single hypothesis were retained in the models.

$$\text{Dependent Measures (PRQ1-6 and TSPRQ1-6)} = b_0 + b_1\text{Treatment} + b_2\text{PQ1} + b_3\text{TSPQ1} + b_4\text{PQ2} + b_5\text{TSPQ2} + b_6\text{PQ3} + b_7\text{TSPQ3} + b_8\text{PQ4} + b_9\text{TSPQ4} + b_{10}\text{PQ5} + b_{11}\text{TSPQ5} + b_{12}\text{PQ6} + b_{13}\text{TSPQ6} + b_{14}\text{AGE} + b_{15}\text{GEN} + b_{16}\text{SS} + b_{17}\text{GPA} + b_{18}\text{SAT} + b_{19}\text{FTWE} + b_{20}\text{PTH} + b_{21}\text{ARWE} + b_{22}\text{HE} + b_{23}\text{SMT} + b_{24}\text{TSMRT} + b_{25}\text{MW} + e.$$

Hypothesis H2a has two dependent measures of accuracy − the score on the second question (PRQ2), and the score on the fourth question (PRQ4). Hypothesis H2b uses the time spent on each of the tasks used to test hypothesis H2a to derive two dependent measures of efficiency − the time spent in seconds answering PRQ2 (TSPRQ2), and PRQ4 (TSPRQ4).

Hypothesis H2c has four dependent measures of accuracy − the score on the first question (PRQ1), the score on the third question (PRQ3), the score on the fifth question (PRQ5), and the score on the sixth question (PRQ6). Hypothesis H2d uses the time spent on each of the tasks used to test hypothesis H2c to derive four dependent measures of efficiency − the time spent in seconds answering PRQ1 (TSPRQ1), PRQ3 (TSPRQ3), PRQ5 (TSPRQ5), and PRQ6 (TSPRQ6).

The results of the analysis indicated that for hypothesis H2a the following covariates were significant at p-values < 0.05 in one or more of the models, and were therefore retained for testing of H2a: 1) time spent on practice question five (TSPQ5), 2) practice question six (PQ6), and 3) the time spent in seconds by each participant when answering the Mental Rotations Test (TSMRT). For hypothesis H2b the following covariates were significant at p-values < 0.05 in one or more of the models, and were therefore retained for testing of H2b: 1) time spent on practice question one (TSPQ1), 2) time spent on practice question five (TSPQ5), 3) time spent on practice question six (TSPQ6), 4) student status (SS), 5) part time working hours (PTH), 6) the score on the Mental Rotations Test (SMRT), 7) and the time spent in seconds by each participant when answering the Mental Rotations Test (TSMRT).

Analysis indicated that for hypothesis H2c the following covariates were significant at p-values < 0.05 in one or more of the models, and were therefore retained for testing of H2c: 1) the score on practice question one (PQ1), 2) the score on practice question two (PQ2), 3) the score on practice question six (PQ6), 4) time spent on practice question five (TSPQ5), 5) gender of the participants (GEN), 6) SAT score (SAT), 7) part time working hours (PTH), 8) accounting related working experience (ARWE), and 9) mental workload (MW). For hypothesis H2d the following covariates were significant, and therefore retained for testing of H2d: 1) the score on practice question four (PQ4), 2) the score on practice question six (PQ6), 3) time spent on practice question four (TSPQ4), 4) age of the participants (AGE), 5) SAT score (SAT), 6) the score on the Mental Rotations Test (SMRT), and 7) the time spent in seconds by each participant when answering the Mental Rotations Test (TSMRT).

The next section will discuss the results of testing hypotheses H2a, H2b, H2c, and H2d.

5.5.4 Results of H2a

Two dependent variables are used to test the pattern recognition task hypothesis H2a − the score on the second question (PRQ2), and the score on the fourth question (PRQ4). In constructing the models to test hypothesis H2a, three covariates − practice question six (PQ6), time spent on practice question five (TSPQ5), and the time spent in seconds by each participant when answering the Mental Rotations Test (TSMRT) − were included in the model (see section 5.5.3) along with the manipulated variable Treatment.

The first dependent variable used to test hypothesis H2a was PRQ2, which asked (see Table 4) participants to separate companies one through six into two groups based on similar financial characteristics. Table 41, Panel A indicates that on average those participants viewing the 3-D perspective displays were the most accurate (had the highest score) on this pattern recognition task. Those participants viewing a tabular display (mean score 4.534) were 14% less accurate than those viewing the 3-D perspective display (mean score 5.255) and those using the 2-D displays (mean score 4.560) were also 14% less accurate than those viewing the 3-D perspective display (mean score 5.255).

The second dependent variable used to test hypothesis H2a was PRQ4, which asked participants to select one of the six companies if the goal is to have high profitability, high turnover but low leverage at the same time (see Table 4). Table 41, Panel A indicates that on average those participants viewing the 3-D perspective display

were the most accurate (had the highest score) on this pattern recognition task. Those participants viewing a tabular display (mean score 0.255) were 48% less accurate than those viewing the 3-D perspective display (mean score 0.488) and those using the 2-D displays (mean score 0.219) were 56% less accurate than those viewing the 3-D perspective display (mean score 0.488). The mean results from this and the preceding paragraph are as hypothesized in H2a, which predicted that participants viewing a single 3-D perspective display will be the most effective in recognizing patterns of accounting data when compared to participants using a set of 2-D displays or participants using a table. Prior to presenting ANCOVA results for the four dependent variables used to test hypothesis H1a, a MANCOVA analysis was conducted. As shown (Table 41, Panel B), the overall F-statistic for the manipulated variable Treatment is significant ($p < 0.001$) using Pillai's Trace. The results are also significant ($p < 0.001$) using Hotelling's Trace. These significant results allow for analysis of the univariate results, which are provided on Panel C of Table 41.

Panel C results indicate that manipulation of the presentation formats (Treatment) is significantly ($p < 0.001$) associated with the accuracy of the participants in separating companies one through six into two groups based on similar financial characteristics. A paired comparison test (Table 41, Panel D) was conducted to determine if the participants viewing the 3-D perspective display were more effective or accurate than those participants reviewing the tabular or 2-D displays. Results revealed that the participants viewing the 3-D perspective display not only were significantly more effective or accurate than those participants viewing the tabular display ($p < 0.001$) but also those participants viewing the 2-D display ($p < 0.001$) on this pattern recognition task (PRQ2).

Thus, the paired comparison tests provide support for H2a, which predicted that participants viewing a single 3-D perspective display will be the most effective in recognizing patterns of accounting data when compared to participants using a set of 2-D displays or participants using a table.

Panel C indicates that the covariates, TSPQ5 ($p = 0.033$), and TSMRT ($p = 0.002$) are significantly associated with PRQ4. Results also indicate that manipulation of the presentation formats (Treatment) has a significant ($p = 0.004$) effect on the accuracy of the participants in selecting one of the six companies if the goal is to have high profitability, high turnover but low leverage at the same time (PRQ4). A paired comparison test was conducted to determine if the participants viewing the 3-D perspective display were more effective or accurate than those participants reviewing the tabular or 2-D displays. Results revealed that the participants viewing the 3-D perspective display not only were significantly more effective or accurate than those participants viewing the tabular display ($p = 0.007$) but also those participants viewing the 2-D displays ($p = 0.016$) on this pattern recognition task (PRQ4). Thus, the paired comparison tests provides support for H2a, which predicted that participants viewing a single 3-D perspective display will be the most effective in recognizing patterns of accounting data when compared to participants using a set of 2-D displays or participants using a table.

Participants answering PRQ4 could either score one point for a correct response or zero for an incorrect response. A logit analysis was performed treating PRQ4 as a dichotomous dependent variable. Results of the logit analysis of the dichotomous PRQ4 variable, with the same covariates, shows that participants viewing the 3-D display had

the highest likelihood of responding correctly compared to participants viewing a tabular display and participants viewing the 2-D displays.

Table 41

Test Results of H2a

(Participants viewing a single 3-D perspective display will be the most effective in recognizing patterns of accounting data**)**

MANCOVA Model on Effectiveness (Accuracy) in Trend Analysis Task

'Please separate companies 1 through 6 into 2 groups based on similar characteristics.' (PRQ2)

'Assuming you cannot select a company solely because of a single variable, for example higher profitability. Comparatively, if it is better to have a higher profitability, higher turnover but lower leverage at the same time, which company you will select?' (PRQ4)

Tests of Between-Subjects Effects on Effectiveness

Panel A: Mean Scores on the Tasks.

Dependent Variable	Treatment	Actual Mean	MANCOVA Adjusted Mean*
PRQ2	Tabular Display (n=43)	4.534	4.518
	2-D Displays (n=41)	4.560	4.563
	3-D Perspective Display (n=43)	5.255	5.271
PRQ4	Tabular Display (n=43)	0.255	0.220
	2-D Displays (n=41)	0.219	0.246
	3-D Perspective Display (n=43)	0.488	0.499

*Adjusted Mean is for the effect of the covariate.

Panel B: Multivariate Tests

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.699	139.220	<0.001
TSPQ5	Pillai's Trace	0.037	2.319	0.103
PQ6	Pillai's Trace	0.014	0.844	0.432
TSMRT	Pillai's Trace	0.078	5.095	0.008
Treatment	Pillai's Trace	0.185	6.163	<0.001

PQ6=score on practice question 6.

TSPQ5 = time spent on practice question five. PTH = part time working hours.

TSMRT = the time spent in seconds by each participant when answering the Mental Rotations Test

Table 41: Test Results of H2a (Continued)

Panel C: ANCOVA Results Using Scores as the Dependent Variables.

Dependent Variables	Source of Variation	Type III SS	DF	Mean Square	F stat	p value*
PRQ2	Corrected Model	15.233	5	3.0477	4.285	0.001
	Intercept	199.370	1	199.370	280.423	<0.001
	TSPQ5	0.007	1	0.007	0.010	0.921
	PQ6	0.429	1	0.429	0.604	0.439
	TSMRT	0.601	1	0.601	0.845	0.360
	Treatment	14.877	2	7.439	10.463	<0.001
	Error	86.027	121	0.711		
	Total	3012.000	127			
	Corrected Total	101.260	126			
PRQ4	Corrected Model	4.090	5	0.818	4.180	0.002
	Intercept	0.759	1	0.759	3.879	0.051
	TSPQ5	0.906	1	0.906	4.630	0.033
	PQ6	0.259	1	0.259	1.322	0.253
	TSMRT	1.968	1	1.968	10.058	0.002
	Treatment	1.988	2	0.994	5.080	0.004
	Error	23.764	121			
	Total	41.000	127			
	Corrected Total	27.764	126			

PRQ2 Adjusted R Squared = 0.244.
PRQ4 Adjusted R Squared = 0.301.
PQ6= score on practice question 6.
TSPQ5 = time spent on practice question five. PTH = part time working hours.
TSMRT = the time spent in seconds by each participant when answering the Mental Rotations Test
*Treatment p-values are one-tail, all others are two-tail.

Panel D: Bonferroni Pairwise Comparisons for Test H2a

Dependent Variables	(I) Treatment (J) Treatment	Mean Difference (I-J)	Std Error	p value*
PRQ2	3-D Displays Tabular Display	0.754	0.185	<0.001
	2-D Display	0.708	0.187	<0.001
PRQ4	3-D Displays Tabular Display	0.279	0.097	0.007
	2-D Display	0.254	0.098	0.016

*p-values are one-tail.

5.5.5 Results of H2b

Hypothesis H2b has two dependent measures of efficiency − the time spent in seconds by each participant when answering the second question (TSPRQ2), and the time spent in seconds by each participant when answering the fourth question (TSPRQ4). In constructing the models to test hypothesis H2b, the following seven covariates were included in the model (see section 5.5.3) along with the manipulated variable Treatment: time spent on practice question one (TSPQ1), time spent on practice question five (TSPQ5), time spent on practice question six (TSPQ6), student status (SS), part time working hours (PTH), score on mental rotations test (SMRT), and time spent on mental rotations test (TSMRT).

The first dependent variable used to test H2b was TSPRQ2 the time spent in seconds by each participant when separating companies one through six into two groups based on similar financial characteristics (see Table 4). Table 42, Panel A indicates that on average those participants viewing the tabular display were the most efficient (used the least time in seconds) on this pattern recognition task. Those participants viewing a 3-D perspective display (mean seconds 82) used 7% more time (in seconds), than those viewing the tabular display (mean seconds 76). But, those participants viewing the 3-D perspective display (mean seconds 82) used 17% less time (in seconds) than those viewing the 2-D displays (mean seconds 96).

