2009

Exploring efficient design approaches for display of multidimensional data to facilitate interpretation of information

Chitra S. Pathiavadi
University of South Florida

Exploring Efficient Design Approaches for Display of Multidimensional Data To Facilitate Interpretation of Information

by

Chitra S. Pathiavadi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Secondary Education
College of Education
University of South Florida

Major Professor: Ann E. Barron, Ed.D.
James A. White, Ph.D.
Jeffrey Kromrey, Ph.D.
John Ferron, Ph.D.

Date of Approval
July 10, 2009

Keywords: Multivariate Data, Data Displays, Graphical Displays, Eye Tracking, Conjunctions

Dedication

This dissertation is dedicated to my parents, Lalitha and P.Subramani. Their encouragement, support and confidence in my abilities have played a key role in motivating me to accomplish the goal of finishing my doctoral degree. Their desire to see me succeed and their constant support in looking out for possible opportunities for me to grow, develop my skills and reach my desired goals have helped me stay focused. They have worked hard to instill in me core values of commitment; perseverance and dedication in everything I do, which has helped me finally reach this point in my life and career.

Acknowledgements

I would like to thank Dr. Ann Barron my major professor for her support, guidance and encouragement through my program and research studies. I would also like to thank Dr. William Kealy, who was my major professor during his tenure here at the University of South Florida. He helped me formulate the concept for the study and translate it into research. I would also like to express my gratitude to Drs. Kromrey and Ferron for their support and insights into the research design and statistical analysis and Dr. White for his support throughout the degree program and the research study.

I would like to acknowledge three of my cohorts who have now graduated from the doctoral program. Dr. Lisa Anderson whose help and support I am very grateful for and from whom I continue to learn every day, Drs. Haunstetter and Singh, who helped me through the data collection process. Their assistance and support was invaluable.

Table of Contents

List of Tables

List of Figures

Exploring Efficient Design Approaches for Display of Multidimensional Data

To Facilitate Interpretation of Information

Chitra S. Pathiavadi

ABSTRACT

Prescriptions for effective display of quantitative information involving more than two variables are not available. To explore the effectiveness of retinal variables in facilitating the interpretation of information and decision making when used in conjunction, a study with 135 participants was conducted. The study involved the use of color shape, color value, and value shape as retinal variables in interactive displays that required participants to answer nine questions in three levels of complexity (identification of data points, analyses of local comparisons and global trends). Time-on-task scores and performance scores were measured. In addition, a View Clamp eye tracker system was used and 12 out of the 135 participants completed the task of answering questions while their eye movements were recorded. Repeated measures analysis followed by multiple comparisons of means showed that participants in the color and shape group performed significantly better and faster than color/value and shape/value groups only for questions that involved studying global trends and decision making (level 3). The shape and value group was significantly faster than color and shape group in answering level 1. Color and value as retinal variables produced results that indicated that the two variables when used in conjunction could be suitable for display of data that involved comparison. This needs to be explored further. Eye movements provided further evidence to the feature

integration theory (Treisman, 1982) and showed feature search occurred right away as participants entered the display. 78% of those who reported mental strategies indicated that they identified the features used in the display first.

Chapter One

Introduction

Prescriptions for effective display of quantitative information involving more than two variables are not available. In order to facilitate effective comprehension of the interactions and relationships in data displays involving more than two variables, and to improve the ability of people to interpret graphical displays, there is a need for design guidelines and approaches. Research into perceptual and cognitive processes have mainly focused on the ways in which people process the information presented. A study is proposed based on Jacque Bertin's image theory supported by Visual Attention and Feature Integration theories. The study attempts to explore the effectiveness of retinal variables used in conjunction to display multivariate data at different levels of graph complexities and propose design recommendations.

The use of images and visuals dates back to several thousand years. Early humans created the first information graphics, cave paintings, pteroglyphs and maps. Map-making began several millennia before writing. Graphics are pictures that resemble a person, thing or a place. Some of them are crude drawings, while others consist of photographic images. In other words, a graphic is a representation of an object on a two-dimensional surface according to mathematical rules of projection. Graphic representations have been used for several years now and constitute one of the most efficient sign systems conceived by the human mind for the purposes of understanding, communicating and storing essential information. Fang (1996) described images as serving to help (a) establish the setting, (b) define and develop characters, (c) develop a plot, (d) provide different viewpoints (e) contribute to text coherence and (f) reinforce the text. Levin et al (1987) proposed that pictures serve four conventional functions (decorational, representational, organizational, and interpretational) and one unconventional function (transformational). Representational pictures literally depict text content. Organizational pictures provide structural framework for text content. Interpretational pictures are used as clarifiers of difficult to understand materials; and transformational pictures include mnemonic illustrations of the text.

Graphs are also communicative tools, which encode data into visual images (Lohse, Walker, Biolsi & Reuter, 1991). Several formats have been used extensively in pedagogy over the years to display different types of quantitative information. Graphs and displays help identify specific data values, allow comparative judgments, and study trends by identifying relationships between variables. The graphic potrayal of quantitative information also dates back several hundred years. It arose from geometry diagrams and making of maps to aid in navigation and exploration. The most ancient graphical representations date back to the third millineum, during which time geographical maps were engraved in clay (Berg ,1983).

Over the 18th and the 19th century, numbers pertaining to people,social, medical and economic statistics began to be gathered. The use of such data and its usefulness for planning, making decisions, and soliciting responses were identified (Friendly & Denis, 2004). Graphical displays such as plots, graphs, maps and diagrams used to depict data and display information were first introduced by William Playfair. The most popular and the well-known use of graphs came in September 1954, when Dr. John Snow plotted the locations of death from cholera in central London. Deaths were shown by dots and crosses marked on a map of the area's eleven water pumps. As shown in Figure 1, the graphic display provided enough evidence for the water pumps to be removed, as the contaminated water was causing cholera.

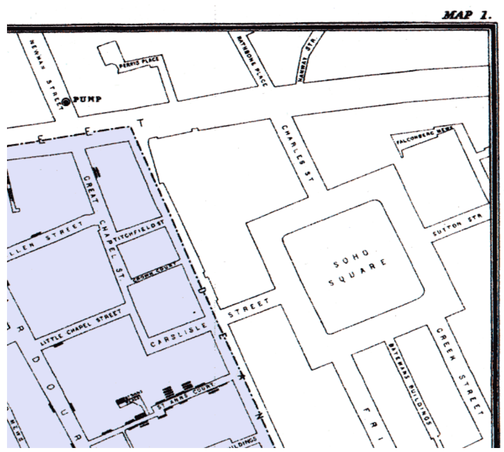


Figure 1 Snow Plot Showing Death by Cholera
Note: Obtained from
http://www7.nationalgeographic.com/ngm/data/2001/08/01/html/ft_20010801.6.html

The birth of statistical thinking was also accompanied by a rise in visual thinking; diagrams and graphs to illustrate the data. Early statisticians, (Arbuthnot, 1710; Brakenridge, 1755; Graunt, 1662) studied the London Bills of Mortality and analyzed it to inform the authorities about the possible outbreak of epidemics. Zabell (1976) used simple graphical methods to show data characteristics of clerical errors in the London bills. This also proved the immense use of graphical displays in conveying information.

Organized spatial displays such as maps and diagrams are used extensively as supporting materials in education today (Verdi, Stock, Ritschoff & Johnson, 1996). Most recently advances in statistical computation, graphical displays and human computer interactions have allowed new approaches to displaying information in a more dynamic way. Information visualization is the broadest term that covers all the advances in information displays (Friendly & Davis 2004). It includes any information presented that is organized. Tables, graphs, maps, diagrams and even text help present information in a way that explains relationships, trends and interactions. Finding effective design approaches to represent text information and data trends are thus critical to developing useful instructional materials that learners can comprehend and apply in the real world.

Context of the Research Problem

The basic characteristics of information today is multidimesionality. For example, environmental data include several variables such as location, altitude, temperature, rainfall, and pollution levels. Design and organization of such data has always been a challenge. All of the variables, their different relationships and interactions need to be displayed in a manner such that the user can identify them very easily and interpret the results. The complexities of the details and the relationships in a multidimensional data display contribute to processing limits of the working memory and cognitive load. The two-dimensional surface of a sheet of paper or a computer monitor poses a problem and representing information in commonly used visuals such as graphs and diagrams is a challenge. In this case, it is easy to perceive data values for two variables along the spatial axes, but there is difficulty when more than two variables are involved. Multidimensional data involve more than two variables at different levels, under different conditions or situations, which makes it difficult to make inferences and decisions.

In order to help improve the graphical efficacy and computational efficiency of users of graphical displays, perceptual and cognitive research studies have tried to understand the different ways in which people read graphs and interpret them.

Comparitive studies analyzing the effectiveness of the different types of formats to the information analyzed have been conducted. Researchers and practitioners have proposed several ways in which the layout of graphical displays could reduce cognitive load and help improve graphical efficacy. Factors such as preattentive processing, perceptual precedence and their role in comprehension and interpretation of displays have also been studied. Studies have demonstrated that reasoning with graphs involves a complex interaction between the perceptual and cognitive abilities, the visual properties of the graphs, and the specific task requirements (Peeble & Cheng, 2002). However, there are no guidelines or prescriptions to help improve the design that may facilitate processing of the information presented. *Semiologie Graphic*, published in 1965 by Jacque Bertin stressed the need to call upon our natural intelligence and intuition, in order to understand data and relationships. However, the problem lies in augmenting this intelligence using methods that support our perception and facilitate decision making. Bertin (1983) suggested that these methods should be able to transcribe large amounts of data, display groupings of objects and characterisitcs, which guide the interpretation.

According to Bertin (1983), there are two functionally different classes of variables, planar and retinal. X and Y axis components in a graphical display contribute the two spatial dimensions of the ―plane" and are the planar variables. Besides the planar variables, components can be represented by ―six retinal variables" such as color, size, value, shape, orientation, texture and brightness. Bertin suggests that viewers comprehend and extract information from visualizations through perceptions of relationships between the variables representing different dimensions. The relationships between the variables define trends, comparisons and aid in decision making. Thus there are eight variables, two planar and six retinal variables, available to display information (Cleveland 1984). For example height and weight components in displays can be extended to a third dimension, gender, using a value component representing males and females using different values (Green, 1998). However, no empirical data is available to support Bertin's view. The purpose of this study is to evaluate the use retinal variables in conjunction and analyze their effectiveness in facilitating comprehension of the interactions and relationships among the data.

Advanced user interactions and the real time visualization of results have been used extensively to present multidimensional data. Diverse solutions for displaying high dimensional data in low dimensional environments such as on the computer screen have been proposed (Enns, 1990, Grinstein et al., 1989; Ware & Beatty, 1988). However these research studies on multidimensional displays have mainly focused on the display formats, the cognitive processes involved in the interpretation of the data, computer models, and algorithms. There are no significant contributions made from the perspective of design of such displays.

This inquiry proposes an approach to design of multivariate data displays using the two dimensional computer screen and retinal variables in conjunction. The study will provide insight into the effectiveness of the retinal variables in design and presentation of multidimensional data. It will also look at how useful the displays are in facilitating comprehension of information presented. An eye tracker system will be used to identify the scan path and measure the eye gaze time and saccade movements. Several research studies suggest different approaches to viewing data displays- holistic, bottom up, top down, but there are no conclusions on what the learner actually does. This study will provide an insight into the problem solving approaches taken by the participants by analyzing the scan path recorded by an eye tracker, the amount of time spent looking at the different data, the saccade movements and their relationship to task performance.

Purpose of Research

The purpose of the research is to explore the effectiveness of retinal variables (color, value and shape) in facilitating interpretation of information when used in conjunction for the display of multidimensional data and their relationships. The goal of the study is to also find differences in the effectiveness of the variables used in conjunction in answering questions of different levels of complexity based on the displays as described by Wainer (1994). Effectiveness is measured as a function of performance scores and time on task. The study involves three treatment methods that include the three retinal variables, color, value and shape in conjunction with each other (color/value, color/shape and shape/value). Using an eye tracker, the researcher will also attempt to identify the participants' regions of interest in pre-defined areas in the displays as they attempt to complete the given task.

Research Questions

The overall research question is: Are retinal variables when used in conjunction, effective in facilitating interpretation of information and decision making. More specifically, the research questions in the study are

$R_{1:}$ Is there a significant difference in performance measured as number of correct responses based on the retinal variables (color, value, and shape) used in conjunction in the display?

H_1: Different retinal variables used in conjunction (color/shape, color/value and shape/value) significantly affect user performance measures as number of correct responses.

$R_{2:}$ Is there a significant difference in performance measured as number of correct responses based on the complexity level (location, local relationships and global trends) of questions answered using the different retinal variables?

H_2: Differences in the levels of complexity of questions answered using a display created with retinal variables used in conjunction (color/shape, color/value and shape/value) significantly affects user performance measures as number of correct responses.

R_3: Is there a significant difference in total time taken to answer questions based on the retinal variables (color, shape, value) used in conjunction in the displays?

H_3: Different retinal variables used in conjunction (color/shape, color/value and shape/value) significantly affect time taken to complete the given task.

R_4: Is there a significant difference in total time taken to answer questions based on the level of complexity (location, local relationships and global trends) of questions presented using the different retinal variables?

H_4: Differences in the levels of complexity of questions answered using a display created with retinal variables used in conjunction (color/shape, color/value and shape/value) significantly affects time taken to complete the given task.

The study will also analyze the following related correlations.

Correlations

Q_1. What is the direction and strength of the relationship between performance scores and time on task scores based on the retinal variables used in conjunction in the displays?

Q_2. What is the direction and strength of the relationship between usefulness rating and the performance scores measured as the number of correct responses, based on the different retinal variables (color, shape, value) used in conjunction in the displays?

Q_3.What is the direction and strength of the relationship between usefulness rating and the time taken to answer questions at different levels of complexity (location, local relationships and global trends) of questions presented using the different retinal variables?

In addition to exploring statistically significant differences and correlations, the study also explores the sequence of the eye movements and the fixations scores that relate to each of the regions of interest defined by the researcher in each of the displays. Thus the study also investigates the following questions

Q_4. What is the direction and strength of the relationship between eye gaze time, saccade scores and total time taken to complete the task based on the different retinal variables (color, shape, value) used in conjunction in the displays?

Q_5. Is there a difference in the sequence in which participants look at the regions of interests defined in the displays in the three treatment groups as they perform their tasks?

Definition of Variables

Identification of appropriate variables both independent and dependent, are critical for the success of the study. The study includes the following variables

Independent Variables

Three treatment methods (color/value, color/shape, shape/value) that include different combinations of the retinal variables in conjunction with one another to display multivariate data. The three different treatment groups are described in detail.

1. Color/Value - variables represented by cyan and magenta in a graphical display. Cyan and magenta are the colors of choice since they are located opposite to one another on the color wheel. A darker shade (value) of the color represents higher data values and a lighter shade (value) represents lower data values.

2. Shape/Value - circles and squares displayed in different degrees of combinations of lightness and shadow(white and black) as seen in the gray scale. The lowest level of the scale is white and the highest level is black with different shades of gray inbetween representing the different data values. The lighter shades represent lower data values and the darker shades represent higher values.

3. Color/Shape - circles and squares displayed in colors of magenta and cyan represent the different data values. Both circles and shapes have the same color value.

The study also includes the difficulty level of questions used a within subjects repeated measures. Questions of different levels of complexity based on the display will be answered by the participants.

The difficulty levels of questions are:

1. Level 1 includes questions that require identification of specific locations of data points.

2. Level 2 includes questions that require comparison of data values and identification of relationships between them.

3. Level 3 includes questions that require the identification of global trends and their comparisons. Level 3 requires decision making when answering the questions.

Dependant Variables

The effectiveness of the retinal variables used in conjunction is measured as a function of the performance scores and the time taken to answer the questions. The following are the dependant variables in the study.