The second dependent variable TSPRQ4 used to test H2b was the time spent in seconds by each participant when selecting one of the six companies if the goal is to have high profitability, high turnover but low leverage at the same time (see Table 4). Table 42, Panel A indicates that on average those participants viewing the 3-D perspective

display were the most efficient (used the least time in seconds) on this pattern recognition task. Those participants viewing a tabular display (mean seconds 59) used 4% more time than those viewing the 3-D perspective display (mean seconds 57). Those participants viewing the 2-D displays (mean seconds 69) used 21% more time than those viewing the 3-D perspective display (mean seconds 69). The mean results for the two dependent variables from this and the preceding paragraph provide mixed evidence of support for H2b, which predicted that participants viewing a single 3-D perspective display will be the most efficient in recognizing patterns of accounting data when compared to participants using a set of 2-D displays or participants using a table.

Prior to presenting ANCOVA results for the two dependent variables used to test hypothesis H2b, a MANCOVA analysis was conducted. As shown (Table 42, Panel B), the overall F-statistic for the manipulated variable Treatment is weakly significant ($p = 0.053$) using Pillai's Trace. The results are also weakly significant ($p = 0.055$) using Hotelling's Trace. These results allow for analysis of the univariate results which are provided on Panel C of Table 42.

Panel C indicates that the covariates TSPQ5 ($p < 0.001$), PTH ($p = 0.011$), and TSMRT ($p = 0.013$) are significantly associated with the time spent in seconds by each participant (TSPRQ2) when answering the second question on this pattern recognition task. Contrary to expectations, the results suggest that manipulation of the presentation formats (Treatment) does not have a significant ($p = 0.120$) effect on the time spent in seconds by each participant (TSTAQ2) when separating companies one through six into two groups based on similar financial characteristics. Since there is not a significant main effect a paired comparison test was not conducted.

Panel C indicates that the covariates TSPQ6 ($p < 0.001$), SS ($p = 0.009$), and TSMRT ($p < 0.001$) are significantly associated with the time spent in seconds by each participant (TSPRQ4) when answering the fourth question of this pattern recognition task. Results indicate that manipulation of the presentation formats (Treatment) is significantly ($p = 0.016$) associated with the time spent in seconds by each participant when selecting one of the six companies if the goal is to have high profitability, high turnover but low leverage at the same time. A paired comparison test (Table 42, Panel D) shows that there was no significant difference ($p = 0.121$) in the efficiency (time used in seconds) between participants viewing the tabular display and participants viewing the 3-D perspective display. But those participants viewing the 3-D perspective display were significantly ($p = 0.019$) more efficient or used less time than those participants viewing the 2-D displays when answering the fourth question on this pattern recognition task. Thus, hypothesis H2b predicting that participants viewing a single 3-D perspective display will be the most efficient in recognizing patterns of accounting data, when compared to participants using a set of 2-D displays or participants using a table, was partially supported.

Table 42
Test Results of H2b
(Participants viewing a single 3-D perspective display will be the most efficient in recognizing patterns of accounting data**)**
MANCOVA Model on Efficiency (Less Time) in Trend Analysis Task
'Please separate companies 1 through 6 into 2 groups based on similar characteristics' (TSPRQ2)
'Assuming you cannot select a company solely because of a single variable, for example higher profitability. Comparatively, if it is better to have a higher profitability, higher turnover but lower leverage at the same time, which company you will select?' (TSPRQ4)

Tests of Between-Subjects Effects on Efficiency

Panel A: Mean Time Spent on the Tasks.

Dependent Variable	Treatment	Actual Mean (Seconds)	MANCOVA Adjusted Mean* (Seconds)
TSPRQ2	Tabular Display (n=43)	76.767	77.413
	2-D Displays (n=41)	96.317	94.727
	3-D Perspective Display (n=43)	82.744	83.615
TSPRQ4	Tabular Display (n=43)	59.465	64.350
	2-D Displays (n=41)	69.847	68.185
	3-D Perspective Display (n=43)	57.488	53.871

*Adjusted Mean is for the effect of the covariate

Panel B: Multivariate Tests

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.042	2.543	0.083
TSPQ1	Pillai's Trace	0.022	1.330	0.269
TSPQ5	Pillai's Trace	0.109	7.090	0.001
TSPQ6	Pillai's Trace	0.107	6.933	0.001
SS	Pillai's Trace	0.061	3.756	0.026
PTH	Pillai's Trace	0.093	5.956	0.003
SMRT	Pillai's Trace	0.029	1.709	0.186
TSMRT	Pillai's Trace	0.177	12.438	<0.001
Treatment	Pillai's Trace	0.078	2.374	0.053

TSPQ1 = time spent in seconds by each participant when answering practice question 1.
TSPQ5 = time spent in seconds by each participant when answering practice question 5.
TSPQ6 = time spent in seconds by each participant when answering practice question 6.
SS= student status. PTH = part time working hours. SMRT = the score on Mental Rotations Test
TSMRT = the time spent in seconds by each participant when answering the Mental Rotations Test

Table 42: Test Results of H2b (Continued)

Panel C: ANCOVA Results Using Time Spent as the Dependent Variables.

Dependent Variables	Source of Variation	Type III SS	DF	Mean Square	F stat	p value*
TSPRQ2	Corrected Model	101971.625	9	11330.181	7.076	<0.001
	Intercept	4722.844	1	4722.844	2.949	0.089
	TSPQ1	13.994	1	13.994	0.009	0.926
	TSPQ5	22902.634	1	22902.634	14.303	<0.001
	TSPQ6	4108.135	1	4108.135	2.566	0.112
	SS	0.646	1	0.646	0.000	0.984
	PTH	10667.610	1	10667.610	6.662	0.011
	SMRT	1394.431	1	1394.431	0.871	0.353
	TSMRT	10174.204	1	10174.204	6.354	0.013
	Treatment	4611.560	2	2305.780	1.440	0.120
	Error	187348.044	117	1601.265		
	Total	1209106.000	127			
	Corrected Total	289319.669	126			
TSPRQ4	Corrected Model	48962.334	9	5440.259	9.012	<0.001
	Intercept	2074.348	1	2074.348	3.426	0.066
	TSPQ1	1472.688	1	1472.688	2.440	0.121
	TSPQ5	530.459	1	530.459	0.879	0.350
	TSPQ6	8127.412	1	8127.412	13.463	<0.001
	SS	4280.842	1	4280.842	7.091	0.009
	PTH	1561.159	1	1561.159	2.586	0.111
	SMRT	1924.010	1	1924.010	3.187	0.077
	TSMRT	13997.534	1	13997.534	23.187	<0.001
	Treatment	4300.291	2	2150.146	3.562	0.016
	Error	70629.540	117	603.671		
	Total	608276.000	127			
	Corrected Total	119591.874	126			

TSPRQ2 Adjusted R Squared = 0.303

TSPRQ4 Adjusted R Squared = 0.364

TSPQ1 = time spent in seconds by each participant when answering practice question 1.

TSPQ5 = time spent in seconds by each participant when answering practice question 5.

TSPQ6 = time spent in seconds by each participant when answering practice question 6.

SS= student status. PTH = part time working hours.

SMRT = the score on Mental Rotations Test

TSMRT = the time spent in seconds by each participant when answering the Mental Rotations Test

*Treatment p-values are one-tail, all others are two-tail.

Table 42: Test Results of H2b (Continued)

Panel D: Bonferroni Pairwise Comparisons for Test H2b

Dependent Variables	(I) Treatment (J) Treatment	Mean Difference (I-J)	Std Error	p value
TSPRQ4	3-D Displays Tabular Display	-10.479	5.952	0.121
	2-D Display	-14.287	5.653	0.019

*p-values are one-tail.

5.5.6 Results of H2c

Four dependent variables are used to test the pattern recognition task hypothesis H2c − the score on the first question (PRQ1), the score on the third question (PRQ3), the score on the fifth question (PRQ5), and the score on the sixth question (PRQ6). In constructing the models to test hypothesis H2c, nine covariates − the score on practice question one (PQ1), the score on practice question two (PQ2), the score on practice question six (PQ6), time spent on practice question five (TSPQ5), gender of the participants (GEN), SAT score (SAT), part time working hours (PTH), accounting related working experience (ARWE), and mental workload (MW) − were included in the model (see section 5.5.3) along with the manipulated variable Treatment.

The first dependent variable used to test hypothesis H2c was PRQ1, which asked (see Table 4) participants what the data differences were between companies one and six. Table 43, Panel A indicates that on average those participants viewing the tabular display were the most accurate (had the highest score) on this pattern recognition task. Those participants viewing the 3-D perspective display (mean score 2.697) were 7% less accurate than those viewing the tabular display (mean score 2.906). Those participants viewing the 3-D perspective display (mean score 2.697) were also 3% less accurate than those using the 2-D displays (mean score 2.731).

The second dependent variable (PRQ3) used to test hypothesis H2c asked participants to describe the pattern of financial ratios they were seeing in group one compared to group two (see Table 4). Table 43, Panel A indicates that on average those participants viewing the 2-D displays were the most accurate (had the highest score) on this trend analysis task. Those participants viewing a 3-D perspective display (mean score

2.116) were 3% less accurate than those viewing the 2-D displays (mean score 2.195). However, those using the tabular display (mean score 1.953) were 8% less accurate than those viewing the 3-D perspective display (mean score 2.116).

The third dependent variable used to test hypothesis H2c was PRQ5, which asked (see Table 4) participants what the data differences were between company four and six by selecting choices from a template. Table 43, Panel A indicates that on average those participants viewing the tabular display were the most accurate (had the highest score) on this pattern recognition task. Those participants viewing the 3-D perspective display (mean score 2.720) were 6% less accurate than those viewing the tabular display (mean score 2.906). Those participants viewing the 3-D perspective display (mean score 2.720) were also 2% less accurate than those using the 2-D displays (mean score 2.780).

The fourth dependent variable (PRQ6) used to test hypothesis H2c asked participants to describe the pattern of financial ratios they were seeing in group two compared to group one by selecting choices from a template (see Table 4). Table 43, Panel A indicates that on average those participants viewing the 2-D displays were the most accurate (had the highest score) on this trend analysis task. Those participants viewing a 3-D perspective display (mean score 2.627) were 5% less accurate than those viewing the 2-D displays (mean score 2.780). Those participants viewing the 3-D perspective display (mean score 2.627) were also 3% less accurate than those viewing the tabular display (mean score 2.720).

The mean results for the dependent variables from this and the preceding paragraph provide little support for hypothesis H2c, which predicted that participants using a single 3-D perspective display will be the most effective in generating hypotheses

for what caused the emerged patterns when compared to participants using a set of 2-D displays or participants using a table.

A MANCOVA analysis was conducted. As shown (Table 38, Panel B), the overall F-statistic for the manipulated variable Treatment is not significant ($p = 0.506$) using Pillai's Trace. The results are also not significant ($p = 0.503$) using Hotelling's Trace. Analysis of the univariate results is not necessary. Thus, hypothesis H2c predicting that participants using a single 3-D perspective display will be the most effective in generating hypotheses for what caused the emerged patterns, when compared to participants using a set of 2-D displays or participants using a table, was not supported.

Table 43
Test Results of H2c
(Participants using a single 3-D perspective display will be the most effective in generating hypotheses for what caused the emerged patterns.)
MANCOVA Model on Effectiveness (Accuracy) in Trend Analysis Task
'What are the differences between companies 1 and 6?' (PRQ1)
'Group one includes companies 1, 3, 4, and 6, and group two includes companies 2 and 5. Compared to group two, what are the patterns of the financial ratios you are seeing in group one?' (PRQ3)
'What are the differences between companies 4 and 6?' (PRQ5)
'Group one includes companies 1, 3, 4 and 6, and group two includes companies 2 and 5. Compared to group one what are the patterns of the financial ratios you are seeing in group two?' (PRQ6)

Tests of Between-Subjects Effects on Effectiveness

Panel A: Mean Scores on the Tasks.

Dependent Variable	Treatment	Actual Mean	MANCOVA Adjusted Mean*
PRQ1	Tabular Display (n=43)	2.906	2.885
	2-D Displays (n=41)	2.731	2.757
	3-D Perspective Display (n=43)	2,697	2.695
PRQ3	Tabular Display (n=43)	1.953	1.863
	2-D Displays (n=41)	2.195	2.194
	3-D Perspective Display (n=43)	2.116	2.208
PRQ5	Tabular Display (n=43)	2.906	2.853
	2-D Displays (n=41)	2.780	2.820
	3-D Perspective Display (n=43)	2.720	2.738
PRQ6	Tabular Display (n=43)	2.720	2.616
	2-D Displays (n=41)	2.780	2.803
	3-D Perspective Display (n=43)	2.627	2.712

*Adjusted Mean is for the effect of the covariate

Table 43: Test Results of H2c (continued)

Panel B: Multivariate Tests

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.353	15.243	<0.001
PQ1	Pillai's Trace	0.098	3.055	0.020
PQ2	Pillai's Trace	0.253	9,472	<0.001
PQ6	Pillai's Trace	0.043	1.256	0.292
TSPQ5	Pillai's Trace	0.067	2.025	0.096
GEN	Pillai's Trace	0.066	1.967	0.104
SAT	Pillai's Trace	0.100	3.115	0.018
PTH	Pillai's Trace	0.051	1.492	0.209
ARWE	Pillai's Trace	0.032	0.923	0.453
MW	Pillai's Trace	0.088	2.685	0.035
Treatment	Pillai's Trace	0.063	0.914	0.506

PQ1 = practice question 1.
PQ2 = practice question 2.
PQ6 = practice question 6.
TSPQ5 = time spent on practice question five.
GEN = gender.
SAT = scores of SAT.
PTH = part time working hours.
ARWE = accounting related working experience.
MW = mental workload.

5.5.7 Results of H2d

Four dependent variables are used to test the pattern recognition task hypothesis H2d − the time spent in seconds by each participant when answering the first question (TSPRQ1), the time spent in seconds by each participant when answering the third question (TSPRQ3), the time spent in seconds by each participant when answering the fifth question (TSPRQ5), and the time spent in seconds by each participant when answering the sixth question (TSPRQ6).

In constructing the models to test hypothesis H2d, seven covariates − the score on practice question four (PQ4), the score on practice question six (PQ6), time spent on practice question four (TSPQ4), age of the participants (AGE), SAT score (SAT), the score on the Mental Rotations Test (SMRT), the time spent in seconds by each participant when answering the Mental Rotations Test (TSMRT) − were included in the model (see section 5.4.3) along with the manipulated variable Treatment.

The first dependent variable used to test H2d was TSPRQ1, the time spent in seconds by each participant when describing what the data differences were between companies 1 and 6. Table 44, Panel A indicates that on average those participants viewing the tabular display were the most efficient (used the least time in seconds) on this pattern recognition task. Those participants viewing a 3-D perspective display (mean seconds 115) used 37% more time, than those viewing the tabular display (mean seconds 83).Those participants viewing the 3-D displays (mean seconds 115) used 29% more time than those viewing the 2-D displays (mean seconds 89).

The second dependent variable used to test H2d was TSPRQ3 the time spent in seconds by each participant when describing the pattern of financial ratios in group one.

Table 44, Panel A indicates that on average those participants viewing the 3-D perspective display were the most efficient on this pattern recognition task. Those participants viewing the tabular display (mean seconds 121) used 37% more time than those viewing a 3-D perspective display (mean seconds 88). Those participants viewing the 2-D displays (mean seconds 119) used 36% more time than those viewing the 3-D perspective display (mean seconds 88).

The third dependent variable used to test H2d was TSPRQ5 the time spent in seconds by each participant when describing what the data differences were between companies 4 and 6. Table 44, Panel A indicates that on average those participants viewing the tabular display were the most efficient (used the least time in seconds) on this pattern recognition task. Those participants viewing a 3-D perspective display (mean seconds 45) used 43% more time than those viewing the tabular display (mean seconds 31). Those participants viewing the 3-D perspective display (mean seconds 45) used 34% more time than those viewing the 2-D displays (mean seconds 34).

The fourth dependent variable used to test H2d was TSPRQ6 the time spent in seconds by each participant when describing the pattern of financial ratios in group two. Table 44, Panel A indicates that on average those participants viewing the 2-D displays were the most efficient on this pattern recognition task. Those participants viewing a 3-D perspective display (mean seconds 43) used 3% more time than those viewing the 2-D displays (mean seconds 41). Those participants viewing the tabular display (mean seconds 44) used 3% more time than those viewing the 3-D perspective display (mean seconds 43). The mean results for the dependent variables indicate that there may be mixed support for H2d, which predicted that participants viewing a single 3-D

perspective display will be the most efficient in generating hypotheses for what caused the emerged patterns when compared to participants viewing a set of 2-D displays or participants using a table.

Prior to presenting ANCOVA results for the four dependent variables used to test hypothesis H2d, a MANCOVA analysis was conducted. As shown (Table 44, Panel B), the overall F-statistic for the manipulated variable Treatment is significant ($p < 0.001$) using Pillai's Trace. The results are also significant ($p < 0.001$) using Hotelling's Trace. These significant results allow for analysis of the univariate results, which are provided on Panel C of Table 39.

Panel C, indicates that the covariates TSPQ4 ($p < 0.001$), AGE ($p = 0.030$), and SAT ($p = 0.008$) are associated with the time spent in seconds by each participant (TSPRQ1) when answering the first question in the pattern recognition task. Results suggest that the manipulation of the presentation formats (Treatment) is significantly ($p = 0.002$) associated with the time spent by participants (TSPRQ1) when describing the data differences between companies 1 and 6. A paired comparison test (Table 44, Panel D) shows that both the participants viewing the tabular display ($p = 0.002$) and the participants viewing the 2-D displays ($p = 0.011$) were more efficient, using less time than participants viewing the 3-D perspective display when describing the data differences between companies 1 and 6. Thus, hypothesis H2d predicting that participants viewing a single 3-D perspective display will be the most efficient in generating hypotheses for what caused the emerged patterns, when compared to participants viewing a set of 2-D displays or participants using a table, was not supported.

Panel C, indicates that the covariate TSPQ4 ($p < 0.001$) is associated with the time spent in seconds by each participant (TSPRQ3) when answering the third question in the pattern recognition task. Results suggest that the manipulation of the presentation formats (Treatment) is significantly ($p < 0.001$) associated with the time spent by participants (TSPRQ3) when describing the pattern of financial ratios in group one. A paired comparison test (Table 44, Panel D) shows that the participants viewing the 3-D perspective display were not only significantly more efficient or used less time than participants viewing the tabular display ($p = 0.001$), but also significantly more efficient than participants viewing the 2-D displays ($p = 0.002$) when describing the pattern of financial ratios in group one. Thus, hypothesis H2d predicting that participants viewing a single 3-D perspective display will be the most efficient in generating hypotheses for what caused the emerged patterns, when compared to participants viewing a set of 2-D displays or participants using a table was supported using TSPRQ3.

Panel C indicates that the covariates TSPQ4 ($p < 0.001$), AGE ($p = 0.017$), SMRT ($p = 0.016$), and TSMRT ($p = 0.032$) are associated with the time spent in seconds by each participant (TSPRQ5) when answering the fifth question in the pattern recognition task. Results suggest that the manipulation of the presentation formats (Treatment) is significantly ($p < 0.001$) associated with the time spent by participants (TSPRQ4) when describing the data differences between companies 4 and 6. A paired comparison test (Table 44, Panel D) shows that both the participants viewing the tabular display ($p < 0.001$) and the participants viewing the 2-D displays ($p < 0.001$) were more efficient or used less time than participants viewing the 3-D perspective display when describing the data differences between companies 4 and 6. Thus, hypothesis H2d was not supported.

Panel C, contrary to expectations, suggests that manipulation of the presentation formats (Treatment) does not have a significant ($p = 0.35$) effect on the time spent by the participants when describing the pattern of financial ratios in group two (TSPRQ6). Since there is no significant main effect a paired comparison test was not conducted. Thus, TSPRQ6 does not support hypothesis H2d.

Table 44
Test Results of H2d
(Participants viewing a single 3-D perspective display will be the most efficient in generating hypotheses for what caused the emerged patterns)**.**
MANCOVA Model on Efficiency (Less Time) in Trend Analysis Task
'What are the differences between companies 1 and 6?' (TSPRQ1)
'Group one includes companies 1, 3, 4, and 6, and group two includes companies 2 and 5. Compared to group two, what are the patterns of the financial ratios you are seeing in group one?' (TSPRQ3)
'What are the differences between companies 4 and 6?' (TSPRQ5)
'Group one includes companies 1, 3, 4 and 6, and group two includes companies 2 and 5. Compared to group one what are the patterns of the financial ratios you are seeing in group two?' (TSPRQ6)

Tests of Between Subjects Effects on Efficiency

Panel A: Mean Time Spent on the Tasks.

Dependent Variable	Treatment	Actual Mean (Seconds)	MANCOVA Adjusted Mean* (Seconds)
TSPRQ1	Tabular Display (n=43)	83.558	85.458
	2-D Displays (n=41)	89.097	90.215
	3-D Perspective Display (n=43)	115.279	112.300
TSPRQ3	Tabular Display (n=43)	121.418	122.700
	2-D Displays (n=41)	119.902	119.900
	3-D Perspective Display (n=43)	88.139	86.876
TSPRQ5	Tabular Display (n=43)	31.976	33.228
	2-D Displays (n=41)	34.097	34.469
	3-D Perspective Display (n=43)	45.953	44.348
TSPRQ6	Tabular Display (n=43)	44.604	45.377
	2-D Displays (n=41)	41.756	42.462
	3-D Perspective Display (n=43)	43.325	41.880

*Adjusted Mean is for the effect of the covariate

Table 44: Test Results of H2d (Continued)

Panel B: Multivariae Tests

Variables	Multivariate Test	Value	F stat	|p value|
Intercept	Pillai's Trace	0.181	6.317	<0.001
PQ4	Pillai's Trace	0.057	1.707	0.153
TSPQ4	Pillai's Trace	0.342	14.802	<0.001
PQ6	Pillai's Trace	0.035	1.022	0.399
AGE	Pillai's Trace	0.092	2.901	0.025
SAT	Pillai's Trace	0.076	2.356	0.058
SMRT	Pillai's Trace	0.053	1.611	0.176
TSMRT	Pillai's Trace	0.076	2.350	0.058
Treatment	Pillai's Trace	0.356	6.225	<0.001

PQ4 = score on practice question 4.
TSPQ4 = time spent in seconds by each participant when answering practice question 4.
PQ6 = score on practice question 6.
AGE= age.
SAT = scores of SAT
SMRT = the score on Mental Rotations Test
TSMRT =the time spent in seconds by each participant when answering the Mental Rotations Test

Panel C: ANCOVA Results Using Time Spent as the Dependent Variables

Dependent Variables	Source of Variation	Type III SS	DF	Mean Square	F stat	p value*
TSPRQ1	Corrected Model	128262.789	9	14251.421	11.274	<0.001
	Intercept	3934.007	1	3934.007	3.112	0.080
	PQ4	584.685	1	584.655	0.463	0.489
	TSPQ4	40991.664	1	40994.664	32.429	<0.001
	PQ6	2307.231	1	2307.231	1.825	0.179
	AGE	6089.008	1	6089.008	4.817	0.030
	SAT	9141.717	1	9141.717	7.232	0.008
	SMRT	229.995	1	229.995	0.182	0.670
	TSMRT	349.556	1	349.556	0.277	0.600
	Treatment	14993.777	2	7496.888	5.931	0.002
	Error	147895.258	117	1264.062		
	Total	1448703.000	127			
	Corrected Total	276158.047	126			

Table 44: Test Results of H2d (Continued)

Panel C: ANCOVA Results Using Time Spent as the Dependent Variables

Dependent Variables	Source of Variation	Type III SS	DF	Mean Square	F stat	p value*
TSPRQ3	Corrected Model	99253.737	9	11028.193	5.566	<0.001
	Intercept	4437.345	1	4437.345	2.239	0.137
	PQ4	5372.637	1	5372.637	2.711	0.102
	TSPQ4	29529.177	1	29529.177	14.902	<0.001
	PQ6	5252.440	1	5252.440	2.651	0.106
	AGE	1.782	1	1.782	0.001	0.976
	SAT	3833.227	1	3833.227	1.934	0.167
	SMRT	462.676	1	462.667	0.233	0.630
	TSMRT	1486.263	1	14862.63	0.750	0.388
	Treatment	28980.825	2	14490.413	7.313	<0.001
	Error	231836.703	117	1981.510		
	Total	1858345.000	127			
	Corrected Total	331090.441	126			
TSPRQ5	Corrected Model	10610.296	9	1178.922	10.208	<0.001
	Intercept	2238.411	1	2238.411	19.382	<0.001
	PQ4	207.424	1	207.424	1.796	0.183
	TSPQ4	2191.325	1	2191.325	18.975	<0.001
	PQ6	14.364	1	14.364	0.124	0.725
	AGE	679.073	1	679.073	5.880	0.017
	SAT	153.426	1	153.426	1.329	0.251
	SMRT	685.363	1	685.363	5.935	0.016
	TSMRT	543.457	1	543.457	4.706	0.032
	Treatment	2710.509	2	1355.254	11.735	<0.001
	Error	13512.019	117	115.487		
	Total	201705.000	127			
	Corrected Total	24122.315	126			

PQ4 = score on practice question 4.
TSPQ4 = time spent in seconds by each participant when answering practice question 4.
PQ6 = score on practice question 6.
AGE= age.
SAT = scores of SAT.
SMRT = the score on Mental Rotations Test.
TSMRT = the time spent in seconds by each participant when answering the Mental Rotations Test.
*Treatment p-values are one-tail, all others are two-tail.