1. Performance scores are measured as the accuracy of responses to the questions at different levels of complexity

2. The total time taken to answer questions in each of three levels of complexity

3. Usefulness rating of the graphical displays provided to the participants based on a likert scale.

4. Fixation and saccade cores measured as time in seconds obtained from the eye tracker.

Limitations/Delimitations

Retinal variables used in the study are limited to shape, color and value. These variables were chosen since they are more common than the others. They are also effective in displaying changes over time. The participants in the study are from a population of pre-service teachers. Hence, the results cannot be generalized across all disciplines and professions. All participants are assumed to have basic computer skills and experience using the computer. Participants include only those without any prior knowledge or experience in engineering, computer science, compiling and analyzing multivariate data. The design includes physical control of these participants, which in turn will allow control of the extraneous variable prior knowledge.

Implications to Education and Learning Environments

Graphical displays and semiotics (use of graphics as a sign system) have been used extensively in pedagogy. Studies in numerous elementary schools (Gimeno, 1980) show that “graphics can introduce into all disciplines the basis of logic and the essential processes of analysis and decision making” (p192). Graphics can stimulate exceptional motivation, foster better questions, aid in constructing text, and reveal the intelligence of the so called “poor students.” Graphs and graphical displays have been extensively used in text books and educational software to help students interpret data (Shah & Hoeffner, 2002).

Graphical displays have been used to supplement text and visually present data interactions and relationships for a number of years now. Organized spatial displays such as maps and diagrams are used as supporting materials in schools today. Several studies have shown that effective visualization can aid in learning. Instructional design theories and models that support constructivism (Mayer, 1992) suggest that graphical representation of information, their organization and structure, are the key factors that facilitate computer based learning. However, in most cases, instructional and/or learning materials involve discussion of several variables and their interactions together at one time in order to study trends, make decisions, and solve problems. This presents a challenge to the learner because many people do not have the needed computational expertise (Wainer, 1980). Thus, exploring design guidelines for display of multidimensional data to help the user interpret the interactions and relationships is of immense importance in the field of visualization, statistics, scientific research and education.

Definitions of Terms

In order to understand the concepts and research studies this investigation is based on, it is important to have knowledge of the terms involved in the study. This section includes a list of terms and their definitions in alphabetical order.

*Configurations and Discriminations***.** William Winn in his attempt to derive a theoretical framework for learning from maps and diagrams proposed that manipulation of symbol systems to change how objects and concepts are shown (discriminations), and how they are placed relative to each other causes (configurations) cause psychological processes to act differently. These processes determine which objects appear to form clusters, and the sequences in which objects are processed (Winn, 1991). Conjunction (a discrimination) is one that is made up of two or more features only one of which is contained in each of the distracters (all non-target objects are considered distracters)

Data Visualization. Data Visualization is a joint function of perception and graphics. It is the transformation of data into useful information. It involves two stages. The first is the map from numbers (the data and the processes) to images, and the next is the translation of the images into insight or information by perception.(Green, 1998).

Eye Tracking. The collection and analysis of a participant's eye gaze data in a specific learning environment. These data are indicative of specific gaze patterns and physical eye movements.

Feature Integration Theory. The feature integration theory assumes that features come first in perception and in parallel across the visual field, while the objects are identified later. This theory forms the basis of studies by Treisman (1977). According to the theory, the type of features or discrimination factors used to represent the data and their relationships influence interpretation of the displays and inferences that are made from it.

Gaze Fixations Score. Fixations scores are numeric data that are collected from eye tracking systems when the eye fixes attention to a specific object. This is recorded through an eye tracking system software, ViewPoint.

Glance Frequency. Fixation scores measured as time in seconds as the eye fixates on a region of interest consecutively through its movements.

Graphics. Based on the rationale of imagery, graphics differ from both figurative representation and mathematics. In graphics, a diagram or a map, each element is defined beforehand.

Graphical displays. Quantitative studies and phenomenon have been displayed in a variety of ways, the most effective of which have been the graphical methods proposed by William Playfair (Weiner, 1991). Tabular representations, graphs and maps, are the most common ways in which data are displayed. Tilling(1975) in his study of experimental graphs, found that graphs and maps are not interchangeable. In order to plot a good graph, we need to first analyze and understand the information that needs to be displayed. Figures 2, 3, and 4 provide a few examples of the different types of representations (Table, Graph and Map) of data:

Trans- Portation- Year	Parks & wildlife areas	National areas	Defense areas	Urban areas	Farmstead areas
1945	22.6	27.6	24.8	15.0	15.1
1949	22.9	27.6	21.5	18.3	15.1
1954	24.5	27.5	27.4	18.6	12.2
1959	25.2	46.9	31.1	27.2	11.4
1964	26.0	75.5	31.9	29.3	10.5
1969	26.0	81.3	25.6	31.0	10.3
1974	26.3	87.5	25.0	34.8	8.1
1978	26.6	98.0	24.9	44.6	8.4
1982	26.7	211.0	24.0	50.2	8.0

Figure 2 Example of a Tabular Representation

Source: U.S. Department of Agriculture, Economic Research Service. Major Uses of Land in the United States: 1987, Appendix Table 4, p. 247, AER 643 (Washington, DC:USDA/ERS, (1991)

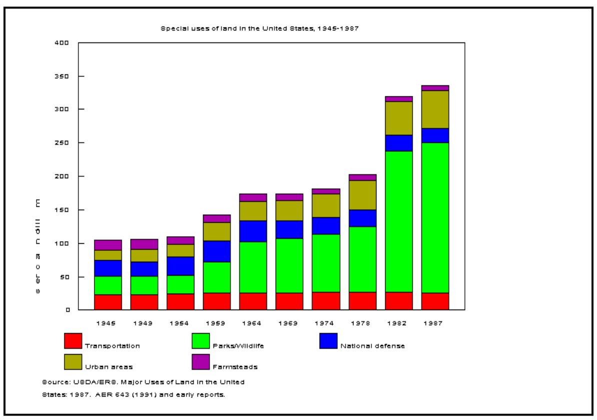


Figure 3 Example of Graphical Representations using a Graph

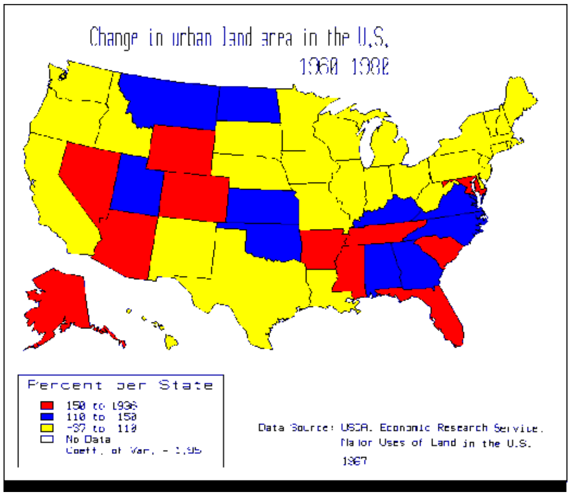


Figure 4 Example of Graphical Representation using a Map

The basic characteristics of information today is multidimesionality. Data involve more than two variables. Hence it is a challenge to represent information in commonly used visuals such as graphs and diagrams. According to Gorban and Zinovyev (2005), it is easy to perceive data values for two variables along the spatial axes, but we have difficulty at higher dimensions.

Perception and Imagery. Images are constructed (Chandler, 2001). Perception consists of decoding what we see. Perception is the registration of physically present stimuli; whereas, imagery is seeing patterns that arise from memory in the absence of the appropriate sensory inputs (Kosslyn, 1995). Imagery requires the brain to reconstruct, present from recognition and put together information from past first hand experiences (Bertin, 1983).

Preattentive Processing: Healy, Booth and Enns (1996), proposed a new method for performing rapid and accurate numerical estimations from preattentive processes. Preattentive processing refers to an initial organization of the visual field based on cognitive operations believed to be rapid, automatic and spatially parallel. Examples include high-speed target detection, boundary identification and region detection using preattentive features such as hue and orientations

Regions of Interests. The View Point software used along with the eye tracker in this research study allows for the source of the stimulus (the three displays used in the study), to be divided up into regions of interest (ROI). These are very useful to know whether or not a participant's gaze was in a certain area. It is possible to specify up to 100 ROI boxes. In this study, there are 12 ROI areas defined by the researcher.

Retinal Variables. A visible mark on a sheet of paper, representable or printable, can vary in position. Fixed at a given point in the plane, a mark, provided it has certain dimensions, can be drawn in different modes. It can vary in color, size, texture, value orientation and shape. A mark then can also express a correspondence between its planar position and its position in the series constituting each variable. A designer thus has eight variables to work with to express data in the non-spatial areas besides the two axes. They have different properties and different capacities of portraying information. These variables are called the retinal or the visual variables (Bertin, 1965). One of the most common methods of representing multidimensional data is where combinations of the retinal variables are used. Figure 4 represents the retinal variables as defined by Bertin.

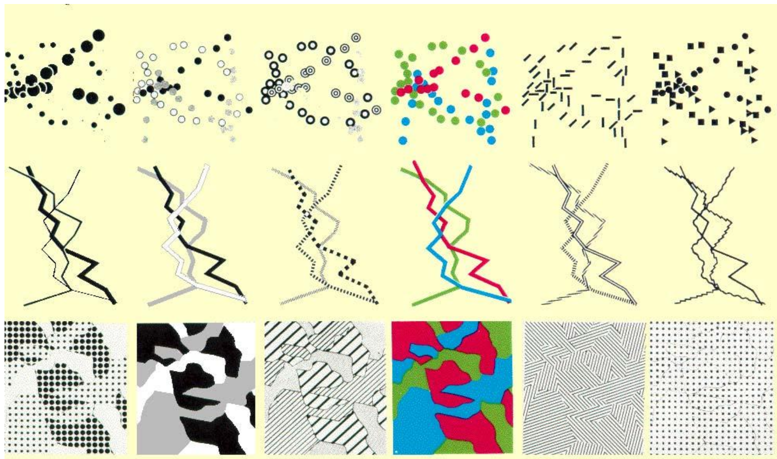


Figure 5 Retinal Variables as Defined by Jaques Bertin
Size, value, texture, color, direction and shape from left to right
Note: Obtained from Semiology of Graphics: Diagrams, Networks, Maps. Bertin, J 1983, University of Wisconsin, Madison, WI

Value . Shades and tones of a color, also known as Hue. An example is the monochromatic scale or the gray scale, which ranges from white through different values of very light, light, gray, and dark gray, finally ending in black.

Shapes Different shapes on the displays used to represent points or marks and values of the variables.

Color. Any of the 256 system colors that are used to indicate the values of the variables such as red, green, blue. These are distinguishable different colors and not hues or values of the same color, Examples red, blue, green, yellow

Saccade Scores. Numeric data that are collected by the eye tracking system when the eye scans and takes in information prior to fixating upon a specific area. It happens 3-4 times every second.

Visualization. Most vision theories suggest that visualization occurs in three stages, which include formation of the retinal image, the decomposition of the retinal image into specialized representations, and the reassembly of the information into object perception.

According to Green (1991), an image or a visual is multi-dimensional in that it contains a large number of properties such as color, shape and motion and true visualization involves a formal theory of computer graphics and a theory of human perception. Non textual symbols such as graphics and images that are used to convey information are called visuals and the presenting and transmitting of such visuals, is what we call visual communications (Moonen, 1999).

Summary

This chapter provided an introduction to the field of graphics and data visualization. The various theories and the implications of the science on presentation of data and their interpretation provide a good foundation and rationale for the research study. Graphical efficacy continues to be a major challenge for all audiences and users of data and this study is an attempt to address the issue.

Chapter Two

Literature Review

The power and the large-scale use of computers have increased our abilities to store, and manipulate numbers (Green, 1999). However, much of the improvements are not implemented since humans are not very proficient in gaining insight from data presented to them. As a result, research in data visualization takes on an important role of transforming the large amounts of data into information that humans can understand, explore and analyze. As a science, visualization involves theories of computer graphics and human perception (Greenberg, 1988). The proposed research study involves a theoretical framework that includes the different areas of information processing, perception, data visualization, information visualization and the theories underlying the display of multidimensional data. In order to further understand the rationale for the study and the research design, a discussion of the different theoretical frameworks is provided in this chapter.

Information Processing

Information processing has developed prodigiously. We now believe ―understanding" to be breaking down and simplification of vast amount of data to the smaller building blocks of information that we are capable of dealing with in resolving a particular problem (Berg, 1983). Our forerunners, who had no access to computers, still attempted the simplification process to understand the data displayed to make decisions.

However, the lack of resources and the time taken to complete these simplifications limited the scope and scale of the research. Various areas of cognitive psychology, vision, visualization and technological research have over the years tried to identify the best level and method of the simplification process. Early research studies in visualization such as those of Hanning (1962) show that humans are poor at gaining insight from data presented in numerical form. Bertin(1965) suggested that an artificial memory or method should be able to transcribe large amounts of data and display groupings of objects and characteristics that guide interpretation.

Perception

Human perception plays an important role in visualization. There has been a steady increase in awareness of the importance of perception as a factor in the design of good visualizations (Lekowitz & Herrman, 1992; Merwin & Wickens, 1993). An understanding of perception can help improve the quality and the quantity of information that is being displayed (Ware, 2000). According to Healy and Enns (2002), visual perception is more a dynamic and ongoing construction process, where the products built are very short lived models of the external world that are relevant and important to the user's tasks on hand. Perception is also the registration of physically present stimuli. The visual perception is not given, it is constructed (Kosslyn, 1995). Humans, as a species, desire to make meaningful connections of everything they observe. This is the underlying assumption of the process of visual perception (Chandler, 2002). However, some images are more open to interpretation than others as seen in Figure 5. Whatever the case may be, once we realize what we are looking for we see meaningful patterns in them.

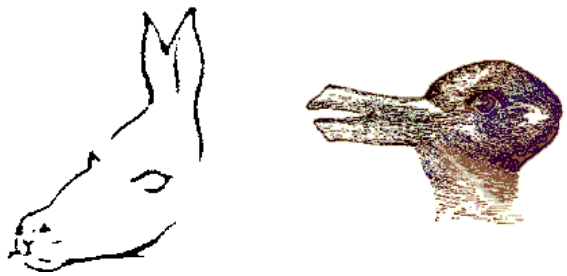


Figure 6 Example used by David Chandler to Explain Perception
http://www.aber.ac.uk/media/Modules/TF12710/visper01.html

The meanings and patterns we recognise from an image or display, however, depends on our perceptual set, which are our own preferred interpretations of the image. In the context of the media theory (McLuhan, 1980), the openness of most images to interpretation suggests that meaning is not a universal fixed content, but it is constructed within certain constraints. Perception and meaning are influenced by culture, the exposure and knowledge one has, and their personality as well as gender. Thus perception and recall of what might seem to be the same image or event involves a very active construction of differing realities (Hastorf & Cantril, 1954). Besides these factors, various kinds of context also shape our interpretation of what we see. Progress and changes over the years (historical context), cultural differences (cultural context), our life experiences (situational context) and structural context, contribute to our meaning and interpretation. Besides our life experiences, culture and context, a schema, a kind of mental framework is also used by humans to make sense of what they see. Based on the context, circumstances activate schemata, which then set up various standard expectations (Bruner, 1986). Furthermore, they also allow us to make inferences based on our experiences of things that are not seen. The application of schemata and the expectation represents a top down process in perception, while the activation of schemata by sensory data is a bottom up process.

Perception is selective. There are physical as well as cognitive limits. We cannot see everything at once. We also tend to ignore the details that are irrelevant to our purpose. Thus, selectivity involves omission. Furthermore, using minimal data we are able to fill in the gaps and recognize what could fill them up. According to Newcomb (1952), selectivity also involves organization. We rearrange objects and events based on our frames of reference and these influence how perception is structured. Selective perception is based on what seems to stand out. Much of the standing out however is related to our purpose, interest, expectations, past experience, and current demands of the situation (Chandler, 1997).