Table 44: Test Results of H2d (Continued)

Panel C: ANCOVA Results Using Time Spent as the Dependent Variables

Dependent Variables	Source of Variation	Type III SS	DF	Mean Square	F stat	p value*
TSPRQ6	Corrected Model	5333.715	9	592.635	1.571	0.132
	Intercept	2701.757	1	2701.757	7.163	0.009
	PQ4	433.582	1	433.582	1.150	0.286
	TSPQ4	840.912	1	840.912	2.229	0.138
	PQ6	421.258	1	421.258	1.117	0.293
	AGE	12.316	1	12.316	0.033	0.857
	SAT	78.688	1	78.688	0.209	0.649
	SMRT	613.989	1	613.989	1.628	0.205
	TSMRT	2569.261	1	2569.261	6.812	0.010
	Treatment	270.051	2	135.025	0.358	0.350
	Error	44130.222	117	377.181		
	Total	287047.000	127			
	Corrected Total	49463.937	126			

TSPRQ1 Adjusted R Squared = 0.423
TSPRQ1 Adjusted R Squared = 0.246
TSPRQ1 Adjusted R Squared = 0.397
TSPRQ1 Adjusted R Squared = 0.039
PQ4 = score on practice question 4.
TSPQ4 = time spent in seconds by each participant when answering practice question 4.
PQ6 = score on practice question 6.
AGE= age.
SAT = scores of SAT.
SMRT = the score on Mental Rotations Test.
TSMRT = the time spent in seconds by each participant when answering the Mental Rotations Test
*Treatment p-values are one-tail, all others are two-tail.

Panel D: Bonferroni Pairwise Comparisons for Test H2d

Dependent Variables	(I) Treatment (J) Treatment	Mean Difference (I-J)	Std Error	p value*
TSPRQ1	3-D Displays Tabular Display	26.856	8.260	0.002
	2-D Display	22.099	8.166	0.011
TSPRQ3	3-D Displays Tabular Display	-35.856	10.341	0.001
	2-D Display	-32.976	10.225	0.002
TSPRQ5	3-D Displays Tabular Display	11.120	2.497	<0.001
	2-D Display	9.879	2.468	<0.001

*p-values are one-tail.

5.6 Post Hoc Analysis

As indicated earlier in section 3.7, the 3-D perspective display of DuPont analysis is a newly created display format of financial ratios that has never been empirically tested. For this reason, the study adopted six survey questions from Murthy, Schafer and Fuller (2007) to survey participants' opinions on the usefulness and ease of use of their randomly assigned display format.

The six questions are: 1) using the tables (graphs) was frustrating, 2) the tables (graphs) displayed the task information in a readable format, 3) I found the tables (graphs) useful in how they presented the data for decision making, 4) the tables (graphs) helped me to understand the task data to make a better decision, 5) the tables (graphs) fit the way I needed to view the task information to make better decisions, and 6) overall, I am satisfied with the tables (graphs) in providing the information I needed to complete this task (see Table 9). Participants were asked to select the scale number from one to seven that indicates the extent to which they agree with each of the six questions. (1 = highly disagree, 2 = moderately disagree, 3 = slightly disagree, 4 = neither agree nor disagree, 5 = slightly agree, 6 = moderately agree, 7 = highly agree.).

Chapter 6 will discuss the implications of the results of these survey questions or the opinion from participants in regards to the usefulness and ease of use of their randomly assigned display format.

5.6.1 Trend Analysis Task Post Hoc Analysis

Table 45 indicates that both the participants viewing the tabular display (mean responses 3.023) and the participants viewing the 2-D displays (mean responses 3.725) disagreed that using the table (graphs) was frustrating. Those participants viewing the 3-D perspective display (mean responses 4.690) neither agreed nor disagreed that using the graph was frustrating. Results of one-way ANOVA testing suggest that mean responses of those participants viewing the tabular display ($p < 0.001$), and the mean responses of those participants viewing the 2-D displays ($p = 0.043$), were significantly different from the mean responses of those participants viewing the 3-D perspective display.

Table 45 indicates that the participants viewing the tabular display (mean responses 5.785) slightly agreed that the table displayed the task information in a readable format. Both the participants viewing the 2-D displays (means responses 4.700) and the participants viewing the 3-D perspective display (mean responses 4.380) neither agreed nor disagreed that the graphs displayed the task information in a readable format. Results of one-way ANOVA testing suggest that mean responses of those participants viewing the tabular display were significantly different from the mean responses of those participants viewing the 2-D displays ($p < 0.001$), and from the mean responses of those participants viewing the 3-D perspective display ($p < 0.001$).

Table 45 indicates that both the participants viewing the tabular display (mean responses 5.190) and the participants viewing the 2-D displays (mean responses 5.050) slightly agreed that the table (graphs) was useful in how it presented the data for decision making. Those participants viewing the 3-D perspective display (means responses 4.071) neither agreed nor disagreed that the graph was useful in how it presented the data for

decision making. Results of one-way ANOVA testing suggest that mean responses of those participants viewing the tabular display ($p = 0.006$), and the mean responses of those viewing the 2-D displays ($p = 0.022$), were significantly different from the mean responses of those participants viewing the 3-D perspective display.

Table 45 indicates that both the participants viewing the tabular display (mean responses 5.357) and the participants viewing the 2-D displays (mean responses 5.075) slightly agreed that the table (graphs) helped them understand the task data to make a better decision. Those participants viewing the 3-D perspective display (means responses 4.190) neither agreed nor disagreed that the graph helped them understand the task data to make a better decision. Results of one-way ANOVA testing suggest that mean responses of those participants viewing the tabular display ($p = 0.001$), and the mean response of those participants viewing the 2-D displays ($p = 0.018$), were significantly different from the mean responses of those participants viewing the 3-D perspective display.

Table 45 indicates that the participants viewing the tabular display (mean responses 5.000) slightly agreed that the table fit the way the participants needed to view the task information to make a better decision. Those participants viewing the 2-D displays (means responses 4.450) neither agreed nor disagreed that the graphs fit the way the participants needed to view the task information to make a better decision. However, those participants viewing the 3-D perspective display (means responses 3.761) slightly disagreed that the graphs fit the way the participants needed to view the task information to make a better decision. Results of one-way ANOVA testing suggest that mean responses of those participants viewing the tabular display were significantly different

from the mean responses of those participants viewing the 3-D perspective displays ($p = 0.001$).

Table 45 indicates that the participants viewing the tabular display (mean responses 5.214) slightly agreed that they were satisfied with the table in providing the information they needed to complete the task. Both the participants viewing the 2-D displays (means responses 4.350) and the participants viewing the 3-D perspective display (mean responses 4.095) neither agreed nor disagreed that they were satisfied with the graphs in providing the information they needed to complete the task. Results of one-way ANOVA testing suggest that mean responses of those participants viewing the tabular display were significantly different from the mean responses of those participants viewing the 3-D perspective displays ($p = 0.006$).

Table 45
Mean, Standard Deviation, Range and ANOVA Pairwise Comparison (Mean) of the Survey Questions Assessing Display Usefulness and Ease of Use, for the Trend Analysis Task.

	Tabular Display (n=42)	2-D Displays (n =40)	3-D Display (n =42)	Tabular & 2-D Display Mean Difference	Tabular & 3-D Display Mean Difference	2-D & 3-D Display Mean Difference
"Using the tables (graphs) was frustrating."	3.023 (1.569) 1.00-6.00	3.725 (1.782) 1.00-7.00	4.690 (1.960) 1.00-7.00	n/s	($p < 0.001$)	($p = 0.043$)*
"The tables (graphs) displayed the task information in a readable format."	5.785 (1.200) 2.00-7.00	4.700 (1.697) 1.00-7.00	4.380 (1.974) 1.00-7.00	($p = 0.001$)*	($p < 0.001$)	n/s
"I found the tables (graphs) useful in how they presented the data for decision making."	5.190 (1.485) 1.00-7.00	5.050 (1.518) 1.00-7.00	4.071 (1.839) 1.00-7.00	n/s	($p = 0.006$)	($p = 0.022$)*
"The tables (graphs) helped me understand the task data to make a better decision."	5.357 (1.077) 2.00-7.00	5.075 (1.327) 2.00-7.00	4.190 (1.783) 1.00-7.00	n/s	($p = 0.001$)	($p = 0.018$)*
"The tables (graphs) fit the way I needed to view the task information to make a better decision."	5.000 (1.361) 2.00-7.00	4.450 (1.600) 1.00-7.00	3.761 (1.664) 1.00-7.00	n/s	($p = 0.001$)	n/s
"Overall, I am satisfied with the tables (graphs) in providing the information I needed to complete the task."	5.214 (1.353) 2.00-7.00	4.350 (1.687) 1.00-7.00	4.095 (1.791) 1.00-7.00	n/s	($p = 0.006$)	n/s

n/s = non significant results of the pairwise comparison of means at $p <= 0.05$

5.6.2 Pattern Recognition Task Post Hoc Analysis

Results of one-way ANOVA testing suggest that there was no significant difference between mean responses of the treatment groups in regard to the question whether using the table (graphs) was frustrating (see Table 46).

Table 46 indicates that the participants viewing the tabular display (mean responses 5.860) slightly agreed that the table displayed the task information in a readable format. Both the participants viewing the 2-D displays (means responses 4.829) and the participants viewing the 3-D perspective display (mean responses 4.441) neither agreed nor disagreed that the graphs displayed the task information in a readable format. Results of one-way ANOVA testing suggest that mean responses of those participants viewing the tabular display were significantly different from the mean responses of those participants viewing the 2-D displays ($p = 0.005$), and from the mean responses of those participants viewing the 3-D perspective display ($p < 0.001$).

Table 46 indicates that the participants viewing the tabular display (mean responses 5.395) slightly agreed that the table (graphs) was useful in how it presented the data for decision making. Both the participants viewing the 2-D displays (mean responses 4.804) and the participants viewing the 3-D perspective display (means responses 4.279) neither agreed nor disagreed that the graph was useful in how it presented the data for decision making. Results of one-way ANOVA testing suggest that mean responses of those participants viewing the tabular display were significantly different from the mean responses of those participants viewing the 3-D perspective display ($p = 0.003$).

Table 46 indicates that both the participants viewing the tabular display (mean responses 5.232) and the participants viewing the 2-D displays (mean responses 5.243)

slightly agreed that the table (graphs) helped them understand the task data to make a better decision. Those participants viewing the 3-D perspective display (means responses 4.472) neither agreed nor disagreed that the graph helped them understand the task data to make a better decision. Results of one-way ANOVA testing suggest that mean responses of those participants viewing the tabular display ($p = 0.015$), and the mean responses of those participants viewing the 2-D displays ($p = 0.015$), were significantly different from the mean responses of those participants viewing the 3-D perspective display.

Table 46 indicates that both the participants viewing the tabular display (mean responses 4.906) and the participants viewing the 2-D displays (means responses 4.634) neither agreed nor disagreed that the graphs fit the way the participants needed to view the task information to make a better decision. However, those participants viewing the 3-D perspective display (means responses 3.953) slightly disagreed that the graphs fit the way the participants needed to view the task information to make a better decision. Results of one-way ANOVA testing suggest that mean responses of those participants viewing the tabular display were significantly different from the mean responses of those participants viewing the 3-D perspective displays ($p = 0.008$).

Table 46 indicates that the participants viewing the tabular display (mean responses 5.209) slightly agreed that they were satisfied with the table in providing the information they needed to complete the task. Both the participants viewing the 2-D displays (means responses 4.853) and the participants viewing the 3-D perspective display (mean responses 4.441) neither agreed nor disagreed that they were satisfied with the graphs in providing the information they needed to complete the task. Results of one-

way ANOVA testing suggest that mean responses of those participants viewing the tabular display were significantly different from the mean responses of those participants viewing the 3-D perspective displays ($p = 0.027$).

Chapter Six will discuss the results in detail.

Table 46
Mean, Standard Deviation, Range and ANOVA Pairwise Comparison (Mean) of the Survey Questions on Display Usefulness and Ease of Use, for the Pattern Recognition Task.

	Tabular Display (n =42)	**2-D Displays (n =40)**	**3-D Display (n =42)**	**Tabular & 2-D Display Mean Difference**	**Tabular & 3-D Display Mean Difference**	**2-D & 3-D Display Mean Difference**
"Using the tables (graphs) was frustrating."	3.162 (1.462) 1.00-6.00	3.756 (1.894) 1.00-7.00	3.651 (1.837) 1.00-7.00	n/s	n/s	n/s
"The tables (graphs) displayed the task information in a readable format."	5.860 (1.245) 2.00-7.00	4.829 (1.610) 1.00-7.00	4.441 (2.015) 1.00-7.00	(p = 0.005)*	(p < 0.001)	n/s
"I found the tables (graphs) useful in how they presented the data for decision making."	5.395 (1.311) 1.00-7.00	4.804 (1.720) 1.00-7.00	4.279 (2.027) 1.00-7.00	n/s	(p = 0.003)	n/s
"The tables (graphs) helped me understand the task data to make a better decision."	5.232 (1.377) 2.00-7.00	5.243 (1.462) 1.00-7.00	4.472 (1.964) 1.00-7.00	n/s	(p = 0.015)	(p = 0.015)*
"The tables (graphs) fit the way I needed to view the task information to make a better decision."	4.906 (1.268) 1.00-7.00	4.634 (1.624) 1.00-7.00	3.953 (1.951) 1.00-7.00	n/s	(p = 0.008)	n/s
"Overall, I am satisfied with the tables (graphs) in providing the information I needed to complete the task."	5.209 (1.225) 2.00-7.00	4.853 (1.711) 1.00-7.00	4.441 (1.776) 1.00-7.00	n/s	(p = 0.027)	n/s

n/s = non significant results of the pairwise comparison of means at p <=0.05.

Chapter 6: Discussion

6.1 Summary of Hypothesized Results

Hypothesis one was tested in several parts (1a-1d) using several dependent variables, as was hypothesis two (H2a-H2d). After providing a summary of the results for the two hypotheses a discussion of the results will be provided in 6.2. Table 43 summarizes the information presented in this section.

Hypothesis H1a predicted that participants using a set of 2-D displays would be the most effective in generating hypotheses for what caused the changes in the trend of accounting data when compared to participants using a single 3-D perspective display or participants using a tabular display. This hypothesis has four dependent measures of accuracy − the score on the first question (TAQ1), which asked what the data differences were between years 1 and 4; the score on the second question (TAQ2), which asked what participants perceived to be occurring in the data when going from year 2 to year 3 and year 4; the score on the fifth question (TAQ5), which asked participants to select from a template to indicate the differences in data between years 2 and 4; and the score on the sixth question (TAQ6), which asked participants to select from a template to indicate changes in data when going from year 1 to year 2 to year 3.

Hypothesis H1a was partially supported by the results for TAQ1 and TAQ6. Participants viewing the 2-D display were significantly ($p = 0.001$) more effective than participants viewing the 3-D perspective display in describing the differences between

years 1 and 4 (TAQ1). Participants viewing the 2-D display were also significantly ($p < 0.001$) more accurate than participants viewing the 3-D perspective display in indicating changes in data when going from year 1 to year 2 to year 3 by selecting choices from a template (TAQ6).

Hypothesis H1b predicted that participants using a set of 2-D displays would be the most efficient in generating hypotheses for what caused the changes in trend of accounting data when compared to participants using a single 3-D perspective display or participants using a tabular display. This hypothesis has four dependent measures of efficiency − the time spent in seconds by each participant when answering what the data differences were between years 1 and 4 (TSTAQ1), the time spent in seconds by each participant in describing what was occurring in the data when going from year 2 to year 3 and year 4 (TSTAQ2), the time spent in seconds by each participant when selecting from a template the differences in data between years 2 and 4 (TSTAQ5), and the time spent in seconds by each participant when selecting from a template changes in data when going from year 1 to year 2 to year 3 (TSTAQ6).

Hypothesis H1b was not supported since only one of the four dependent variables significantly increased when participants viewed 2-D displays. Participants viewing the 2-D displays were significantly ($p < 0.001$) more efficient or used less time in seconds than participants viewing the 3-D perspective display when selecting from a template the differences in data between years 2 and 4 (TSTAQ5).

Hypothesis H1c predicted that participants using a set of 2-D displays would be the most effective in an accounting judgment involving estimation of values when compared to participants using a single 3-D perspective display or participants using a

tabular display. This hypothesis has five dependent measures of accuracy − the score on the third question (TAQ3), which asked the participants to estimate what the ROE would be in year 6 (TAQ3) if each of the variables comprising ROE in year 5 doubles; and the scores on parts 'a' 'b' 'c' and 'd' of the fourth question (TAQ4a, b, c, d), which asked the participants to estimate the average of turnover (TAQ4a), profitability (TAQ4b) and leverage (TAQ4c) for the years 1, 2, 4 and 5, and use the estimated average to calculate a new ROE (TAQ4d).

Hypothesis H1c was not supported. Contrary to what hypothesis H1c predicted, participants viewing the 3-D perspective display were more effective ($p = 0.031$) or accurate than those participants viewing the 2-D display in estimating the ROE for year 6 (TAQ3) if each of the variables comprising ROE in year 5 doubles.

Hypothesis H1c was also not supported in that both the participants viewing the tabular display ($p = 0.032$) and the participants viewing the 3-D perspective display ($p = 0.047$) were more effective or accurate than those participants viewing the 2-D displays in estimating the average of leverage for the years 1, 2, 4 and 5 (TAQ4c). On top of the aforementioned result, hypothesis H1c was further not supported by the fact that the participants viewing the tabular display were more ($p = 0.026$) effective or accurate than those participants viewing the 2-D display in estimating the average of turnover, profitability and leverage for the years 1, 2, 4 and 5, and in using the estimated average to calculate a new ROE (TAQ4d).

Hypothesis H1d predicted that participants using a set of 2-D displays would be the most efficient in an accounting judgment involving estimation of values when compared to participants using a single 3-D perspective display or participants using a

tabular display. This hypothesis has two dependent measures of efficiency − the time spent in seconds by each participant when estimating what the ROE would be in year 6 if each of the variables comprising ROE in year 5 doubles (TSTAQ3); and the time spent in seconds by each participant when estimating the average of turnover, profitability and leverage for the years 1, 2, 4 and 5, and using the estimated average to calculate a new ROE (TSTAQ4). Hypothesis H1d was not supported as the MANCOVA F-statistic for the manipulated variable Treatment was not significant ($p = 0.751$) using Pillai's Trace.

The remainder of this section summarizes the second hypothesis. Hypothesis H2a predicted that participants using a single 3-D perspective display would be the most effective in recognizing patterns of accounting data when compared to participants using a set of 2-D displays or participants using a tabular display. This hypothesis has two dependent measures of accuracy − the score on the second question (PRQ2), which asked the participants to separate companies one through six into two groups based on similar financial characteristics; and the score on the fourth question (PRQ4), which asked the participants to select one of the six companies if the goal is to have high profitability, high turnover but low leverage at the same time.

Hypothesis H2a was fully supported, in that significant results were found for both dependent variables. Participants viewing the 3-D perspective display not only were significantly more effective or accurate than those participants viewing the tabular display ($p < 0.001$) but were also more effective than participants viewing the 2-D display ($p < 0.001$) when it came to separating companies one through six into two groups based on similar financial characteristics (PRQ2). Participants viewing the 3-D perspective display were also significantly more effective or accurate than those

participants viewing the tabular display ($p = 0.007$) and those viewing the 2-D displays ($p = 0.016$) in selecting which of the six companies had high profitability, high turnover but low leverage at the same time (PRQ4).

Hypothesis H2b predicted that participants using a single 3-D perspective display would be the most efficient in recognizing patterns of accounting data when compared to participants using a set of 2-D displays or participants using a tabular display. This hypothesis has two dependent measures of efficiency − the time spent in seconds by each participant in separating companies one through six into two groups based on similar financial characteristics (TSPRQ2); and the time spent in seconds by each participant in selecting one of the six companies if the goal is to have high profitability, high turnover but low leverage at the same time (TSPRQ4).

Hypothesis H2b was partially supported. Participants viewing the 3-D perspective display were only significantly more efficient ($p = 0.019$) than those participants viewing the 2-D displays when selecting which of six companies had highest profitability, highest turnover but lowest leverage at the same time (TSPRQ4). There was not a significant difference between 3-D display and tabular format, nor was there support for H2b using the dependent measure of efficiency TSPRQ2 − the time spent in seconds by each participant in separating companies one through six into two groups based on similar financial characteristics.

Hypothesis H2c predicted that participants using a single 3-D perspective display would be the most effective in generating hypotheses for what caused the emerged patterns when compared to participants using a set of 2-D displays or participants using a tabular display. This hypothesis has four dependent measures of accuracy − the score on

the first question (PRQ1), which asked the participants what the data differences were between companies one and six; the score on the third question (PRQ3), which asked the participants to describe the pattern of financial ratios they were seeing in group one compared to group two; the score on the fifth question (PRQ5), which asked the participants what the data differences were between companies four and six by selecting choices from a template; and the score on the sixth question (PRQ6), which asked the participants to describe the pattern of financial ratios they were seeing in group two compared to group one by selecting choices from a template. Hypothesis H2c was not supported as the MANCOVA F-statistic for the manipulated variable Treatment was not significant ($p = 0.506$) using Pillai's Trace.

Hypothesis H2d predicted that participants using a single 3-D perspective display would be the most efficient in generating hypotheses of what caused the emerged patterns when compared to participants using a set of 2-D displays or participants using a tabular display. This hypothesis has four time measures related to the dependent measures identified for H2c; they are TSPRQ1, TSPRQ3, TSPRQ5 and TSPRQ6.

Hypothesis H2d had mixed results. No support was found for hypothesis H2d using TSPRQ6 – time spent by each participant when describing the pattern of financial ratios they were seeing in group two compared to group one by selecting choices from a template. Contrary support was found using TSPRQ1 (time spent by each participant in describing what the data differences were between companies one and six) and TSPRQ5 (time spent by each participant in describing what the data differences were between companies four and six by selecting choices from a template). With TSPRQ1 both the participants viewing the tabular display ($p = 0.002$) and the participants viewing the 2-D

displays ($p = 0.012$) were more efficient, or used less time, than participants viewing the 3-D perspective display when describing the data differences between companies 1 and 6. Again, with TSPRQ5 both the participants viewing the tabular display ($p < 0.001$) and the participants viewing the 2-D displays ($p < 0.001$) were more efficient, or used less time, than participants viewing the 3-D perspective display when describing the data differences between companies 4 and 6 by selecting choices from a template.

However, the time spent by each participant in describing the pattern of financial ratios they were seeing in group one compared to group two (TSPRQ3) support hypothesis H2d. Participants viewing the 3-D perspective display were not only significantly more efficient than participants viewing the tabular display ($p = 0.001$), but also significantly more efficient, or used less time, than participants viewing the 2-D displays ($p = 0.002$) when describing the pattern of financial ratios in group one.

Table 47 and Table 48 provide a summary of the results of hypotheses H1 and H2, respectively.

Table 47
Summary of the Results of the Trend Analysis Task

Hypothesis	Dependent Variables	Test*	Findings
H1a			
Subjects using 2-D displays	Score on the 1st question (TAQ1)	S	2-D more accurate than 3-D
will be the most effective	Score on the 2nd question (TAQ2)	N/S	
(accuracy) in generating	Score on the 5th question (TAQ5)	N/S	
hypotheses for what caused	Score on the 6th question (TAQ6)	S	2-D more accurate than 3-D
the changes in the trend of			
accounting data			
H1b			
Subjects using 2-D displays	Time spent on the 1st question (TSTAQ1)	N/S	
will be the most efficient	Time spent on the 2nd question (TSTAQ2)	N/S	
(less time) in generating	Time spent on the 5th question (TSTAQ5)	S	2-D more efficient than 3-D
hypotheses for what caused	Time spent on the 6th question (TSTAQ6)	N/S	
the changes in the trend of			
accounting data			
H1c			
Subjects using 2-D displays	Score on the 3rd question (TAQ3)	N/S*	3-D more accurate than 2-D
will be the most effective	Score on the 4th 'a' question (TAQ4a)	N/S	
(accuracy) in the estimation	Score on the 4th 'b' question (TAQ4b)	N/S	
of values	Score on the 4th 'c' question (TAQ4c)	N/S*	Table,3-D more accurate
	Score on the 4th 'd' question (TAQ4d)	N/S*	Table more accurate than 2-D
H1d			
Subjects using 2-D displays	Time spent on the 3rd question TSTAQ3)	N/S	
will be the most efficient	Time spent on the 4th question (TSTAQ4)	N/S	
(less time) in the estimation			
of values			

TAQ1 = score on the 1st question which asked the participants what the data differences were between years 1 and 4.
TAQ2 = score on the 2nd question which asked what the participants perceived to be occurring in the data when going from year 2 to year 3 and year 4.
TAQ3 = score on the 3rd question which asked the participants to estimate what the ROE would be in year 6 if each of the variables comprising ROE in year 5 doubles.
TAQ4'a' to 'd' = score on question which asked the participants to estimate the average of turnover (TAQ4a). profitability (TAQ4b), and leverage (TAQ4c) for the years 1, 2, 4, and 5, and use the estimated average to calculate a new ROE (TAQ4d).
TAQ5 = score on the 5th question which asked the participant to select from a template to indicate the differences between year 2 and 4.
TAQ6 = score on the 6th question which asked the participants to select from a template to indicate changes in data when going from year ` to year 2 to year 3.