William Winn (1994) describes the cognitive and perceptual processes that are involved in the organization, interpretation and comprehension of graphics. According to him, the notational symbol system of graphics provides a one to one relationship between each symbol as well as clear indications among them. Perceptual processes acting on the symbols and their relationships allows the viewer to detect, discriminate and configure graphic symbols into patterns. Cognitive processes influenced by perception leads to identification, interpretation and comprehensions of what is seen. Salomon (1979) proposed that when a symbol system used to convey a message interacts with psychological processes active in the person who receives it, comprehension occurs. In the graphics system, symbols are highly notational. This was proposed by Goodman (1968), who identified that each of the discrete symbols, bears a unique and unambiguous relationship to an object in the domain the graphic describes and has a unique set of spatial relationships with other symbols.

Thus, harnessing the theories of human visual perception for data visualization requires that we construct displays that draw attention to the most important parts. We must effectively map the data, its features and the interactions. The first step is to help the viewer make a mental representation of the display or the scene that outlasts the next glance that comes into view (Healy, 2002). With the use of the eye tracker, the researcher hopes to identify the path the participants take as they try to answer the questions, thereby suggesting a top down or a bottom up approach to problem solving. Most research studies in the area of data vizualisation have used mathematical algorithms and computer models to display multidimensional data, but eye tracker capabilities have not been used.

Cognitive Architecture

In 1994, Wainer noted that the graphical interpretation abilities of humans is very low. One of the reasons for such inefficiencies is the fact that data displays in many instances involve higher order thinking skills and decision making where many factors need to be looked at together. This can be further explained by taking a look at the cognitive architecture of humans. Visual cognition is the central point of human cognition (Sweller, 1996). Understanding the cognitive architecture of humans is critically important in data vizualisation and greatly affects the approaches one may take for the design of displays. The cognitive architectures include a short term memory of limited capacity and duration, visual and auditory channels, and an infinite long term memory that holds several schema or mental frameworks. Baddeley and Hitch (1974) define the short term memory as the working memory. Working memory has a limited capacity (Miller, 1956) and an extremely limited duration (Peterson & Peterson, 1959). These major constraints of working memory should be taken into consideration in the design of multivariate data displays due to the fact that when faced with new high element interactive materials, one cannot process it effectively (Sweller, 1996). In other words, we fail to understand the material if it is too complex. Comprehension of such materials requires other mechanisms and learning strategies. The work of DeGroot (1965) on chess players demonstrated the critical importance of long term memory in higher order thinking functions. The performance of the chess masters was based on their ability to quickly identify patterns vizualising enormous numbers of boards held as schemas in their long term memory. Long term memory is not only important for recall but also for the processing of high level tasks. Studies on schema formations and problem solving by Larkin, Simon and Simon (1980) and Chi, Glaser and Rees (1982) demonstrated that problem solving expertise and higher order thinking demands the acquisition of several domain specific schemas. These schemas allow experts to recognize patterns and problem states and their appropriate resolutions.

Thus interpretation of multidimensional data displays that involve higher order thinking skills and problem solving should present information in a way that facilitates a deep understanding of the relationships and interactions among the data, so as to be effectively stored as knowledge in schemas, which then would enable the users to recognize patterns efficiently. This study uses retinal variables in conjunction and analyze their effectiveness in the design of the display to promote such comprehension.

Image Theory

Bertin's (1973) image theory demonstrates the close connection between visualization and theories in cognitive psychology, perception and physiology. His theory analyzes vizualization into sets of primitive components and procedures for combining the components and for creating good visuals. In addition, he defines image as a fundamental perceptual unit and states that an ideal visualization will contain only one single image in order to enhance speed and efficiency. In the real world, there is very rarely a visualization that has a single image. All of the data that is presented is usually multidimensional. If the number of images is a factor that effects speed and efficiency, it is critical to find design approaches that work to help enhance comprehension and aid interpretation of the data interactions and relationship. Bertin proposed two classes of variables- planar (variables in the x,y plane) and retinal (shape, color, value, hue, orientation, size, texture). He used the term *retinal* based on the assumption that humans have automatic, preconceived reactions to these variables at the retinal processing level (Fillipakopoulou et al.,2001). The image theory (Bertin, 1973) suggests that good visuals use variables with proper length and level of organization. Length refers to the number of categories or steps. A visual or a retinal variable must have a length (distinguishably different colors, brightness, etc.) that is equal to or greater than the components it represents. The major challenge is to match the variable to the appropriate length. The variables may be represented in different levels of data scale, nominal, ordered, or ratio. The nominal level data can occur at two levels of organization, associative and selective. The associative level is the lowest organization level, allowing grouping of all elements of a variable in spite of different values. Texture, color, orientation and shape are associative. The next higher level above the associative is selective, where the viewer can select one category or component, observe all the locations of the objects related to that category and ignore the rest. The highest level of organization is quantitative, which permits the direct extraction of ratios, without needing to consult a legend. The ratio of the variables directly maps to the ratio of data values. Only size and planar dimensions are quantitative (Green, 1991)

Bertin proposed the seven visual variables and associated syntatic rules to match each of the variables to characteristics of data represented. However, Bertin proposed his theory solely based on introspection and did not make any attempts to prove them through empirical research. He based his conclusions on the image itself and did not consider perception and psychology of the viewer. Perhaps the only study that has evaluated directly the selectivity and effectiveness of the retinal variables is the study by Fillipakopoulou et.al (2001). They conducted a series of tests in an attempt to examine the level of perceptual organization supported by the four basic retinal variables hue, value, shape and size. They specifically examined the following

1. Can shape be selective and symbols of the same complexity pop out from, or be grouped?

2. Does complexity affect selectivity? Pictures with six abstract, six composite and six pictorial symbols were used. Each picture was displayed for a short period of time (7 sec), and then, the participants of the test were asked to recall the shape they saw. Circles and triangles were the most recalled, and hexagons and rectangles were the least recalled.

3. Does hue affect selectivity? They studied the relative selectivity of specific hues such as red, yellow, green, blue, magenta and orange, and whether the hue of the background such as ochre, green and gray affects the selectivity. Recall after the picture was displayed for 7 seconds was also evaluated. Red, yellow, blue and green were recalled significantly more times than magenta and orange.

4. Do hue and complexity affect map perception? The participants in the study were asked to answer a questionnaire based on a chloropeth map. The researchers found that gradual variations in the value and saturation indicated differences and increases exponentially rather than in a linear fashion.

5. Whether the visual variables of hue and shape can activate schemata classification of data to map users with no education on map syntactics.

6. Whether the visual variables of size and value can activate the schemata of ordering of data. For the first part of the test, students aged six to nine, and for the second test, students aged 14-15 and 17-18 were asked to participate. Participants had to portray themes with data of the nominal and ordinal scale of measurement using point symbols; whereas, at the second test, data of nominal, ordinal and ratio scale were used. The first test used a map on a display and the second a paper map. Most pupils applied size and hue for ordinal data and high schoolers could portray ordinal and ratio data. Size was applied more for quantity than value.

Fillipakopoulou et.al (2001) made the following conclusions from the study. The dichotomy of the visual response to the variable (whether selective or not selective, ordered or not ordered) is strict. Shape can enhance the selectivity of point symbols in a map. However selectivity also depends on the number of vertices, the relative dimensions, the surface and its area, and the spatial frequency. Recall of pictorial and abstract symbols is easier than composite symbols. Red, green, yellow and blue were more selective than all other colors. Value-saturation affects the retrieval of information. Perception of differences based on value are more effective with less number of data classes. Hue and shape can affect the schemata of classification and size can affect the schemata of ordering. Even though the study was a direct attempt to evaluate the variables, the results were based only on recall and identification of features and did not involve information processing. A simple chloropeth map was used in the study.

This dissertation study focuses on studying the effectiveness of the variables in identification of the interactions and the trends among the data. The study involves not just the identification of data locations, but also information processing at more complex levels. It also uses an interactive data display on screen and the basic variables such as color, value and shape, and does not involve recall. The study uses a data display with shape, value and color as retinal variables and not a map.

In any graphical display, the identification of the number of components is the first step in the analysis of data. The different identifiable parts of the components are the categories or elements. The complexity of the display is linked to the number of categories and their relationships. According to Bertin (1965), graphics and data analysis work together. At the stage of interpretation, the two languages complement each other. Quantitative data that is to be used to acquire information can be thus depicted through graphics for easy comprehension. In his book *Semiologie Graphique,* Bertin (1983, p11) suggests that ―one does not see a map or a graphic, but one asks questions about it". What are the x and y components of the data? What are the groups in x and y around which the data is built, and what are the exceptions in these groups that we perceive? Graphical images are visuals that depict the real world. According to Bertin (1973), graphical displays answer the following three levels of questions

1. Elementary levels that involve data extraction. For example, "On a given date, what was the price of stock x?" (Bertin1973, 10)

2. Intermediate levels that involve trends seen in the data. For example, "In the first three years of sales, what were the highest sales recorded?" (Bertin1973, 10)

3. Overall levels that involve the understanding of the deep structure of the data using groupings and comparisons. For example "During the entire period of ten years the company was in existence, what were the trends in sales prices?" (Bertin1973, 10)

Visual Attention Theory

Visual attention has been studied for many years. James (1981) proposed a definition for the term *attention*. According to him, attention is taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneous possible objects or trains of thought. Focalization and concentration of consciousness is of its essence. Attention implies withdrawal from some things in order to deal with others effectively. Many objects that are seen are filtered into a few for perception. The faculty implied as the filter is that of attention. This implies that, as humans, we cannot pay attention to many things at once. Our capacity for information processing is limited. The brain processes sensory inputs by concentrating on specific components of the entire sensory realm so that interesting sights, sounds and smells are examined in greater detail than peripheral stimuli. Thus, human vision is a piecemeal process relying on the perceptual integration of small regions to construct a coherent representation of the whole (Chun & Wolfe, 2000).

The Components of Visual Attention

Heijden (1992) proposed that attention consists of two main components, the ―what” and the ―where”. The *where* viewpoint of visual attention emphasizes the location that subjects are orienting their visual focus. The *what* viewpoint of visual attention specifically focuses on the item or item group that is being attending to. Von Helmhotltz (1925), in his *Treatise on Physiological Optics*, notes that the only way we can see all the individual parts of a field as distinctly as possible is by letting our eyes roam continually over the visual field. He also proposed that attention could be controlled by a conscious and voluntary effort allowing attention to the peripheral objects without making eye movements to spatial location. In sharp contrast, James (1981) believes attention to be more related to imagination. He defined attention in terms of the "what", the identity, the meaning and the expectations associated with the item or the item group that is being attended to. James favored the active and voluntary aspects of visual attention.

What and where theories of visual attention are significant. The what and the where of attention correspond to the foveal (James, 1981) and the parafoveal (Helmholtz, 1925) aspects of visual attention. This explains the bottom up, feature driven approach of humans where our attention is focused on certain parts of the displays or images. These are the regions that are perceived parafoveally, and drive the attention in terms of where to look next and identify what for a detailed explanation. Many computational algorithms and models for display of multidimensional data are based on the what and where model. However, higher level problem solving and decision-making tasks involved more than just features. Some of them include intentional factors related to the voluntary preconceived cognitive factors that drive attention.

Gibson (1941) proposed the third component in visual attention, the how factor, which is based on intention. The *how* viewpoint focuses on specifics of how subjects are able to attend to information in the visual plane. During the design of experiments, it is important to consider the perceptual expectations of the stimulus, which are influenced by the researcher's instructions. It involves the viewer's advance preparation as to whether to react, and if so, how. In order to prepare the participant for this research study, the researcher will explain the purpose of the study and the task before the participant looks at the display. A practice display will also be provided, and the participant will be allowed to look at it as long as they want to familiarize themselves both with the display and the interactivity involved. The area of focus and the scan path of the eye will be further recorded by the eye tracker, which will provide an in depth analysis of where the participant was looking and for how long.

Preattentive Theory

Landman (2004) describes visual attention as being commonly divided into two successive stages--preattentive and attentive processing. Preattentive processing is a quick and basic feature analysis of the whole visual field, on which attention can subsequently operate. According to the theory, attention can directly select objects, and the number of objects selected is limited by the capacity of information processing (Neiser, 1967). It includes a set of visual properties that are detected very rapidly and accurately by the low level visual system. Tasks that can be performed on a large multi-element display in less than 200-250 milliseconds are considered preattentive. Eye movements take about 200 milliseconds to initiate, and random locations of elements in the display ensure that attention cannot be prefocused on any one particular location. However, viewers seem to complete the task with little effort (Healey, 2006).

An example of a preattentive task is the detection of a red circle in a group of blue circles. The target object has the visual property red, which the distracters do not. At a glance one can tell if the target object is present or not. Here the identification of the target is made through the difference in hue. Another example would be the identification of a red circle among red squares, which are distracters. Here the identification is made through the differences in form. Unique visual properties such as form and color help the target object pop out of a display.

Another example is the use of conjunction features, such as those used in the proposed study. In a conjunction search as shown in Figure 6, the red circle target is made up of two features: red and circular. One of the features is also present in the distracter objects: red squares and blue circles. There is no unique property here that helps the target object pop out of the display. If a viewer searches for the red item, the visual system always returns true because of the red squares in the display. Similarly a search for the circular items sees blue circles. Such items cannot be processed preattentively. Viewers must perform serial searches through the displays to identify the targets. Preattentive search is easily achieved in a display that represents a single feature. In situations where serial conjunction searches that involve more than one feature, (Sagi & Julesz, 1985; Green 1991) require the integration of the object with the spatial coordinates. Green(1992), suggested that if the user can locate a target effortlessly, then it is easy to spot the location of the target on the x and y axes and study the relationship of the objects with others.

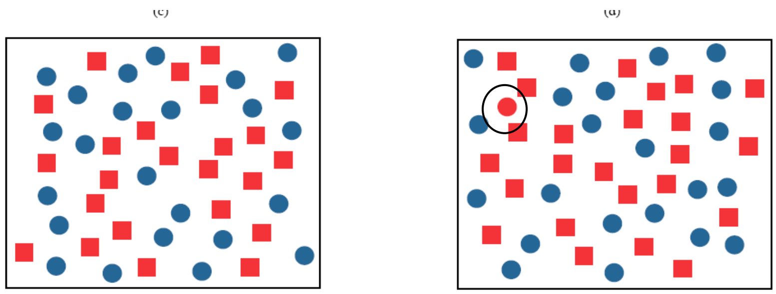


Figure 7 An Example of Conjunction Use in the Search for a Target Red Circle. Target Absent in the First Image and Present in the Second
Note: Obtained from http://www.csc.ncsu.edu/faculty/healy/PP/index.html

Research literature reveals that only for a small number of attributes is there a consensus that they are preattentive features. Color, size and orientation are without a doubt preattentive features. Shape, line termination, closure, curvature and topological status can also be considered preattentive features.

One of the most important and significant theories about the relationship between attention and vision is the Feature Integration Theory proposed by Treisman and Gelade (1980). The theory states that certain basic features are represented as individual retinotopic maps for color, orientation and motion. Attention provides the glue that binds and integrates the separated features in a particular location into a master map so that the object can be perceived as a unified whole. The master map specifies where things are but not what they are. Treisman (1991), Gormican (1988) and Souther (1986) ran experiments using target and boundary detection methods to classify preattentive features. For target detection, participants had to detect a target object in a background of distracter elements. As illustrated in Figure 7, boundary detection involved placing groups of target elements with a unique visual feature within a set of distracters to see if the boundary can be preattentively detected. Task performance was measured by accuracy and speed. The display was shown for a fixed duration and removed from the screen. The numbers of distracters were constantly increased. If the participants were able to perform the task accurately regardless of the number of distracters, and were consistent, the feature used to define the target was considered preattentive. Treisman found that feature identification is instantaneous. Features used in conjunction cannot be detected by accessing an individual feature map. Here in order to locate the targets, one must search serially through a master map of locations, and then use focused attention to look for an object and its associated combination of features. The amount of differentiation between the target and the distracters for a given feature increases the search time.