S = results of testing of the hypothesis indicate that the Mean difference is significant at the 0.05 level and thus supporting the hypothesis in the predicted direction.
N/S = results of testing of the hypothesis indicate that the Mean difference is insignificant at the 0.05 level.
N/S* = results of testing of the hypothesis indicate that the Mean difference is significant at the 0.05 level but in the opposite direction of the predicted.

Table 48
Summary of the Results of the Pattern Recognition Task

Hypothesis	Dependent Variables	Test	Findings*
H2a		*	
Subjects using 3-D	Score on the 2nd question (PRQ2)	S	3-D more accurate than table, 2-D
display will be the most	Score on the 4th question (PRQ4)	S	3-D more accurate than table, 2-D
effective (accuracy) in			
recognizing patterns of			
accounting data			
H2b			
Subjects using 3-D	Time spent on the 2nd question (TSPRQ2)	N/S	
display will be the most	Time spent on the 4th question (TSPRQ2)	S	3-D more efficient than 2-D
efficient (less time) in			
recognizing patterns of			
accounting data			
H2c			
Subjects using 3-D	Score on the 1st question (PRQ1)	N/S	
display will be the most	Score on the 3th question (PRQ3)	N/S	
effective (accuracy) in	Score on the 5th question (PRQ5)	N/S	
generating hypotheses for	Score on the 6th question (PRQ6)	N/S	
what caused the emerged			
patterns			
H2d			
Subjects using 3-D	Time spent on the 1st question (TSPRQ1)	N/S*	Table, 2-D more efficient than 3-D
display will be the most	Time spent on the 3th question (TSPRQ3)	S	3-D ,more efficient than 2-D Table
efficient (less time) in	Time spent on the 5th question (TSPRQ5)	N/S*	Table, 2-D more efficient than 3-D
generating hypotheses for	Time spent on the 6th question (TSPRQ6)	N/S	
what caused the emerged			
patterns			

PRQ1 = score on the 1st question which asked the participants what the data differences were between companies one and six.
PRQ2 = score on the 2nd question which asked the participants to separate companies one through six into two groups based on similar financial characteristics.
PRQ3= score on the 3rd question which asked the participants to describe the pattern of financial ratios they were seeing in group one compared to group two.
PRQ4 = score on the 4th question which asked the participants to select one of the six companies if the goal is to have high profitability, high turnover, but low leverage at the same time.
PRQ5 = score on the 5th question which asked the participants what the data differences were between companies four and six by selecting choices from a template.
PRQ6 = score on the 6th question which asked the participants to describe the pattern of financial ratios they were seeing in group two compared to group one by selecting choices from a template.

S = results of testing of the hypothesis indicate that the Mean difference is significant at the 0.05 level and thus supporting the hypothesis in the predicted direction.
N/S = results of testing of the hypothesis indicate that the Mean difference is insignificant at the 0.05 level.
N/S* = results of testing of the hypothesis indicate that the Mean difference is significant at the 0.05 level but in the opposite direction of the predicted.

6.2 Discussion of Results of Hypothesis H1

Hypothesis H1 predicted that participants using a set of 2-D displays would be most effective in two trend analysis tasks – generating hypotheses for what caused the changes in the trend of accounting data, and estimating values – when compared to participants using a single 3-D perspective display or participants using a tabular display. In relation to the two different trend analysis tasks, hypothesis H1 employed a total of nine dependent measures to test for the differences in accuracy between the responses of participants assigned to different treatment groups. Further, hypothesis H1 predicted that participants using a set of 2-D displays would be most efficient in generating hypotheses for what caused the changes in the trend of accounting data and in estimating values, when compared to participants using a single 3-D perspective display or participants using a tabular display. Hypothesis H1 employed a total of six dependent measures of efficiency to test for the differences in efficiency between participants of different treatment groups.

6.2.1 Discussion of Results of Hypothesis H1a

Hypothesis H1a stipulates that those participants viewing the 2-D displays will be the most effective or accurate in generating hypotheses for what caused the changes in the trend of accounting data when compared to those participants viewing the tabular display or those participants viewing a 3-D perspective display. Viewing 2-D displays resulted in significantly more effective or accurate hypotheses by participants for two (TAQ1 and TAQ6) of the four dependent measures. Given the fact that there is no

significant difference in other dependent measures of accuracy between the treatment groups, hypothesis H1a was partially supported.

Additional analysis of the results of hypothesis H1a show some facts worthy of discussion. The four dependent measures of accuracy of hypothesis H1a can be divided into two groups: the first group contains the first (TAQ1) and the fifth question (TAQ5), while the second group contains the second (TAQ2) and the sixth question (TAQ6).

The first question of the trend analysis task asked the participants to use short sentences to describe the data differences between years 1 and 4. Similarly, the fifth question asked the participants what the data differences were between years 2 and 4 by selecting choices from a template. Both questions involved a time period of two years. Woods (1990) would identify the tasks from TAQ1 and TAQ5 as simple tasks because participants did not need to perform a prior processing step or calculation to create an intermediate value and use that intermediate value to derive the final outcome. A simple comparison between years 1 and 4 or years 2 and 4 is sufficient to answer the first and fifth questions correctly.

Though the study is unable to explain why there was no significant differences between those participants viewing the 2-D displays and those viewing the tabular display on all the four dependent measures of accuracy (TAQ1), (TAQ2), (TAQ5), and (TAQ6), the study suggests a possible explanation for why those participants viewing the 2-D displays were significantly ($p = 0.001$) more accurate than participants viewing a single 3-D perspective display on the first question (TAQ1) but not on the fifth question (TAQ5).

Assuming the template of choices benefited all participants equally when answering the fifth question, those participants viewing the 3-D perspective display possibly became more accurate by being more proficient in reading the 3-D perspective display, which was new to them at the beginning of the experiment. Triffett and Trafton (2006) described how meteorologists spatially transformed the position of a low pressure system toward a certain direction even though the actual movement of the low pressure system was not explicitly shown in a graph. In this study, once participants became more proficient in reading the 3-D perspective display, they were possibly able to mentally create a slope line while generating hypotheses for what the data differences were between years 2 and 4. The aforementioned explanation contributes toward the understanding of why there were no significant differences in the mean score of the fifth question (TAQ5) between the three treatment groups. The next few paragraphs discuss the second group of questions, TAQ2 ant TAQ6.

The second question (TAQ2) of the trend analysis task asked the participants to write short sentences about what they perceived to be occurring in the data when going from year 2 to year 3 and year 4; similarly the sixth question (TAQ6) of the trend analysis task asked the participants to select from a template to indicate changes in data when going from year 1 to year 2 to year 3. Both questions involved a time period of three years. Woods (1990) suggests that if a task's final outcome depends on the result of a prior process, calculation, and step, then the task is a complex task. Both the second and the sixth question are complex tasks because participants had to discern what was happening between the first two years before correctly hypothesizing what caused the changes in the trend of accounting data across three time periods (from year 1 to year 2 to

year 3 for example). It is interesting to note that the covariate mental workload (MW) was not significantly associated with the dependent measure of TAQ1 ($p = 0.094$) and only marginally significantly associated with the dependent measure of TAQ5 ($p = 0.050$), but MW was significantly associated with the dependent measure of TAQ2 ($p=0.011$) and TAQ6 ($p = 0.023$). These results support that the second and sixth question of the trend analysis task were both complex tasks. There was no significant difference in the covariate 'mental workload' between the three treatment groups.

Benefiting from the message richness of the slope of line, which portrays the trend relationship explicitly, those participants viewing the 2-D displays should have outperformed participants of the other two treatment groups especially on a complex task involving the analysis of multi time period data. While there was no significant difference in accuracy between the three treatment groups on the second question (TAQ2), participants viewing the 2-D displays were significantly more accurate than participants viewing a 3-D perspective display ($p < 0.001$) in situations such as question six, where a template of choices was provided.

Summarizing the above discussion of hypothesis H1a, the following is offered. For simple tasks, those participants viewing the 2-D displays can sometimes be more accurate than those viewing the 3-D perspective display. However, participants of the 3-D perspective display were able to improve accuracy enough (possibly due to the ability to spatially transform information) to be on the same footing as participants viewing the 2-D displays when performing a simple task involving two time periods. On the other hand, for the complex task involving three time periods (what was occurring in the data when going from year 1 to year 2 to year 3) viewers of the 2-D displays only out

performed participants viewing the 3-D perspective display when a template of choices was available.

Future research should employ methods like verbal protocol analysis to discern what causes the improvement in accuracy in the 3-D perspective display and to understand why participants viewing the tabular display can be more accurate than those viewing the 2-D display when generating hypotheses for what caused the changes in the trend of the data.

6.2.2 Discussion of Results of Hypothesis H1b

Hypothesis H1b stipulates that those participants viewing the 2-D displays will be the most efficient or use less time (in seconds) in generating hypotheses for what caused the changes in the trend of accounting data when compared to those participants viewing the tabular display or those participants viewing a 3-D perspective display. Hypothesis H1b uses the time related to answering the four questions asked to test hypothesis H1a. In terms of these four dependent measures of efficiency, those participants viewing the 2-D displays were significantly more efficient or used less time than those participants viewing the 3-D display in only one dependent measures of efficiency (TSTAQ5).

Given the fact that there are no significant differences in the other dependent measures of efficiency between the treatment groups, hypothesis H1b was not supported. The results of H1b suggest that those participants viewing the 2-D displays were not more efficient than participants viewing the tabular display or participants viewing the 3-D perspective display in generating hypotheses for what caused the changes in the trend

of accounting data, the lack of results suggests no one display type is more efficient than the others.

6.2.3 Discussion of Results of Hypothesis H1c

Hypothesis H1c stipulated that those participants viewing the 2-D displays would be the most effective or accurate in an accounting judgment involving estimation of values when compared to those participants viewing a tabular display or those participants viewing a 3-D perspective display. There was no support for hypothesis H1c using any of the five dependent measures tested. In fact, for the dependent measure TAQ3, participants viewing the 3-D perspective display were more effective than those participants viewing the 2-D displays in estimating what the ROE would be in year 6 if each of the variables comprising ROE in year 5 doubles. This same result was also found for the dependent measure TAQ4c, which asked participants to estimate the average of leverage for the years 1, 2, 4, and 5, both the participants viewing the tabular display and the participants viewing the 3-D perspective display were more effective than those participants viewing the 2-D displays. Lastly, for the dependent variables TAQ4d participants viewing the tabular display were more effective than those participants viewing the 2-D displays.

Though hypothesis H1c was not supported, it is interesting to notice that when performing a trend analysis task that involves the estimation of values, those participants viewing a static three dimensional perspective display of DuPont analysis, which does not have a line of slope connecting data points, can be more effective, or accurate, than those participants viewing the 2-D displays with the line of slope explicitly shown. Dull

and Tegardan (1999) found that participants using three-dimensional line graphs of accounting wealth, momentum and impulse that could be rotated provided the most accurate predictions of a company's future wealth compared to two-dimensional or three dimensional fixed line graphs of equivalent information. It seems that neither the richness of message flags provided by the slope of line (Pinker 1990) nor the smooth continuation of elements as portrayed in a 2-D line displays (Pinker 1990) benefited those participants viewing the 2-D displays in an accounting judgment involving estimation of values. The results of this study indicate that 3-D displays may provide a better fit for such complex tasks, providing a basis for future research investigating whether other types of static or rotatable three-dimensional displays of accounting information that do not have the slope of line connecting data points can improve participant's accuracy of prediction or estimation of values.

6.2.4 Discussion of Results of Hypothesis H1d

Hypothesis H1d stipulated that those participants viewing the 2-D displays would be the most efficient or use less time in an accounting judgment involving the estimation of values when compared to those participants viewing a tabular display or those participants viewing a 3-D perspective display. Hypothesis H1d has two dependent measures of efficiency (TSTAQ3 and TSTAQ4). Contrary to prediction, there was no significant difference between treatment groups in terms of time used in a trend analysis task involving an accounting task that requires an estimation of values. Hypothesis H1d was not supported, indicating that viewing a 2-D display format did not command an

advantage or competitive edge over other participants viewing a different type of display format when performing a trend analysis task involving the estimation of values.

6.3 Discussion of Results of Hypothesis H2

Hypothesis H2 predicted that participants using a 3-D perspective display would be most effective in pattern recognition tasks – recognizing patterns of accounting data and generating hypotheses for what caused the emerged patterns of accounting data – when compared to participants using a set of 2-D displays or participants using a tabular display. In relation to the two pattern recognition tasks, hypothesis H2 employed a total of six dependent measures to test for the differences in accuracy of the responses of participants assigned to different treatment groups. Further, hypothesis H2 also predicted that participants using a 3-D perspective display would be most efficient in recognizing patterns of accounting data and generating hypotheses for what caused the emerged patterns of accounting data when compared to participants using a set of 2-D displays or participants using a tabular display. Hypothesis H2 employed a total of six dependent measures to test for the differences in efficiency between participants of different treatment groups.