Figure 8 Form and Hue Segregation. Hue Boundary is Preattentively Detected even though Forms Varies, but Hue Interferes with the Detection of Form Boundary
Note: Obtained from http://www.csc.ncsu.edu/faculty/healey/PP/index.html

Predictions were tested on a number of paradigms including visual search, texture segregation, identification and localization and using separable dimensions such as color and shape and local elements or parts of Figures such as curves and lines as features to be integrated into complex wholes. Treisman et al (1977) compared search for targets specified by a single feature such as pink in brown and purple distraction in one condition, —O in —N and —T” distracters in another. The stimulus displays were made using letter stencils and colored inks on white cards. The distracters were scattered and placed in random order. Participants were asked to search concurrently for two targets; each defined by a different single feature; a color and a shape. Thus they were forced to attend to both the dimensions as a feature and as a conjunction. From these studies, Treisman, Gelade and Sykes (1971) proposed that attention that addresses the features come first in perception. Features are registered early, automatically and in parallel across the visual field, while objects are identified separately and only at a later stage, which requires focused attention. The visual scene is initially coded along a number of separate dimensions such as color, orientation, spatial frequency, brightness and direction of movement. Focused attention is then required to recombine separate features and ensure the correct synthesis of features for each object in a complex display. Any feature that is present within the same central fixation of attention is combined to form a single object. This study also proposed that without focused attention, features cannot be related to one another. We cannot consciously perceive an unattached shape without giving it color, texture, size and location. Separate items are detected by parallel search. Conjunctions on the other hand, need serial processing and seem to cause no effects on performance. If two physical properties are integral, then they function as a single object allowing parallel search, texture aggregation and detection. The study also helped confirm that more distinctive colors and shapes allow search to proceed nearly three times as fast as the less distinctive.

Conjunction searches involving search tasks such as motion, depth, color and orientation have also been shown to be preattentive (Nakayama & Silverman 1986; Driver et al, 1992; Wolfe et al, 1989). Focal attention, and scanning of successive locations serially, is the means by which the correct integration of features into multidimensional percepts is ensured. The proposed research will use retinal variables such as value, color and shape to display the interactions among variables and use the eye tracker to examine the processing strategies of the participants. The colors used for the study will include cyan and magenta, since they are on the opposite sides of the color wheel and complement each other. Colors such as red and green stand out more significantly than others and are easy to identify.

Similarity Theory

Humphreys and Quinlan (1987) investigated conjunction searchers by focusing on two factors. First, search time may depend on the number of items of information required to identify the target. Second, search time also depends on how easily the target can be detected from its distracters irrespective of unique preattentive features. Later, Duncan, Humphreys and Muller (1990) proposed that the search ability varies depending on both the type of task and the display conditions. Search time is based on two criteria: T-N similarity and N-N similarity. T-N similarity is the amount of similarity between the targets and non-targets. N-N similarity is the amount of similarity within non-targets themselves. As T-N similarity increases, search efficiency decreases and the search time increases. As N-N similarity decreases, search efficiency decreases and search time increases. Duncan and Humphreys (1990) proposed a three-step theory of visual selection:

1. The visual field is segmented into structural units, which share common properties (spatial proximity, hue, shape and motion). Each unit may again be segmented into smaller units. This produces a hierarchical representation of the visual field. Within

the hierarchy, each structural unit is described by a set of properties (location, hue, texture and size).

2. Very little memory resource is allocated among structural units. A template of the information being sought is available. Each structural unit is compared to this template. The better the match, the more the resources allocated to the given structural unit.

3. A poor match also allows for efficient rejection of other units that are strongly grouped to the rejected unit. Structural units with a relatively large number of resources have the highest probability of access to the visual short-term memory. Thus structural units that are closely matched to the template of information that is being sought are presented to the short-term memory. Search speed is a speed of resource allocation and the amount of competition for access to the visual short-term memory.

Graphical Displays and Data Visualization

Finding different ways to improve representation of information has been an area of interest for many cognitive psychologist and researchers. Some of the areas that relate to the proposed dissertation are described.

Semiotics The rapid increase in research and methods to transition information from simple representations to a system that is independent and complete has given rise to semiology, the science that studies the use of graphics as a sign system. Thoughts and ideas have for long been expressed through signs, which have been a means of communication. Barthes (1974) described semiology as a system of signs that include their substance and limits; images, gestures, musical sounds, objects and the complex associations of all of these. According to Morris (1938), semiotics embraces semantics, the relationship between signs and what they stand for. Signs are used in different media, which include different categories as speech, writings, print, and mass media, such as television, newspapers and magazines. Noth (1995) suggested that humans are multi-sensory and interactions between the users and the medium are an important indicator of the experience. Different media provide different frameworks for experiences, and this factor has been put into use in the field of education to address the issue of different learning styles.

Lauretis (1970) suggests that there has been a shift in the focus of semiotics in the past decade. The focus has shifted from classification of sign systems, their levels of structural organizations, towards modes of production of meanings from signs, coding and encoding and effective translation of the meaning. Translation of data, their interactions through meaningful displays to convey trends and information gained importance.

Graphical Displays The early 1900's saw an increase in the number of images and illustrations that were used in children's books (Slythe, 1970). A wide variety of graphics, images, pictures, cartoons, graphs and maps have been a part of instructional strategies. There is a tremendous increase in the number of research studies that attempt to find the relationship between effective instruction and visual elements. The main challenge, however, is to identify when and where it is appropriate to use such displays in instruction. Graphics can help, and cause no effect or harm on an attempt to teach and facilitate understanding. Samuel (1970) suggested that research that shows relations between the intent and results of graphics in education has been often jumbled. Graphical representations bring out many of the hidden patterns and elements in the raw data

(Reiber, 2000). Since humans are good at pattern recognition, efforts are always made to translate data into images or pictures. Charles de Fourcroy (18th century) suggested that the two dimensions on a sheet of paper could be used to represent something other than visible space. One can combine, juxtapose and transpose graphical images in ways that lead to groupings or the integration of information converting the image on paper to a living image, which is a research instrument and a means of justification of information derived from them.

Challenges in design of graphical displays Edward Tufte's pioneering work in the use of graphical displays to represent quantitative data has provided a number of insights and challenges. According to Tufte (1984), the challenges lie in depicting huge amounts of data in a concise, accurate way that is adequate for achieving the goals. In their analysis of research on graphical representations and static diagrams, Scaife and Rogers (1994) studied the work of Larkin and Simon's related to physics (1987) and Bauer and Johnson Laird's study of logic of problems (1993). Larkin and Simon analyzed examples taken from physics and geometry related to pulleys and weights and theorems. Their aim was to develop models that allowed a contrast between processing of sentential and diagrammatic representations of the same problem. It was found that the value of diagrams in such situations is related to experience and expertise. Novice physicists do not make the same inferences as experts (Anzai, 1991). Bauer and Laird (1993) concluded that the participants form a visual representation of the diagram and in their mind's eye they can imagine moving the pieces or switches (carry out visual transformations of images). "Bypassing the construction of the meanings of verbal premises and manipulating visual images appears to reduce the load on working memory and to speed up the process of inference" (Bauer, 1993, p 373). Stenning and Oberlander (1995) proposed the theory of specificity according to which graphical representations such as diagrams limit abstraction and thereby aid processing. This was achieved through the information available in the diagram restricting the possible interpretations of the problem and in doing so guiding the participant to make the right decisions. Thus they believed that certain diagrams are more effective than the others because they exploit better the properties of varying graphical forms.

Graphics and Media Research in computer based instruction and interactive multimedia environments focuses heavily on identifying the factors that help to bridge the gap between what is seen and what is perceived, between what is learned and the process of learning. Through such studies, the proper mapping of data to images and visuals that could help in interpretation and the comprehension of information that is being conveyed is identified as being the key. Instructional design theories and models that support constructivism (Mayer, 1992) suggest that graphical representation of information, their organization and structure are the key factors that facilitate computer based learning. Mayer (1992), suggests that encoding is the process by which the mental representations are constructed in working memory and are stored in long term memory for permanent retention. Encoding occurs when learners are able to make one-on-one connections between corresponding element or components of a representation or display. A great deal of information is conveyed by the way in which items are placed on the page or screen, relative to the frame and relative to each other, and variations in the placement and representation of items have important effects on the perception and understanding of maps and diagrams (Winn, 1990). Since they are used to illuminate data, different graph layouts can differ dramatically in effectiveness of conveying information (Huang & Eades, 2005). Three major factors are considered in the visual characteristics of graphs- format, function and purpose (Shah & Hoeffner, 2002). Graphs make quantitative information easy to understand. Encoding of the visual information depends on the interactions and relationships represented in the display (Bertin 1983; Kosslyn 1989). While line graphs represent x-y trends, bar graphs clearly indicate comparisons, and pie charts indicate relative proportions. However, studies on three-dimensional displays that represent multivariate data have been very rare and uncommon.

Multivariate graphical displays One of the most commonly used perceptual characteristics is the third domain where the retinal variables come into play. They introduce a different perspective of the two dimensional displays and tend to show interactions and relationships, which are displayed through the use of visual variables. According to Tufte (2001), visual representations of quantitative evidence should be governed by principles of reasoning about quantitative evidence. Making controlled comparisons also guides the construction of graphical displays. Principles for the design of data graphics include documenting the sources and characteristics of the data, enforcing appropriate comparisons, demonstrating mechanisms of cause and effect, recognizing multivariate nature of analytical problems, and evaluating alternative explanations.

Non-verbal elements such as pictures, images and graphs have been used for a long time in education to depict results and data from sciences and social studies. A well-drawn graph, could answer most of the commonly asked questions. However, the challenge over the years has been the levels of proficiency in understanding quantitative data presented in a graphical way—the so-called individual's graphicacy (Wainer, 1991). According to Zabell (1975), these levels of proficiency are used in combinations with one another and are used in the detection of outliers or unusual data points. However, comprehension at the intermediate level can be more challenging than those in the elementary level, since they involve the study of the relationships between the variables under consideration. The establishments of these connections are the key to understanding quantitative data in graphical displays. Graphics work the best when they have high data densities. Tufte (1983), defines density as the number of entries in the (data matrix) / (area of graphic). Graphics with low-level data densities invoke comprehension in a sequential pattern rather than spatially. Good quality graphics are comparative, multivariate, high density and show interactions (Tufte, 1999). Several instructional design theories such as Constructivism (Mayer, 1992) and the theory of Multiple Intelligences (Gardner, 1984) support the use of graphical and visual elements. They also propose methods of organization, structuring and sequencing of these elements in certain ways to facilitate learning and encoding. The rise in graphical user interfaces and the advancement in technology tools to develop these user interfaces have permitted integration of graphics and text. Development and duplication of materials are not an issue anymore.

Data visualization

According to Bertin, the best visualizations are those that allow immediate interpretation of data. An image is the unit of visualization, which the eye can isolate during an instant of perception. Green (1991) proposed that perception or information processing occurs when the image is broken down into simple representations, which are then processed by the brain. How the brain reassembles all of these simpler pieces is a mystery. Attaneave (1974) suggested that spatial location is the attribute that holds the variables together during reassembly. Space is the only attribute that is shared by all the variables. Visual and cognitive research has shown that at certain times the images are being perceived effortlessly at a faster rate than others. This is often described as the preattentive search (Cleveland, 1984). On other occasions, the user requires concentrated effort and attention to read the display and comprehend what is being conveyed.

Robinson, Robinson and Katayama (1999) conducted studies that showed that the facilitative effects of graphics organizers is due to the spatial encoding. Kirkham, Slemmer and Johnson (Jan. 2002), in their studies, showed that response to visual stimuli involving statistical patterns was invoked in the infancy stages of a child's development. These studies indicted that there was a functional statistical learning mechanism in children. The challenge was to develop and nurture the skills by identifying a method that would aid in improvement. Bertin's Image Theory is a bridge between these findings and the actual translation of multidimensional data into graphical displays.

Visual grouping, which is the process of revealing structures in images by associating common features with one another, serves a computational role in the our interpretations of the worlds (Watt, 1988). File (1979) established that in situations where closely related colors or visually related colors were used in a display, the time taken to fulfill tasks was longer. Visually distinguishable retinal variables as opposed to similar ones should thus aid in comprehension and interpretation of data.

A resemblance between representations for different values along a single dimension and their perceived referents have been demonstrated by Moyer and

Bayer(1976). The time taken to differentiate between two circles decreased as the difference in size increased. Tufte (1983) and others have argued that monochromatic (gray) scale maps are adequate for statistical maps since they provide clear progression from lighter to darker areas, which makes it easier to read them. Tukey (1979) suggests that if geographical units in a map are of greatly varied sizes, the eye is drawn to the larger areas, even though smaller areas have greater or dense populations. Color has long been used for bivariate mapping (Pickle & Hermann, 1993), but research suggests that users have difficulty understanding this format (Wainer & Francolini, 1980). For statistical maps, the use of different hues or values of the same color are not recommended. As mapping has become more of a tool for data analysis, rather than just representations of locations, more complex questions can be asked. The succcess of reading a map is a function of the many variables that affect one or more of our cognitive stages (Pickle & Hermann, 1993). Comprehension of what is being represented in a map occurs when a symbol system said to convey a message interacts with a physiological process active in the person who receives it (Solomon, 1979). According to Goodman (1968), the symbol system in graphics is highly notational. It consists of symbols, each of which bear a unique and discrete relationship with another in the domain the graphics describes and has a unique set of spatial relations with other symbols. Thus the comprehension of graphics involves the identification of the objects for which the graphic objects stand and interpretation of the spatial relations among them with reference to the domain of the graphic. The use of the eye tracker system in this dissertation study will provide an insight into the path the participants take in order to achieve comprehension.

William Winn (1994) proposed that graphics consist of objects that may be represented as words, dots, boxes or other shapes, icons, and drawings. These objects are related to each other and to the whole graphic. For comprehension, the symbols in a graphic have to be detected. Also, the spatial organization of the symbols, implicit patterns formed by the distances between the symbols and their arrangement, have to be discerned. These tasks are achieved by detecting the symbol, discriminating one symbol from another, and configurating the symbols into patterns. Discrimination establishes whether it is the same as or different from other symbols. Configuration determines which other symbols are associated with it. According to Treisman (1988), these are preattentive processes that work in parallel and draw a little upon cognitive resources. The proposed research study uses color, value and shape in conjunction to help discriminate between the variables and their relationships. The use of the eye tracker system to track trace eye scan paths, the eye gaze and the attention to a specific object or data point will provide a means to analyze the participants' mental strategies and information processing approaches. The researcher hopes to obtain data that provides insight on other processes besides detection, discrimination and configuration that may be involved in graphic comprehension.

Eye tracking and visualization

Huang and Eades (2005) studied how people read graphs. They proposed that in order to define user centered design criteria, it is important to know how people read graphs and how a particular layout characteristic can affect people's reading performance. During picture viewing, eye movements are not random; features that are interesting and stand out attract attention (Buswell, 1935; Yarbus, 1967). Huang and

Eades (1999) also used an eye tracking system to discover relationships between eye movement patterns and properties of graph drawings. Thirteen participants were recruited from the student population. They used ten drawings in the form of node edge diagrams where the nodes represented objects and the edges represented the relationships between the objects (Colin Ware et al, 2004). Based on videos gathered from the eye tracking system, they observed that participants showed very similar high-level search strategies. Participants consistently looked for highlighted nodes to determine which one to start with and then searched for possible paths. During this time period, some nodes and edges were repeatedly visited to determine and verify the answers to the searches. Huang and Eades (1999) made the following observations based on the study:

1. For different types of queries, people may use different strategies. Since this research study uses three levels of complexities in questions, use of the eye tracker may provide information as to the strategies used when the questions are of different levels of complexity.

2. The study included only questions related to business relations and friendship relationships represented by the distances and number of nodes. Hence the results of the study cannot be applied to scaled graphs and more complex questions such as those that evaluate global trends. This study is an attempt to address such higher order thinking skills.

Eye Tracking Research

Use of eye tracking systems to gather information on eye movements is relatively new. Eye tracking records the way in which the eye reacts to visual stimuli. As technology has progressed, these measures have become more sophisticated and accurate.

Two major types of eye movements are usually collected and analyzed during eye tracking research. Visual stimuli produce stages in eye movement. When the eye gaze is held constant at a stationary point, fixation occurs; movement between fixation points is termed saccade movement, and these tend to be rapid scanning processes. The velocity of these movements can be as high as 500 degrees per second. The velocity of these movements can rise and drop until the target destination is fixated on. Tasks such as those involved in the study require learners to attend to a variety of visual clues that in turn produce patterns of eye fixations and saccade movements. These patterns of movement make up the eye gaze pattern. Hoffman and Subramaniam's (1995) work supports the theory that attention precedes saccade movement to a specific place. Findings in two different experiments show that participants cannot move their eyes to a specific place while attending to a different one. Research conducted by Knowler, Anderson, Dosher and Blaszer (1995) suggests that attentional movements and saccade movements are coupled in complex tasks that involve information processing. The coupling of attention and eye movement is more related in complex tasks (Duebel & Schneider, 1996). The relationship of saccades to attentional mechanisms provides support for the study of eye movement in research.

Human eye movement in relationship to still images such as pictures, photographs, and paintings has been conducted for some time with varying degrees of sophistication and success. Some of the earliest research conducted with eye tracking, although primitive at the time, provided valuable input to support aspects of visual attention theories and human learning. Yarbus (1967) conducted studies of scan paths using still images. These early results found that the order of progression of what is visually attended to is task dependent and that, given a picture, viewers will view a picture differently depending on what they are in fact searching for. Participants changed eye movement patterns based upon changes in the verbal directions given to them.

Eye movements are fundamental to the operation of the human visual system. Due to their close relation to attention mechanisms, saccade movements of the eye, which are scanning movements that occur 3-4 times every second, provide insight into the cognitive processes such as mental imagery and decision-making. (Richardson & Spivey, 2004). Prior research studies in the area of data visualization have focused mostly on format of the displays and recall of information, but not on decision making or design considerations. Since the proposed study is based on attention theories and information processing, the use of the eye tracker system could provide insight into possible strategies used in the processing of the information presented in the display. Attending to a location in space aids processing of a stimulus at that location (Posner, 1980). Planning a saccade towards a location improves processing at that location (Hoffman & Subramaniam, 1995). The saccade movements and gaze scores recorded when participants interpret the display presented may provide information on the degree of attention and on the approach used in problem solving. The scan path recorded by the eye tracker should inform the order in which the participants look at the different data points to study interactions. Due to their selective and rapid sampling, eye movements must recruit memory processes in order to build up some representation of the visual world (Irwin, Zacks, & Brown, 1990). Conversely, memory processes often recruit eye movements (Spivey & Geng, 2001). Neuroscience suggests that to some degree memory representations recapitulate perceptual processes. (Kosslyn,1995).

Empirical studies have found that frequency of eye movements increases during mental imagery, particularly those of spatial nature (Clark, 1916; Goldthwait, 1933). Recent work suggests that not only are eye movements engaged by a memory of a specific perceptual experience but also by cognitive acts. Spivey et al (2001) asked participants to listen to narratives describing events that extended either horizontally or vertically (stories that involved a train and tall buildings). Participants' eye movements were recorded when they listened to the narratives with their eyes closed. The direction of saccade movements corresponded to the directionality of the narratives, suggesting that eye movements can accompany high-level cognitive processes. According to the theory of Neurolinguistic Programming (Bandler & Grindler, 1975), eye movements to particular locations correspond to the areas of the brain that were being used.

Eye tracking research can also inform the study of complex behaviors. Analyses have examined differences in scan paths between novices and experts. Land and Hayhoe (2001) found that eye movements are linked with goals and sub tasks. This is in contrast to the perception of static images or scenes where the eye movements may be drawn to areas of high contrast or spatial frequency. The relationship between eye movements and cognitive processing has been employed as a tool to improve ergonomic design and computer interfaces (Kramer & McCarley, 2003). Since fixations (area where attention is directed) indicate what parts of a visual are salient to the user, eye movement information can be used in the design of visual displays. According to Kosslyn (1995), given a stimulus such as an image, the entire scene is first seen mostly in the parallel through peripheral vision. At this stage, interesting features stand out in the field of view. Attention is turned off or disengaged from the foveal location and moved towards the striking features that stand out. Once the eyes complete their movement, the fovea is directed at the region of interest and attention is now engaged to the feature under inspection at high resolution.

By recording and analyzing what the human eye looks at and tracking the scan path, the research can provide data on information processing, how the participant approaches problem solving and what part of the displays were important to the participant in problem solving. The eye tracker used in this study records the fixation and saccade movements as scores measured as time in seconds.

Summary

This chapter provided a detailed comprehensive account of the literature that supports and necessitates the study. As the literature review indicates, studies have looked at data using algorithms and computer models and tested low level recall and identification of data points, but have not focused on the more frequently used two dimensional computer screen and higher order decision making skills. This study explores ways to identify design approaches to displaying multidimensional data.

Chapter Three

Method

The study compared the effectiveness of retinal variables used in conjunction for the display multivariate data. The study addressed four research questions and supporting hypotheses and examined the performance scores measured as number of correct responses, and time on task scores. In addition, the study also used an eye tracker and analyzed the fixation and gaze scores from the eye movements. The eye movements are explored by defining specific regions of interests (ROI) in each of the displays used. The purpose of the study was to analyze the implications of design of graphical displays in the interpretation of the data interactions and relationships and hence decision making using multidimensional data. This chapter describes the research design and the various procedures used to accomplish the purpose of the study.

This study includes two research instruments, the interactive display and the eye tracker both of which are discussed in detail in this chapter. An overview of the data collection methods with specific results are described in chapter Five. The findings from the pilot study are provided in Appendix M.

Population and Sample.

The population for this study consisted of undergraduate college students from five sections of a 2000 level undergraduate course in education technology from a major research university in southeastern United States. Volunteers from each of these sections possessed basic computer skills. The researcher contacted the instructors and arranged for the volunteers from each of these sections to participate in the study for extra credit. The interactive displays were installed in the desktop computers in the computer lab where the classes were being taught. After the instructor covered the learning activities and instructional content for the sessions, the volunteers who were interested in participating in the study completed the research task given to them. Participants were given ten extra credit points for their participation.

Selection of Sample

Based on an apriori power analysis (Cohen, 1988) with an alpha value of .05, a medium effect size and a power of .80, the repeated measures analysis for this study called for approximately 78 participants. In order to account for an equal number of participants in each treatment group, N=90 was calculated. However, for investigating the significant differences in interaction effects within each of the groups in the repeated measures analysis and to make inferences based on data from each of the groups independently, the population size was increased from 90 to 120, which resulted in 40 participants in each group. Due to the challenging logistics involved in having 120 participants' eye movements being individually recorded by the researcher, it was decided that twelve participants' eye movements would be recorded. With volunteers from five sections of the 2000 level course, the study had a total of 135 participants. 12 out of the 135 participants completed the task while their eye movements were recorded and 123 participants completed the task with the program loaded on the desktop computers in the lab. Thus there were 41 participants in each of the groups for the interactive displays and 4 participants in each of the treatment groups for the eye tracker study.

The pilot study included nine participants, three in each of the treatment groups. All nine participant's eye movements were recorded individually.

Random Assignment

The study involved random assignment of 135 participants to the different treatment groups. In order to ensure random assignment of the participants, all three displays for the three treatment groups were installed in the desktop computers. Participants were instructed to pick one paper from a basket containing 135 small pieces of paper with numbers one, two and three, written on them and folded twice.
The numbers one, two and three were repeated 45 times each. Thus there were 45, ones, twos and threes each in the basket. The participants completed the task on the treatment groups corresponding to the number they picked.

Research Design

The study was designed to be a true experiment with random assignment of treatment methods, with time on task and performance scores as dependant measures. The experiment used three treatment groups (color/value, color/shape, value/shape), and three levels of questions (location of data point, comparisons, global trends). Nine questions (3 data point location, 3 comparisons and 3 global trends) were used for each of the treatment groups. The research design resulted in three treatment groups, each with the same nine questions of three different complexity levels as seen in Table 1.

Table 1.Research Design using three treatment groups

	Color Value	**Color Shape**	**Shape Value**
Question Level 1	3 questions	3 questions	3 questions
Question Level 2	3 questions	3 questions	3 questions
Question Level 3	3 questions	3 questions	3 questions

Limitations of the Study

One of the limitations of the study was generalizability. Since all participants were from the college of education, the results of the study cannot be generalized. Moreover there is also a possibility that the background and prior knowledge of participants in Math and Science areas may perform better than the others. In order to evaluate this concern, the educational background of the participants was noted, when they signed up for the study.

Control for Extraneous Variables

Prior Knowledge

Prior knowledge and experience interpreting and designing graphical displays showing statistical data and their interactions can affect the performance. Participants in this study include only those without any prior knowledge or experience in engineering, computer science, or compiling and analyzing medical data. Through physical control of participants, the study controlled for the extraneous variable prior knowledge.

Computer Skills

Computer skills and practice are other factors that may affect the performance scores of the participants. For this research, it is assumed that participants were comfortable with the basic tasks (using a mouse, clicking, understanding basic navigational clues and interactions) involved in the use of a computer. The pre-service teachers who were included as participants in the study enrolled in the 2000 level

Introduction to educational Technology courses and were trained in basic computer skills. Participants in this study include only those who were not experienced or proficient in computer science, engineering, and statistical analysis.

Motivation/Interest

Another major factor, which is anticipated to contribute to variability and differences in performance, is the motivation and interest in the study of each participant. The participants were requested to complete the tasks as quickly as possible, allowing them to complete the task without much distraction and lapse of attention. After completion of the study, each participant was awarded ten extra credit points by their instructor. Acquiring extra credits that counts towards their final grades could be a motivating factor to participate in the study.

Design and Development of the Displays

Other Data Display Formats Explored

Prior to deciding on the three treatment groups, the retinal variables and the design of display, several other display types and methods were considered. The entire research study is based on Bertin's proposed image theory (1963), which states that there are two types of variables, planar (x and y) and retinal variables (color, value, shape, direction, size, texture) that can be used for the display of data. The variables shape, color and value were chosen for this study since they are the most commonly used. Cyan and magenta were the colors of choice because they are complementary colors in the color wheel. Prior studies conducted by the researcher (Kealy& Subramaniam, 2005) included red and green colors as retinal variables. Feedback from participants indicated that both colors were equally strong and competed with each other. Hence the two colors chosen for this study were colors that complemented each other.

For the displays, several models were considered. Computational and algorithmic models are commonly used to display multivariate data. Many computer visualization techniques use specialized software to study user interactions and real time visualizations (Hibbard & Santek, 1990; Upson, 1989; Wettering,V, 1990). The application of computer science to problems of data analysis (McCormick et al., 1987) has been recommended and used by the National Science Foundation. However this research study has implications to education and learning situations. Algorithmic or computational models differ from the two dimensional computer screens and paper that are most commonly used in either learning or more common situations to display data. Hence the researcher used the two dimensional computer screen display in the research study.

A table allows people to get a single point values most accurately but provides the least integrative information (Guthrie et al, 1993). Line graphs are good for depicting x and y trends, bar graphs for discrete comparisons, and pie charts are good for relative proportions. Three dimensional displays compared to two dimensional displays proved to be better when questions required integrating information across all dimensions (Wickens et al, 1994). Shah (1995), compared wireframe graph and a line graph depicting the same data. It was found that while viewers were able to describe relationships well using the wireframe, they were inaccurate in reading individual data points and were less familiar in reading such graphs. Three dimensional bars are commonly used in media and textbooks (Spence, 1990; Tufte, 1983). Tufte (1983) labeled the added dimension as chart junk and recommended keeping it to a minimum.

Data in two separate graphs was also considered for the display. This would require the participants to keep track of the graphical elements in the two different displays and comprehend the relationships and interactions. This approach imposes demands on the working memory and adds to the cognitive load. Increased cognitive load decreases processing efficiencies (Sweller, 1975). Carpenter and Shah (1998) also found that two separate graphs required participants to continuously re-examine the labels to refresh their memory and keep track of the names of the variables.

Exploring the different display formats used and the research results from previous studies in data visualization and graphical displays, it was decided that designing retinal variables in conjunction with each other (Winn, 1994) in one display would be most appropriate, effective, and suitable for this study. The display design is also based on the theoretical framework around pre-attentive and top down psychological processes.

Treatment Groups

The study includes three treatment methods designed as interactive self executable programs that were developed using Macromedia Director. The three treatment groups, include:

1. color/shape used in conjunction
2. color/value used in conjunction
3. shape/value used in conjunction

The displays used in the three treatment groups are shown in Figures 9, 10 and 11.

Interactive Displays

An interactive computer based program was developed by the researcher. Such tools have been successfully used in the past. Several research studies with similar computer based tools, have proven to be effective and successful. Using such an interactive tool, Robinson, Robinson and Katayama (1999) examined whether different word displays also differ in terms of the format in which they are represented in memory. Undergraduates studied text, an outline, graphic organizer or a concept map and then were shown either verbal or spatial displays. Kealy (2000) studied the potential of computer-based experimental treatments (CET) and the implications they pose for researchers. Segments from computer programs (developed in Authorware) were shown to demonstrate the capabilities of CET for presenting multimedia stimuli and for collecting a variety of performance data (e.g., recall, latencies, ratings). Authoring programs such as Super Lab have been used for development of experimental treatments and have been found to be an efficient tool. Robinson and Molina, E (2002) studied the relative involvement of visual and auditory working memory when studying adjunct displays using Super Lab. The trends in the development of graphical user interfaces for microcomputers have increased the availability of multimedia authoring programs (Reiber, 2002). A study by Kealy and Subramaniam (2005) used retinal variables such as color, value and shape independent of each other to display data that included more than two variables. They tried to identify the variable that is most effective in facilitating higher order thinking skills and interpretation of information from the data display. The retinal variable shape was found to be the most effective. The display that used color was chosen by participants as most useful. The study also used a similar computer based program that was developed in Macromedia Authorware. The compatibility feature of such object oriented development environments promote the sharing and reuse of objects and element between applications. Instruction can be designed to manipulate elements and a variety of icons, each representing a particular instructional function. Recent developments in technology have made possible development of data driven animations and interactivity. These systems are capable of changing variable and program values based on student input, record interactions and ratings. These are very important in visually based computer applications and learning tools.

Using Macromedia Director 8.0, an interactive computer program was created for each of the three treatment groups. The program included brief instructions, directions for use, practice displays with a question, the treatment group displays and the nine questions at different levels of complexity followed by the usability rating of the display and the description of mental strategy screens. The program was designed for an 800 X 600 screen resolution with functionality for displays and the instructions to appear full screen. The instructions were evaluated by the researcher's major professor. Grammatical errors were corrected and a couple of sentence structures were altered based on the feedback provided. . The interactive program automatically recorded the responses as correct and incorrect. The time to answer each individual question and the total time taken to answer questions at each level were also recorded by the program. The researcher then assigned scores of 0 for the incorrect answers and 4 for the correct answers for levels 1 and 2 questions. Level 3 included open ended questions. While the program recorded the amount of time taken to answer each of the questions and hence the total time, the rating of each of the open ended responses were done manually by two raters.