6.3.1 Discussion of Results of Hypothesis H2a

Hypothesis H2a stipulated that participants viewing the 3-D perspective display would be the most effective or accurate in recognizing patterns of accounting data when compared to those participants viewing a tabular display or those participants viewing the 2-D displays. Hypothesis H2a has two dependent measures of accuracy – separating

companies one through six into two groups based on similar financial characteristics (PRQ2), and selecting one of the six companies if the goal is to have high profitability, high turnover, but low leverage at the same time (PRQ4). In terms of both the dependent measures of accuracy, those participants viewing the 3-D perspective display were significantly more effective or accurate than those participants viewing the tabular display and those participants viewing the 2-D displays.

The implication of having hypothesis H2a fully supported is interesting. Findings of the study have empirically demonstrated that the emergent features portrayed by a 3-D perspective display, as proposed by the Proximity Compatibility Principle (PCP), results in greater decision making accuracy for accounting judgments involving the recognition of patterns in accounting data. Based on the findings of hypothesis H2a, it can be concluded that for accounting tasks involving pattern recognition, the 3-D perspective display (newly developed by the study) has the potential to improve decision accuracy.

6.3.2 Discussion of Results of Hypothesis H2b

Hypothesis H2b stipulated that participants viewing the 3-D perspective display would be the most efficient or use less time in recognizing patterns of accounting data when compared to those participants viewing a tabular display or those participants viewing the 2-D displays. For the pattern recognition task of selecting which of six companies has the highest profitability, highest turnover, but lowest leverage at the same time (PRQ4), participants viewing the 3-D perspective display were significantly more efficient or used less time (in seconds) than those participants viewing the 2-D displays.

However, support was not found for H2b using the dependent variable PRQ2, thus hypothesis H2b was partially supported.

Partial support of hypothesis H2b implies that participants viewing the 3-D perspective display can potentially be more efficient or use less time in forming their decision. The findings indicate that there could be circumstances where decision makers will use less time in decision making by using higher dimension display formats such as the 3-D perspective display.

The joint implication of having support for hypotheses H2a and H2b for the pattern recognition task is important. Those participants viewing the 3-D perspective display were not only being more effective or accurate, but also more efficient than those participants viewing the 2-D displays in selecting one of the six companies if the goal is to have high profitability, high turnover, but low leverage at the same time. The results suggest that, for accounting tasks involving the recognition of patterns in data, a 3-D perspective display not only can improve decision maker's accuracy but can also result in less time being used by decision makers in forming their conclusions.

6.3.3 Discussion of Results of Hypothesis H2c

Hypothesis H2c stipulated that participants viewing the 3-D perspective display would be the most effective (accurate) in generating hypotheses for what caused the emerged patterns when compared to those participants viewing the tabular display or those participants viewing the 2-D displays. Contrary to prediction, there was no significant difference between treatment groups in term of accuracy in generating hypotheses for what caused the emerged pattern.

The implication from the lack of support for hypothesis H2c is that participants viewing a 3-D perspective display do not command an advantage or competitive edge over other participants viewing a different type of display format when performing a pattern recognition task involving the generation of hypotheses for what caused the emerged pattern.

6.3.4 Discussion of Results of Hypothesis H2d

Hypothesis H2d stipulated that participants viewing the 3-D perspective display would be the most efficient or use less time in generating hypotheses for what caused the emerged patterns when compared to those participants viewing the tabular display or those participants viewing the 2-D displays. Hypothesis H2d has mixed results. For two of the dependent variables (TSPRQ1 and TSPRQ5) those participants viewing the tabular display and those participants viewing the 2-D displays used significantly less time than those participants viewing the 3-D perspective display. However, when describing the pattern of financial ratios in group one when compared to group two, participants viewing the 3-D perspective display were significantly more efficient than those participants viewing the tabular display or those viewing the 2-D displays. With such mixed results it is difficult to form any conclusions concerning the use of displays to increase the effectiveness of individuals in generating hypotheses for what caused the emerged patterns.

As a conclusion, the results of the study provide some support for Cognitive Fit Theory. The results support the need for fit between task and display since it was found that a single method of display was not suitable for all types of tasks.

6.4 Comments on the Implication of Significant Covariates

Table 43 shows that the covariates that are significantly associated with the dependent variables of H1a-d and H2a-d include: scores of practice question one (PQ1), scores of practice question three (PQ3), scores of practice question four (PQ4), time spent by each participant when answering practice question four (TSPQ4), time spent by each participant when answering practice question five (TSPQ5), time spent by each participant when answering practice question six (TSPQ6), age of participants (AGE), accounting related working experience (ARWE), highest education (HE), student status (SS), scores of SAT (SAT), mental workload (MW), scores on mental rotations test (SMRT), and the time spent by each participant when answering Mental Rotations Test (TSMRT).

The significant covariates can be grouped into two categories: 1) scores of practice questions and time spent in answering the practice questions (PQ1, PQ3, PQ4 TSPQ4, TSPQ5, and TSPQ6), and 2) individual characteristics (AGE, SS, HE, ARWE, SAT, MW, SMRT, and TSMRT). In the current study, the lack of theory and empirical research prevented hypothesizing relationships between the covariates and visualization; however, the significance of the covariates in the various models indicates that these covariates should be considered in future research on visualization.

6.5 Discussion on the Responses of the Survey Questions

In this study, participants responded to a set of survey questions regarding their reactions to the display format they were presented both for the trend analysis task and the pattern recognition task (Table 40 and 41 respectively). Despite the differences in the types of tasks, it is interesting to note that the participants of each treatment group responded to the survey questions in a similar fashion.

It is also interesting to note that participants viewing the 3-D perspective display in both tasks did not respond negatively towards the survey questions. Tables 47 and 48 show that those participants viewing the 3-D perspective display in both tasks responded to four out of the six survey questions by selecting the scale number four (neither agree nor disagree). Both the participants viewing the 3-D perspective display in the trend analysis task and the pattern recognition task slightly disagreed that the graphs fit the way the participants needed to view the task information to make a better decision. It is understandable that viewers of the 3-D perspective display prefer to use tabular display or 2-D displays, which are more familiar to them. Unlike their counterparts in the trend analysis task who neither agreed nor disagreed that using the graphs was frustrating, participants of the 3-D perspective display in the pattern recognition task actually disagreed that using the graphs was frustrating.

The results seem to show that participants viewing the 3-D perspective display did not overwhelmingly dislike this newly created presentation format. The lack of resistance to the 3-D perspective presented in the experiment indicates the potential for researchers to develop three-dimensional presentations of accounting information that are acceptable to users and can be useful to decision making.

6.6 Contribution

The study contributes to the literature by extending the Cognitive Fit Theory in the context of a single task type, namely the spatial task. Prior studies focused on the study of the effects of tabular or two-dimensional graphical displays upon the symbolic or spatial tasks. The study extends the literature by using two and three-dimensional displays to demonstrate how Cognitive Fit Theory applies in different types of spatial tasks. Findings of hypothesis H1a suggest that 2-D displays are suitable for the spatial task of generating hypotheses for what caused the changes in the trend of accounting data, while findings of hypothesis H1c and hypothesis H2a suggest that 3-D perspective display is suitable for other types of spatial tasks such as the prediction of values and recognition of patterns in accounting data, respectively. Findings of the study demonstrate that use of Cognitive Fit Theory is not limited to examining the relationships between tabular and two-dimensional formats and spatial tasks.

The study contributes to the literature by demonstrating that even without the line of slope connecting data points, multidimensional visual display of complex multidimensional data such as the DuPont analysis can result in greater estimation accuracy of ROE. Prior literature suggests that viewers of 2-D line graphs benefit from richness of message flags provided by the slope of the line (Pinker 1990), which explicitly shows the smooth continuation of elements in a 2-D line display (Pinker 1990). However, findings of hypothesis H1c suggest that those participants viewing a static three-dimensional perspective display of DuPont analysis (without a line of slope connecting data points) can be more effective or accurate than those participants viewing the 2-D displays (with the line of slope explicitly shown) while performing a trend

analysis task that involves the estimation of values of ROE. However, it should be noted that results of hypothesis H1a show that participants viewing the 2-D displays do benefit from the slope of the line (which shows continuation of elements) when performing the spatial task of generating hypotheses for what caused the changes in the trend of accounting data.

Results of the study demonstrate the importance of the fit between presentation format and task, consistent with the tenets of Vessey's Cognitive Fit Theory. Whenever there is a fit between the presentation format and the task, performance will be enhanced regardless of whether the task-related information is explicitly or implicitly shown in the display. The findings of hypothesis H1a suggest that 2-D displays are suitable for the spatial task of generating hypotheses for what caused the changes in the trend of accounting data, because the slope of line in the 2-D plane explicitly showed the changes in the accounting data through changes of the slope. Additionally, the findings of hypothesis H2a suggest that the 3-D perspective display is suitable for the spatial tasks of recognition of patterns in accounting data, because the emergent pattern in the 3-D plane explicitly shows the patterns of accounting data through the relative positioning of the data points in the 3-D plane.

The most interesting finding of the study is the fact that a presentation format can be a better fit for a task even if the information is not explicitly shown in the presentation format. As the findings of hypothesis H1c indicate, while performing a trend analysis task that involves the estimation of values of ROE, participants viewing the 3-D perspective display can be more effective or accurate than those participants viewing the 2-D displays, despite the fact that only the 2-D display has the line of slope explicitly

showing the changes in the accounting data. According to Woods (1990), the spatial task of estimation of values of ROE is a complex task as decision makers need to calculate the values of turnover, profitability, and leverage first before they can estimate the values of ROE. A possible explanation of why participants viewing the 3-D perspective can be more effective in a complex spatial task involving the estimation of values of ROE is the perceptual proximity of having turnover, profitability, leverage and ROE in one single display which facilitates cognitive integrative processing as suggested by Carswell (1992).

Tegarden (1999) indicates the purpose of visualization is not to replace quantitative analysis but to allow decision makers to identify trends, patterns anomalies and relationships in data. AICPA Professional Standards (2007) requires the use of analytical procedures in the planning and overall review stages of all audits. (AU section 329). The results of the study clear demonstrate that student participants can benefit from visualization techniques such as the 2-D displays and the 3-D perspective display while performing various spatial tasks involving DuPont analysis. DuPont analysis is a commonly used analytical procedure in the audit and financial world. If student subjects can benefit from the use of visualization techniques, there is no reason to expect auditors would not also benefit from the use of a proper display format to reason with the combination of all cues, properly generate an accurate hypothesis to explain the causal agent underlying the pattern and relationships among pieces of financial information. The results of this study thus have practical implications for external auditors who must perform analytical review procedures at the beginning of the audit and also during the audit. Specifically, for such analytical review procedures, the results of this study

suggest that visualization techniques such as 2-D line graphs can assist auditors in generating hypotheses for what caused the changes in the trend of accounting data, while the 3-D perspective display can help auditors in identifying patterns within the client's accounting data and in estimating new audit-relevant values .

The 3-D perspective display of the DuPont Analysis, newly developed in this study, is the first of its kind in accounting literature. The concept of representing a company by a point in a three dimensional space, if each company was defined by only three accounting ratios, was first discussed by Altman et al. (1974). The study is the first research that technically develops a 3-D perspective display to represent a company by a point in a three-dimensional space in terms of the DuPont analysis.

This study contributes to the accounting literature by applying findings from human factors research in an accounting context. The study applied the Proximity Compatibility Principle (PCP) in an accounting context and demonstrates how a 3-D perspective display can integrate information from the X, Y, Z axes to help decision makers invoke configural information processing. Findings of hypothesis H2a demonstrate that the emergent features portrayed in a 3-D perspective display could actually benefit decision makers. Specifically, findings of hypothesis H2a demonstrate that those participants viewing the 3-D perspective display not only were more accurate but could also be more efficient (in some cases), compared to those decision makers viewing the tabular display or those decision makers viewing the 2-D displays, when performing the pattern recognition task.

Participants in the Dull and Tegarden (1999) experiment improved their decision accuracy by using the 3-D displays of line graphs but did not use less time in forming

their conclusions when using the higher dimensional display format. A minor contribution of the study is demonstrating that decision makers using higher dimensional displays of accounting data can improve accuracy and efficiency at the same time. Findings of hypothesis H2a and H2b suggest that participants viewing the 3-D perspective display were not only being more effective or accurate, but also more efficient than those participants viewing the 2-D displays in selecting one of the six companies if the goal is to have high profitability, high turnover, but low leverage at the same time.

Another minor contribution to the literature is the results of the survey questions on the usefulness and satisfaction of the users of the 3-D perspective display. Analysis of the responses to the survey questions show that users of the 3-D perspective display did not negatively or overwhelmingly dislike this newly created presentation format. Participants of the study were able to learn and use a display format that they had never seen before – the 3-D perspective display of DuPont analysis.