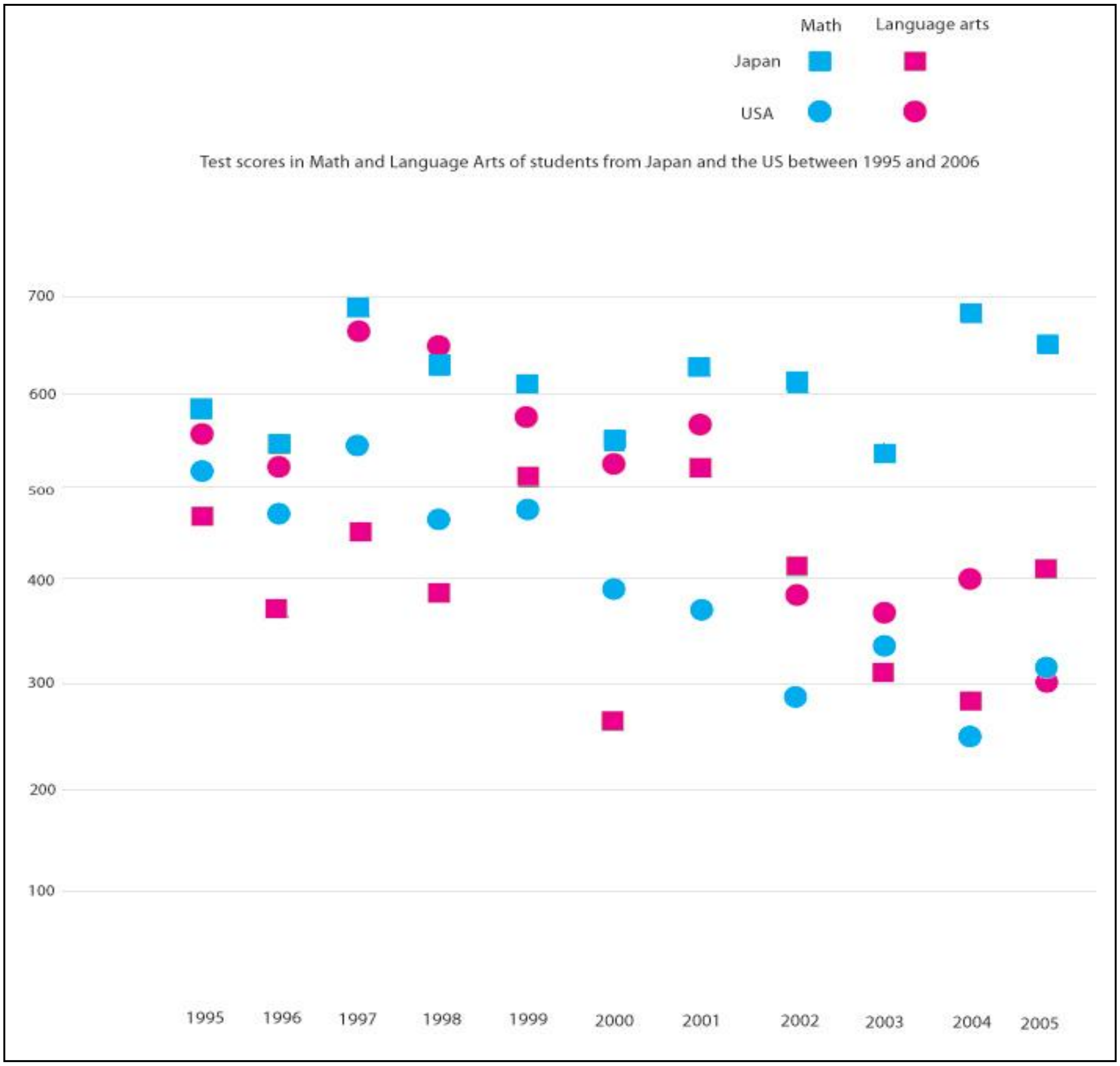


Figure 9 Display for Color Shape Treatment Group

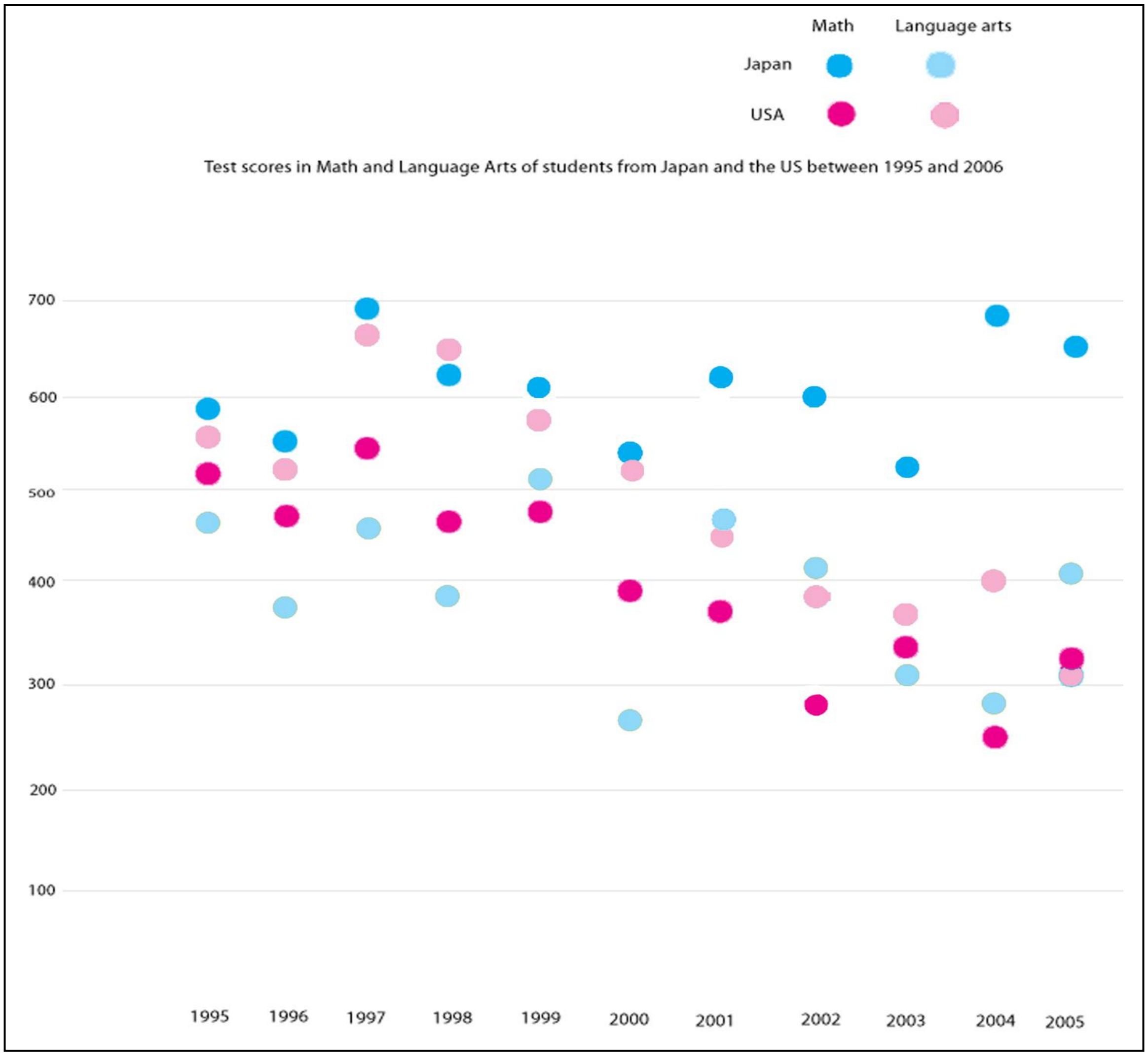


Figure 10 Display for Color Value Treatment Group

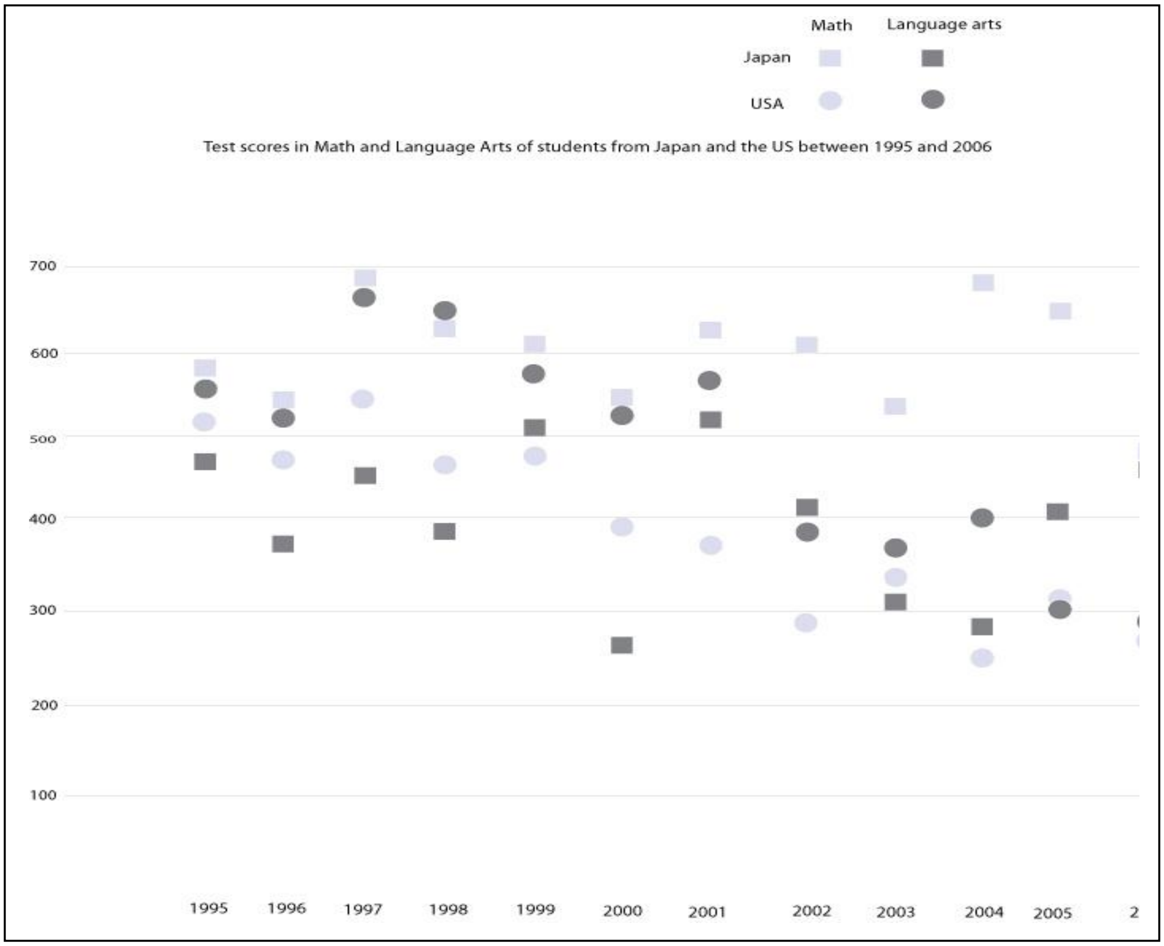


Figure 11 Display for Shape Value Treatment Group

Eye Tracker System

Besides the multimedia authored program with the three different treatment displays, the study also used the Viewpoint Eye Tracker System. The system provides a complete eye movement evaluation environment including stimulus presentation, eye movement monitoring and a software application for recording and analysis of data and its communication to other applications. The system consists of an infrared light source attached to a camera that is clamped to a ViewPoint Quick Clamp, a head positioner that clamps to a working table. The clamp also contains a chin rest that allows for proper positioning of the camera and video capture. The infrared light source serves to illuminate the eye and also to provide a specular reflection from the cornea of the eye. The video signal from the camera is digitized by the video capture device into the computer. Algorithms are applied to the digitized images to locate the areas of the pupil and corneal reflection. A mapping function transforms the eye position signals into gaze coordinates. Interfaces to the eye tracker for data synchronization, communication and control are made possible through the following:

1. Graphical User Interface (GUI)
2. Command Line Interface (CLI)
3. Software Developers Kit (SDK)
4. ViewPoint Client

Every aspect of the eye tracker is controlled by a CLI. These are ASCII string commands that allow the GUI settings to be saved for use again when the equipment is used. The CLI can send the command strings through the software developers kit functions to the ViewPoint programs running on the same or other remote computers via an inter computer link. The ViewPoint client in this study runs on the same computer as the stimulus display and communicates with the eye tracker.

Regions of Interest (ROI)

The stimulus area in the eye tracker can be divided into regions of interests (ROI). These are useful to find out whether or not a participant looked or gazed at a certain area. In the View Point system, it is possible to specify 0-99 regions to aid in data analysis. These regions are associated with the image that they are related to and the settings are stored. Every time the image with the defined ROI is opened, View Point associates the image with the ROI settings. When the gaze position moves inside a ROI, the status window in the eye tracker displays the region number and the value is stored in the data file. If two areas are looked at, at the same time and overlap, both the regions are recorded. Areas that were not looked at are indicated with a -1 in the data file. In this study, the researcher defined 12 regions of interests, based on the answers to each of the questions. Data points that represent the correct responses to the questions in levels one and two were also included in the regions. The display legend was also included as a region of interest to determine how often and how long the participants looked at the legend. Each of the three treatment groups displayed the same regions of interests. Before the participant began the task of answering questions, the image with the associated ROI was opened up in the gaze window.

Regions of Interests (ROI) in eye tracking studies have been used as early as the 1960s. Yarbus (1967), measured participants' eye movements over an image while they answered certain queries. The eye movements recorded showed a sequential pattern in the way the eye fixated on certain regions. Norton and Stark (1971) conducted their own studies based on Yarbus's observations and coined the term scan paths. They observed that even without leading questions, participants fixated on certain areas of interests or informative details. Furthermore, the order of the eye movements was quite variable. Even though there were similarities in the regions that participants looked at, there were differences in the order in which they looked at regions. The research studies suggest that a picture of the visual field is created piece-meal through the assembly of serially viewed regions of interests.

Gaze determines a person's current line of sight or point of fixation. The fixation point is defined as the intersection of the line of sight with the surface of the object being viewed such as the screen. Eye gaze direction can express interests of the user since communication through the direction of the eyes is faster than any other mode of communication (Zhu & Ji,2004). Several studies in cognitive psychology have used regions of interests to identify patterns in response to stimuli. Kiln,et.al (2002), measured visual fixations in four defined regions of the mouth, face, eyes and certain objects in 15 cognitively able males with autism. Statistical analyses compared fixation time on regions of interest between groups and correlation of fixation time with outcome measures of social competence. McCarthy , Sasse., and Reigelsberger,(2005) tested the web design guideline that requires designers to create navigations on the web that are in line with user expectations. McCarthy and his research term found that users rapidly adapt to unexpected screen layouts and hence designers should not be restricted to design conventions and guidelines as long as there is consistency

In this research study, the regions of interests helped to identify the areas of interest, the areas on the display the participants fixated on while completing the given task, and the sequence of the eye gaze in these regions. The trends in the glance frequencies (the total fixation in each of the regions of interests) and identification of patterns that participants followed were investigated. The regions of interests defined in each of the treatment groups are shown in Figures 12, 13 and 14.