6.7 Limitations

The study has several limitations. Student subjects were used as surrogates for real world auditors. As discussed in section 3.8, student subjects are suitable for the task of the study. However, using real world auditors as participants can increase the external validity of the findings of the study.

As discussed in section 4.34, participants in both the 2-D and 3-D treatment conditions had to scroll up and down the screen when answering questions, since the display and the response area did not fit on one screen. Participants viewing the tabular display did not need to scroll up and down the screen when answering questions. The study acknowledges the need to scroll up and down the screens as a limitation.

As discussed in section 5.61 the manipulation check did not operationalize as intended. There was a high error rate in participants' responses to the manipulation check question. Such a high error rate was due to the fact that participants had to rely on their memory when answering the manipulation check questions about display. Participants did not look at the displays while answering the manipulation check. The study acknowledges this as a limitation.

The study employed a single set of data to test the effects of the manipulation of the presentation format in the trend analysis or pattern recognition task. It is beneficial to have several sets of data of similar complexity to test the effects of the manipulation of the presentation format. The study acknowledges the lack of multiple sets of data to test for the effects of the manipulation of the presentation format.

6.8 Future Research

There are ample opportunities for future research in visualization of accounting information. The study employed static 3-D perspective display. Future research can investigate the effects of animation and rotation on those decision makers viewing the 3-D perspective display.

Future research should employ real world auditors and study the effect of 3-D perspective display upon novice auditors. An interesting question for future research is whether 3-D perspective display can improve the decision quality of the novice auditors when compared to experienced auditors.

Future research can develop other types of 3-D display of accounting information. Future research should also investigate the question of what type of three-dimensional display is suitable for which type of accounting information. During the process of developing new three-dimensional displays of multidimensional accounting data, it is crucial to test the newly developed display format with several sets of data of similar nature. Further, future research can investigate the effect of split attention, since the participants viewing the 2-D displays or 3-D perspective display had to scroll up and down the screen while performing the assigned task, by having a control group (of a display format) without the need to scroll up and down the screen.

Future research should employ verbal protocol to record the mental status of participants viewing different types of display formats. It is important to understand how the participants learn, teach themselves and improve in their performance while viewing their randomly assigned display format.

References

Altman, E. I., M. Margaine, M. Schlosser, and P. Vernimmen. 1974. Financial and statistical analysis for commercial loan evaluation: A French experience. *The Journal of Financial and Quantitative Analysis* 9 (2):195-211.

American Institute of Certified Public Accountants. *Professional Standards* (2007, June 1). New York: AICPA.

_____. 1978. *The Cohen Commission on Auditors' Responsibilities: Report, Conclusions, and Recommendations*. New York: AICPA.

_____. 1987. *The Treadway Commission on Fraudulent Financial Reporting*. New York: AICPA.

Ashton, R. H. 1974 An experimental study of internal control judgments. *Journal of Accounting Research* (Spring): 143-157.

Bedard, J. C., and S. F.Biggs. 1991. Pattern recognition, hypotheses generation, and auditor performance in an analytical task. *The Accounting Review* 66 (3): 662-642.

Bennett, K. B., and J. M. Flach. 1992. Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors* 34: 513-533.

Benford, T.L. 2000. Determinants of Audit Performance: An Investigation of Task/Technology Fit and Mental Workload. (Doctoral Dissertation, University of South Florida, 2000).

Bertin, J. 1981. *Graphics and Graphic Information Processing.* Berlin: Walter de Gruyter & Co.

Bierstaker, J. L., J. C. Bedard, and S. F. Biggs. 1999. The role of problem representation shifts in auditor decision processes in analytical procedures. *Auditing: A Journal of Practice and Theory* 18 (1): 18-36.

Bogacz, S., and J. G. Trafton. 2005. Understanding dynamic and static displays: Using images to reason dynamically. *Cognitive Systems Research* 6 (4): 312-319.

Bray, J. H., and S. E. Maxwell. 1985. *Multivariate analysis of variance*. Sage University paper series on quantitative applications in the social science, 07-054. Newbury Park CA: Sage University.

Brown, C. E., and I. Solomon. 1990. Auditor configural information processing in control risk assessment. *Auditing: A Journal of Practice & Theory* 9 (3): 17-37.

_____. 1991. Configural information processing in auditing: The role of domain-specific knowledge. *The Accounting Review* 66 (1): 100-129.

Chandler, P., and J. Sweller. 1992. The split attention effect as a factor in the design of instruction. *British Journal of Education Psychology* 62: 233-246.

Carswell, C .M. 1992. Reading graphs: Interactions of processing requirements and stimulus structure. In B. Burns (Ed.), *Percepts, Concepts and Categories* (pp.605-647). Amsterdam: North-Holland.

Chernoff, H. 1971. *The uses of faces to represent points in N-dimensional space graphical technical report No. 71*, Department of Statistics, Stanford University.

Chung, J., and G. S. Monroe. 1998. Gender differences in information processing: An empirical test of the hypothesis-confirming strategy in an audit context. *Accounting and Finance* 38: 265-279.

Choo, F., and M. Firth. 1998. The effect of time pressure on auditors' configural information processing. *International Journal of Auditing* 2: 21-33.

Cohen, J. 1992. A power primer. *Psychological Bulletin* 112 (1): 155-159.

Cook, R. D., and S. Weisberg. 1982. *Residuals and influence in regression*. New York: Chapman & Hall.

Dickson, G. W., G. DeSanctis, and D. J. McBride. 1986. Understanding the effectiveness of computer graphics for decision support: A cumulative experimental approach. *Communications of the ACM* 29 (1): 40-47.

Dull, R. B., and D. P. Tegarden. 1999. A comparison of three visual representations of complex multidimensional accounting information. *Journal of Information Systems* 13 (2): 117-131.

Faul, F., E. Erdfelder, A. G. Lang, and A. Buchner. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical science. *Behavior Research Methods* 39: 175-191.

Field, A. 2005. *Discovering Statistics Using SPSS*. London: Sage Publications.

http://www.goldensoftware.com

Haskell, I. D., and C. D. Wickens. 1993. Two and three dimensional displays for aviation: A theoretical and empirical comparison. *International Journal of Aviation Psychology* 3: 87-109.

Hygge, S., and I. Knez. 2001. Effects of noise, heat and indoor lighting on cognitive performance and self-reported affect. *Journal of Environmental Psychology* 21: 291-299.

Jarvenpaa, S. L., and G. W. Dickson. 1988. Graphics and managerial decision making: Research based guidelines. *Communications of the ACM* 31 (June): 764-774.

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

Kolata, G. 1982.Computer graphics comes to statistics. *Science*, 919-920.

Kosslyn, S. M. 1994. *Elements of Graph Design*. New York: W.H. Freeman.

Kumar N., and I. Benbasat. 2004. The effect of relationship encoding, task type, and complexity on information representation: An empirical evaluation of 2-D and 3-D line graphs. *MIS Quarterly* (June): 255-281.

Larkin, J. H., and A. H. Simon. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 11: 65-99.

Liu, Y., and C. D. Wickens. 1992. Visual scanning with or without spatial uncertainty and divided and selective attention. *Aeta Psychologica* 79: 131-153.

Lohse, J. 1991. A cognitive model for the perception and understanding of graphs. *Communications of the ACM* 383: 137-144.

Masters, M. S., and B. Sanders. 1993. Is the gender differences in mental rotation disappearing? *Behavior Genetics* 23 (4): 337-341.

Miller, G. A. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review* 63 (2): 81-97.

Moriarity. S. 1979. Communicating financial information through multi-dimensional graphics. *Journal of Accounting Research* 17 (Spring): 205-224.

Murthy, U., B. Schafer, and R. Fuller. 2007. The impact of model representation on error detection and functional use. Working paper, The University of South Florida.

Nissim, D., and S. H. Penman. 2001. Ratio analysis and equity valuation: From research to practice. *Review of Accounting Studies* 6: 109-154.

Nunnally, J. C. 1978. *Psychometric Theory*. New York: McGraw Hill.

Pinker, S. 1990. A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73-126). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Reid. G. B., and T. E. Nygren. 1988. The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 185-218) North Holland: Elsevier Science Publishing Company, Inc.

Sen, A., and W. A. Wallace. 1990. An expert systems assistance to internal audit department evaluations. *Expert Systems with Applications: An International Journal* 3 (1): 51-66.

Shepard, R., and J. Metzler. 1971. Mental rotation of three dimensional objects. *Science* 171: 701-703.

Simon, H. A. 1978. On the forms of mental representation. In Savage, C. W. (Ed.), *Minnesota Studies in the Philosophy of Science. Vol. IX: Perception and Cognition: Issues in the Foundations of Psychology* Minneapolis; University of Minnesota Press.

Sloman, S. A. 1996. The empirical case for two systems of reasoning. *Psychological Bulletin* 119 (1): 3-22.

Smith. C. P. 2005. Content analysis and narrative analysis. In T. R. Harry, and C. M. Judd (Eds.), *Handbook of Research Method in Social and Personality Psychology*, (325). New York: Cambridge University Press.

Smith, C. P., S. C. Field, and C. E. Franze. 1992. Methodological considerations: Steps in research employing content analysis systems. In C. P. Smith (Ed.). *Motivation and Personality: Handbook of Thematic Content Analysis* (pp 515-516). New York: Cambridge University Press.

Speier, C. 2006. The influence of information presentation formats on complex task decision-making performance. *International Journal of Human-Computer Studies* 64: 1115-1131.

Tan. J. K. H. and I. Benbasat. 1993. The effectiveness of graphical presentation for information extraction: A cumulative experimental approach. *Decision Science* 24 (1): 167-191.

Tegarden D. P. 1999. Business information visualization. *Communications of the Association for Information Systems* 1 (4): 1-38.

Totty, M. 2007, September 24. Making sense of it all: New software aims to put corporate data into the hands of decision makers and in a form that's easily understood. *Wall Street Journal (Eastern edition)* pp. R8.

Trafton, J. G., S. B. Trickett, and F. E. Mintz. 2005. Connecting internal and external representation: Spatial transformations of scientific visualizations. *Foundations of Science* 10: 89-106.

Tricket, S. B., and J. G. Trafton. 2006. Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. *Lecture Notes in Computer Science* 4045: 286-300.

Tufte, E.R. 1983. *The Visual Display of Quantitative Information.* Graphics Press: Cheshire, Connecticut.

Vandenberg, S.G., and A. R. Kuse. 1978. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills* 47: 599-604.

Vessey, I. 1991. Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Science* 22 (Spring): 219-241.

_____. 1994. The effect of information presentation on decision making: A cost-benefit analysis. *Information & Management* 27: 103-119.

Vessey, I., and D. Galletta. 1991. Cognitive fit: An empirical study of information acquisition. *Information Systems Research* 2 (1): 63-82.

Washburne, J. N. 1927. An experimental study of various graphic, tabular and textural methods of presenting quantitative material. *Journal of Educational Psychology* 18 (6): 361-376.

White, G. I., A. C. Sondhi, and D. Fried. 1998. The Analysis and Use of Financial Statements. (2nd ed.). Portland, Oregon: Wiley.

Wickens, C. D., and C. M. Carswell. 1995. The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors* 37 (3): 473-494.

Wickens, C. D., D. H. Merwin, and E. L. Lin. 1994. Implications of graphics enhancements for the visualization of scientific data: Dimensional integrality, stereopsis, motion, and mesh. *Human Factors* 36 (1): 44-61.

Wood, R.E.” Task Complexity: Definition of the Construct” Organizational Behavior and Human Decision Processes, 1986, Vol. 37, 60-82.

Wright, W. F. 1995. Superior audit judgment given graphical displays. *Auditing: A Journal of Practice & Theory* 14 (Fall): 144-154.

Wright, W. F., and J. J. Willingham. 1997. A computational model of loan loss judgments. *Auditing: A Journal of Practice & Theory* 16 (Spring): 99-113.

Wright, W. 1997. Business visualization applications. *IEEE Computer Graphics and Applications* 17 (4): 66-70.

Zhang P. (Ed.). 1996. Proceeding of the Second American Conference on Information Systems (AIS 1996): *The Impact of Information Visualization on Human Problem Solving Performance in a Complex Business Domain*.

About the Author

I am from Hawaii with a free spirit. Simplicity of life is what I am pursuing. I had my master's degree before I started the Ph.D. program. My research interest is inter-disciplinary studies on auditing and accounting information system. Visualization is one of my key future research areas.

I have several professional accounting qualifications. My working experience both as an external and internal auditor helps me in teaching the auditing and accounting information system classes. I am confident that I can help my students in passing the examination of the American Institute of Certified Public Accountants on their first attempt.

I enjoy surfing, as I am from Hawaii.

Aloha.