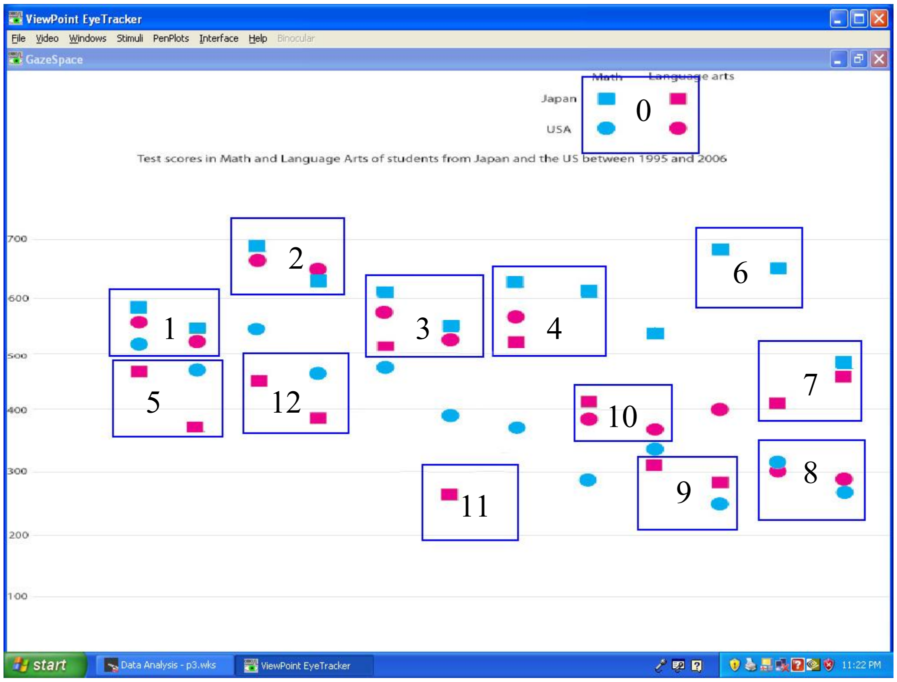


Figure 12 Regions of Interest defined in Color Shape Treatment Group

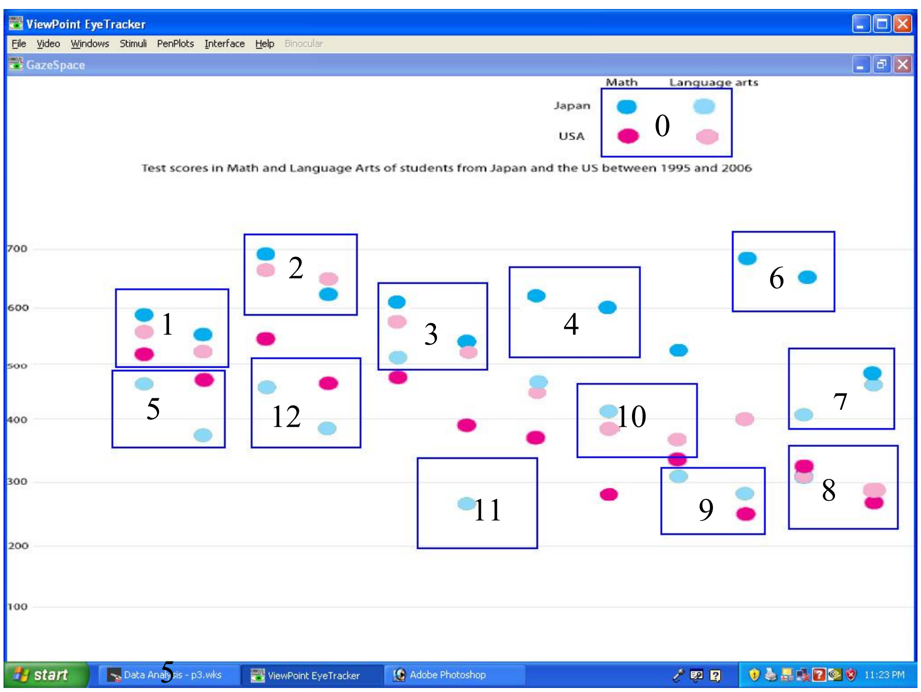


Figure 13 Regions of Interest Defined in Color Value Treatment Group

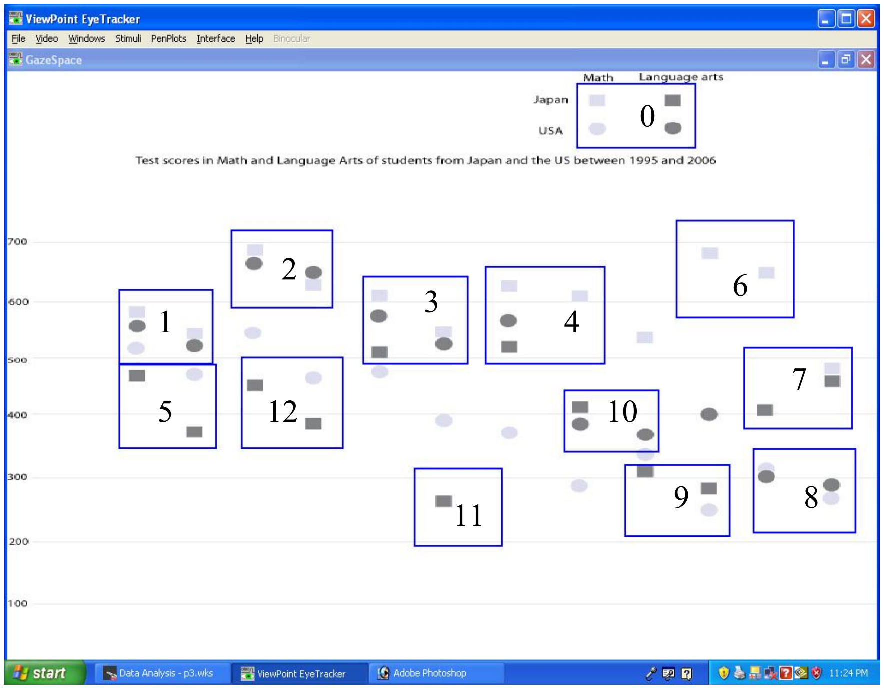


Figure 14 Regions of Interest defined in Shape Value Treatment Group

Criterion Measures

Performance and Time on Task Scores from the Interactive Displays

The treatment groups used retinal variables, in conjunction, to display scores obtained by Japanese and American students in math and language arts between the years 1995-2003. The research study involved 9 questions the participants had to answer based on the display (see Appendix E). The 9 questions addressed the three levels of complexity (Wainer, 1994). Questions in level 1 required participants to identify specific data point locations to answer the question. An example of level 1 question is "Identify the highest score obtained in Language arts among both Japanese and Americans". Questions in level 2 involve the identification of data and their relationships. An example of question 2 would be "Between the years 1997- 2001, in math among Japanese and American students, what is the lowest score? Click on the data point that shows the correct answer. Level 3 questions require participants to compare global trends in the display and make decisions. An example of a level three question would be ―The US department of Education is investigating scores of Americans in math to identify the years where their performance in math declined. Looking at the data display, explain when you decline happened? List the in which the performance was low performance is low. Provide data to support your results". Participants were required to answer all nine questions, which were the same for the three treatment groups. Time taken to answer the different levels of questions and the total time taken were also measured. The performance scores (measured as the number of correct answers) were also recorded

Usability Rating and Mental Strategies

The participants also rated the usability of the data display. The rating scale ranges from not useful to very useful. Weights are assigned to each of the ratings in the scale; 1=not useful, 2=somewhat useful, 3 neutral, 4=useful, 5=very useful. Additionally the participants were asked to describe any mental strategies that they used while performing the task. Based on the research questions that needed to be answered through the study, the responses related to mental strategies were coded into three categories that include

1. Any methods or strategies used to answer the questions that are any mental tasks or approaches the participants used to complete their tasks.
2. Biases or perceptions described in response that relate to the participants' perceived knowledge about graphs, the three levels of questions which might have influenced their approach to completing the task.
3. Any challenges faced by the participants in completing the tasks.

Scores from Level 3 Questions

Level 3 questions used in the study included open ended response questions. Evidence of efficient decision making and higher order thinking skills are showcased by the ability of a participant to articulate their thoughts and provide rationale for their choice. A continuous piece of prose written in response to a question is commonly intended for higher cognitive levels of learning (Biggs, J.1999). Such open ended responses are best demonstrative of the participant's declarative knowledge and their understanding of the core content. This study required participants to take a holistic view of the interactions and relationships in the data display in order to answer the questions. Chin, and Brown, (1999) conducted studies to understand differences between deep and surface learning in Science. They found that when students used a deep approach, they ventured their ideas more spontaneously; gave more elaborate explanations which described mechanisms and cause-effect relationships or referred to personal experiences; asked questions which focused on explanations and causes, predictions, or resolving discrepancies in knowledge; and engaged in "on-line theorizing." Students using a surface approach gave explanations that were reformulations of the questions, or macroscopic descriptions which referred only to what was visible. Their questions also referred to more basic factual or procedural information.

A rubric was used to assess each participant's response. The rubric (see Appendix F) allowed for the evaluation of the responses provided by the participants to the level 3 questions. Two raters scored the responses independently. The scores ranged from 0-4 points. Responses that described in detail the data trends and supporting evidence were awarded four points. Answers with little or no details, and with no data trend explanations were awarded one point. Incorrect answers were awarded zero points. Cronbach's alpha used in the study to measure the inter-rater reliability is discussed later in this chapter.

Scores from the Eye Tracker

The study included saccade and fixation scores from the eye tracker measured as time in seconds as outcome measures. In addition, twelve regions of interests were defined and the scan paths from the eye movements were recorded. Glance frequency defined as the number of times a specific region is being looked at consecutively was also noted. Both the interactive displays and the eye tracker recorded the data in real time as the participants completed the tasks. The researcher then transferred the data into MS Excel, organized and formatted them. SPSS version 16.0 was used for data analysis.

Logistics of the Research and Equipment

A sign in sheet was made available to the participants who entered their background information before they were assigned a treatment group. Once the entire class of volunteers had signed in and their knowledge and experiences were verified, the researcher passed around to the participants a basket with small pieces of paper with the numbers one through forty five written in them. Each participant picked a piece of paper and was randomly assigned to a treatment group.

Computer Equipment used in the Study

The interactive displays, used as the three treatment groups in the research study, were loaded on to the desktop computers in the lab where the sessions of the 2000 level undergraduate course were taught. The researcher completed this task before the sessions began and the instructor arrived. After loading the self executable file, the researcher checked to ensure that each of the files were functional and worked without any errors. All three treatment groups were loaded on to each of the computers to allow for random assignment of the participants. The resolution of the monitors was checked and the researcher ensured that they all had the same screen resolution.

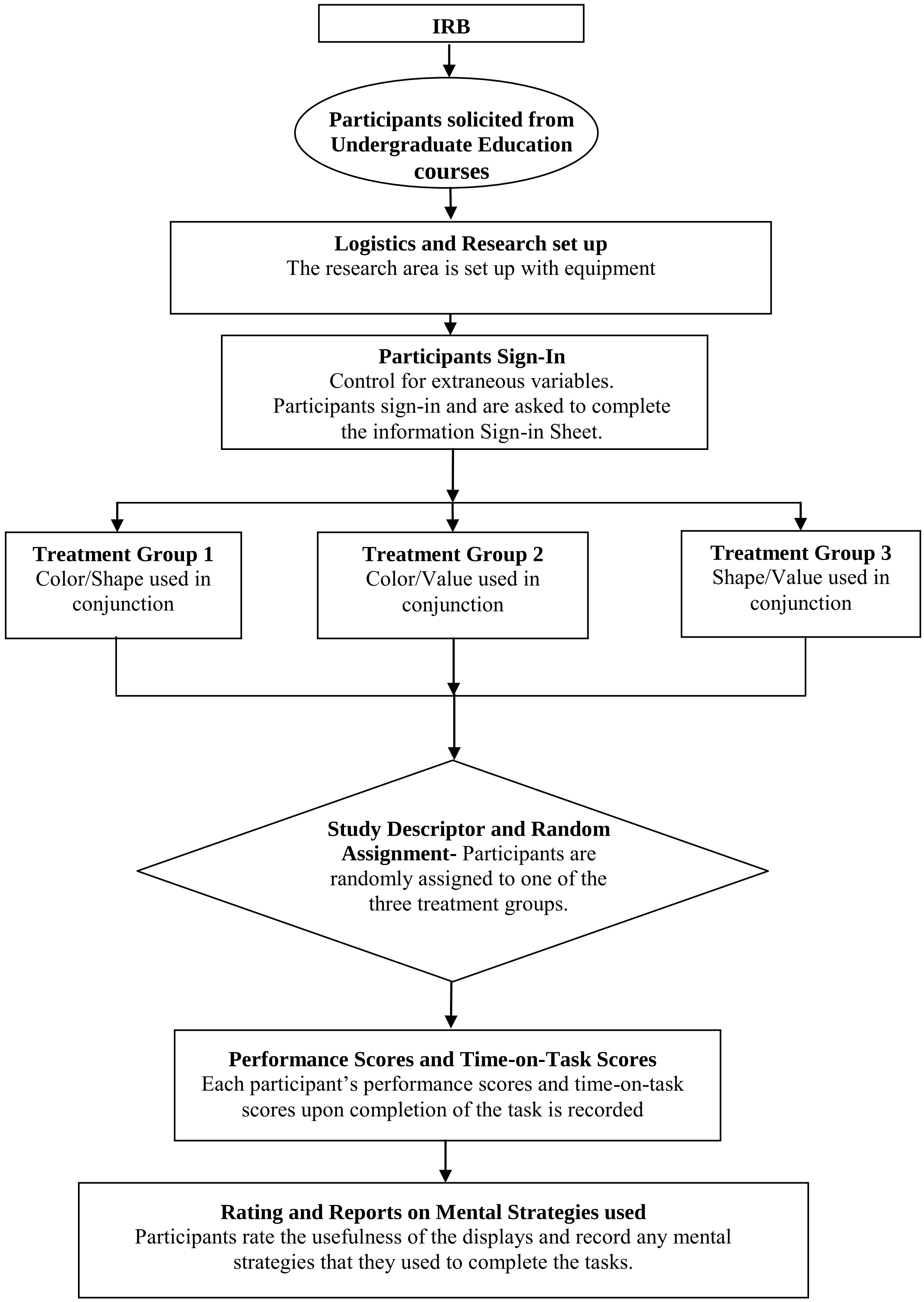
Eye Tracker System Set up

This research study uses a View Point Eye Tracker® system to record the eye movements of participants as they complete their task. The View Point system was purchased from Arrington Research, founded in 1995 as part of the technology transfer initiative at the Massachusetts Institute of Technology. Arrington research eye trackers are used in several research initiatives across the world in areas of Psychology, marketing, multimedia, neuroscience, usability, sports and training. The system applications include diverse projects. The ViewPoint EyeTracker® was recently used in a field study to explore high speed navigation on board a 36 ft. large RIB (rigid inflatable boat) in the Swedish archipelago in speeds up to 45-50 knots, in neuro marketing to study the brain of a shopper, by several academic institutions for studies in neuroscience, psychology and usability analysis. In addition the View Point data analysis software application is integrated into several third party tools such as MatLab Enabled, PsychoToolbox, SuperLab and many others

In order to capture data from the eye tracker, a frame grabber card that converted the eye data was installed on a PCI slot at the back of a Gateway™ Multimedia PC tower was used. A monitor similar to the ones that had the three treatment groups loaded was connected to the tower, along with a keyboard and a mouse as input devices. The ViewPoint software that was used for the analysis of data collected by the eye tracker was installed in the PC. A lipstick camera used by the eye tracker system to capture the movements of the eye and hence the gaze and the fixation scores was mounted on a quick clamp. These scores were fed into the PC tower through the frame grabber card, and the data packets were calculated and reported by the system software. The quick clamp along with the lipstick camera was then mounted on a desk in the back of the classroom. The participants were each seated in a chair in front of the monitor and the chin rest was adjusted to ensure that there was little or no movement of the head as the task was being completed. The stimulus for the eye tracker was defined as the interactive monitor display as the three treatment groups were used based on the random assignment of the participants. The researcher described to the participants the details about the data the eye tracker captures and how such data was going to be used in the study. Once the participants confirmed that they understood the details and had no questions regarding the eye tracker, they were asked to begin the task.

The View Point Eye Tracker® software captures data real time and also makes it available for posthoc analysis. The data is stored in ASCII format and are easily exported to MS Excel format for analysis. The regions of interests are captured, the eye movements are recorded real time and the scan paths are displayed as the participant is performing the task. The eye movements are also recorded and output as a video file for further analysis. Fixations scores at each of the regions of interests are also displayed and available for a posthoc analysis.

Research Procedures



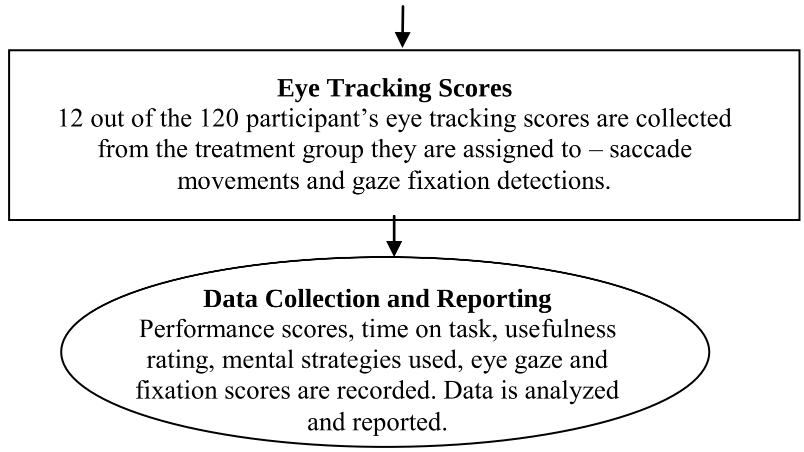


Figure 15 Flow Chart Showing the Research Procedures.

Figure 15 shows a flowchart which lists the research procedures step by step. Each of the procedures is explained in detail in the following section.

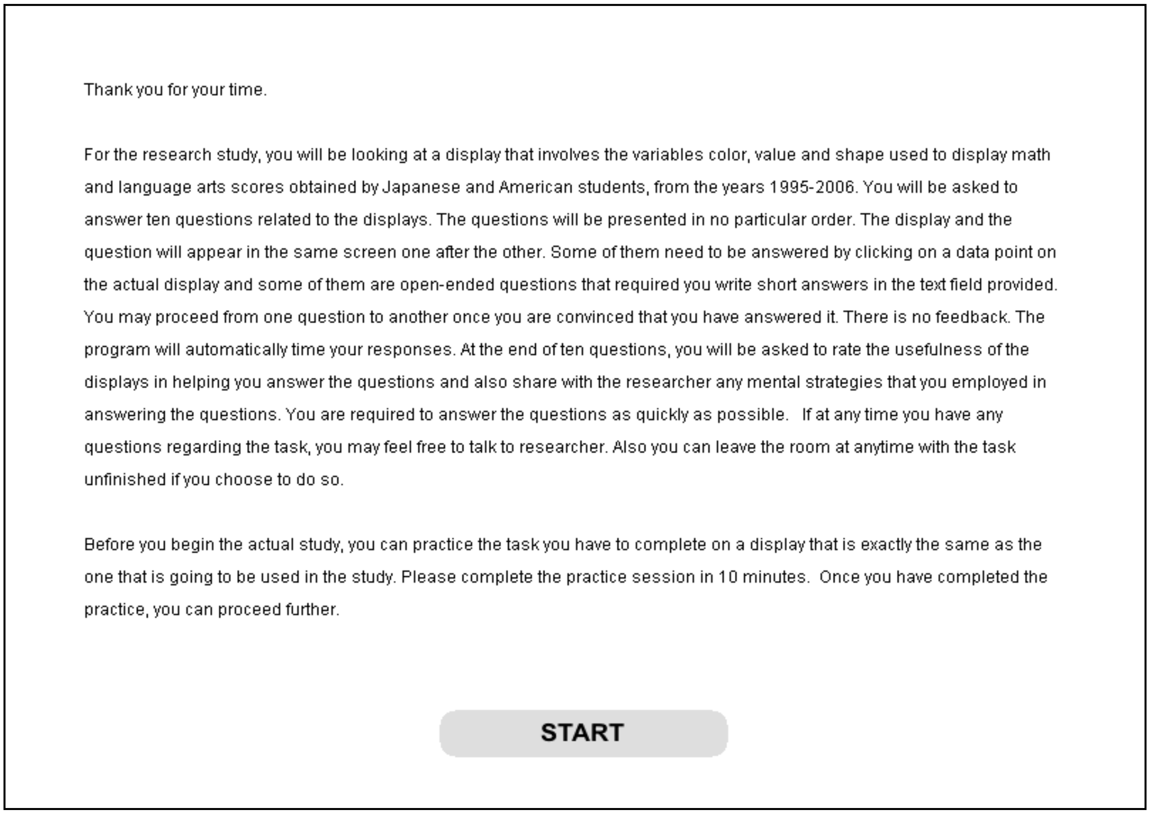
Institutional Review Board

An Institutional Review Board (IRB) application for an exempt research study was submitted to the university. The initial approval was obtained by the researcher before the pilot study was conducted. Based on the results obtained from the pilot study and input from the dissertation committee members during the proposal defense, a modification form was submitted and approved by the IRB office. The final research study was conducted after the IRB approval was obtained.

Methods

After making prior arrangements with instructors, the researcher visited the classes and shared with students the purpose of the study and provided a brief description of the procedures. Participants were informed about the task they were to perform. The researcher let the participants know before the start of the session that they could leave the session, not completing the task, at any time if they wanted to. A sign-in sheet (see appendix A) was passed around in which the volunteers shared with the researcher their educational backgrounds and experience by checking the appropriate choices provided in the sign in sheets. The participants were also informed that they will be given extra credit for their participation. The educational backgrounds and experience details obtained from the sign-in sheet allowed for physical control of those participants who had prior knowledge and proficiency in computer science, statistical analysis and engineering. In this study, none of the participants indicated any such information. Hence all volunteers were included in the study.

Before each participant began the task assigned, the researcher provided a detailed description of the task, the practice displays, the scoring and the criterion measures, the treatment groups involved and what the participants can expect to see on the screens. The participants were also asked to answer the questions as quickly as possible. Figure 13 shows the study descriptor provided in the first screen for the participants. After ensuring that the participants had no questions related to the information shared, and that they understood the task that needs to be performed, they were allowed to proceed to the practice sessions. Figures 14, 15 and 16 show practice displays similar to the actual displays used for the study and also present a question to help the participant understand how the interactivity works and how to answer the questions. The participants were allowed to practice for approximately ten minutes. The practice session was intended to orient the participants towards the task and familiarize them with the displays and the interactions made possible by the researcher to answer the questions



Thank you for your time.

For the research study, you will be looking at a display that involves the variables color, value and shape used to display math and language arts scores obtained by Japanese and American students, from the years 1995-2006. You will be asked to answer ten questions related to the displays. The questions will be presented in no particular order. The display and the question will appear in the same screen one after the other. Some of them need to be answered by clicking on a data point on the actual display and some of them are open-ended questions that required you write short answers in the text field provided. You may proceed from one question to another once you are convinced that you have answered it. There is no feedback. The program will automatically time your responses. At the end of ten questions, you will be asked to rate the usefulness of the displays in helping you answer the questions and also share with the researcher any mental strategies that you employed in answering the questions. You are required to answer the questions as quickly as possible.   If at any time you have any questions regarding the task, you may feel free to talk to researcher. Also you can leave the room at anytime with the task unfinished if you choose to do so.

Before you begin the actual study, you can practice the task you have to complete on a display that is exactly the same as the one that is going to be used in the study. Please complete the practice session in 10 minutes. Once you have completed the practice, you can proceed further.

START

Figure 16 Study Descriptor Screen Explaining the Study and the Task that needs to be Performed

The practice displays were designed to be very similar to the experimental displays. Participants found the practice display very useful in helping them understand how the displays worked and what the task involved.

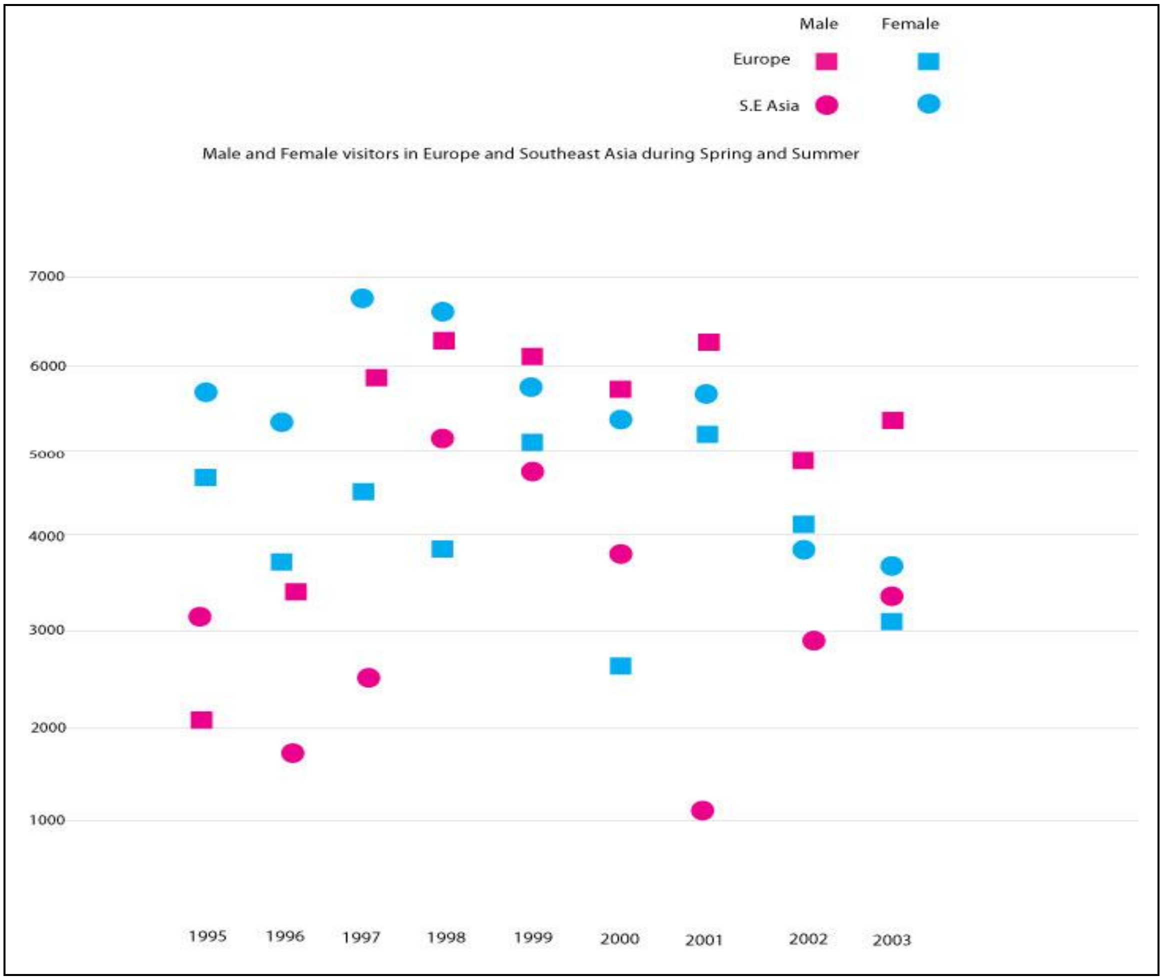


Figure 17 Sample Question in Color Shape Treatment Group

Figure 18 Sample Question in Color Value Treatment Group

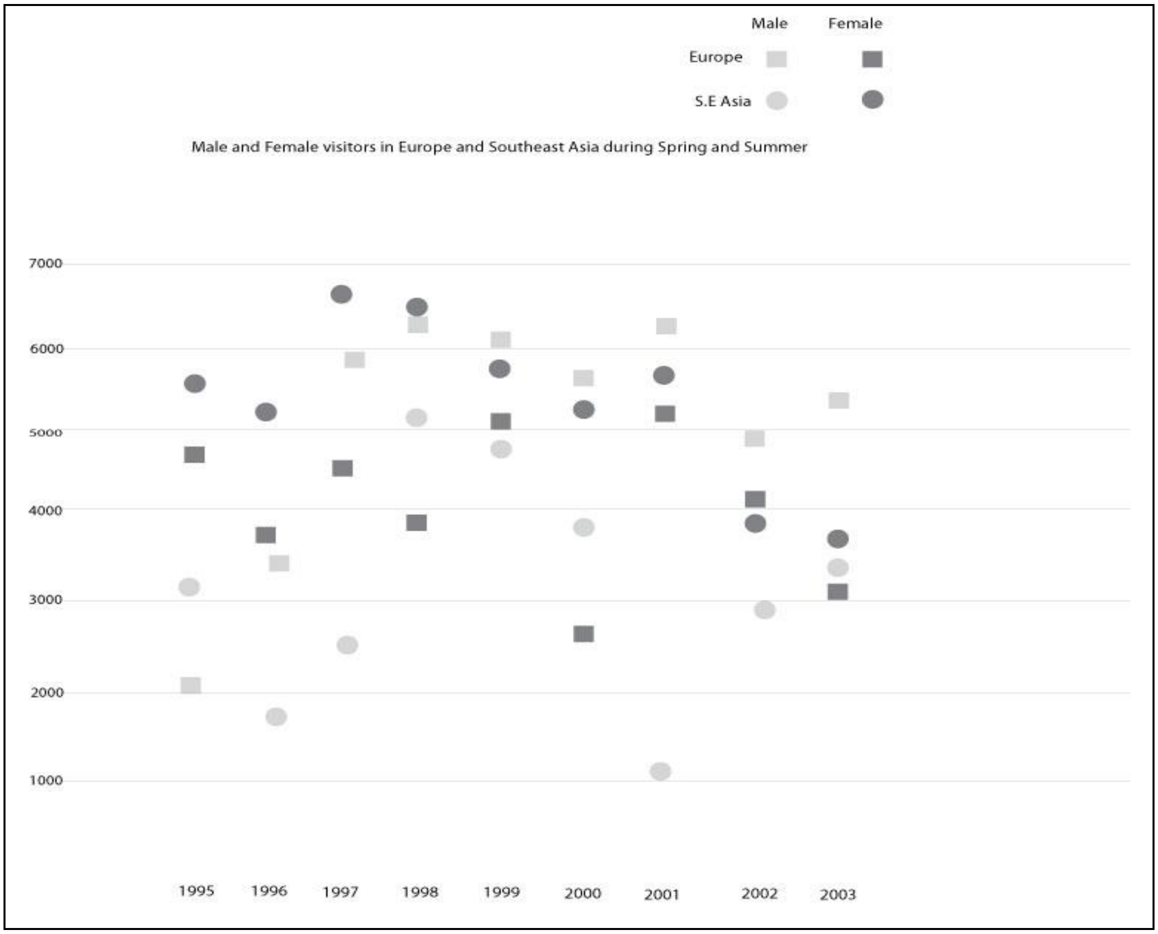


Figure 19 Sample Question in Shape Value Treatment Group

As the participants completed the sample practice question, they were instructed to wait until all others were ready to move to the next screen. When all participants completed their sample questions and ready to proceed, they were asked by the researcher to click the NEXT button on the screen and proceed further. This was done to ensure that all participants started approximately at the same time. The screen in the interactive program showing the NEXT button which leads the participant to the next screen with the experimental questions is seen in Figure 17.

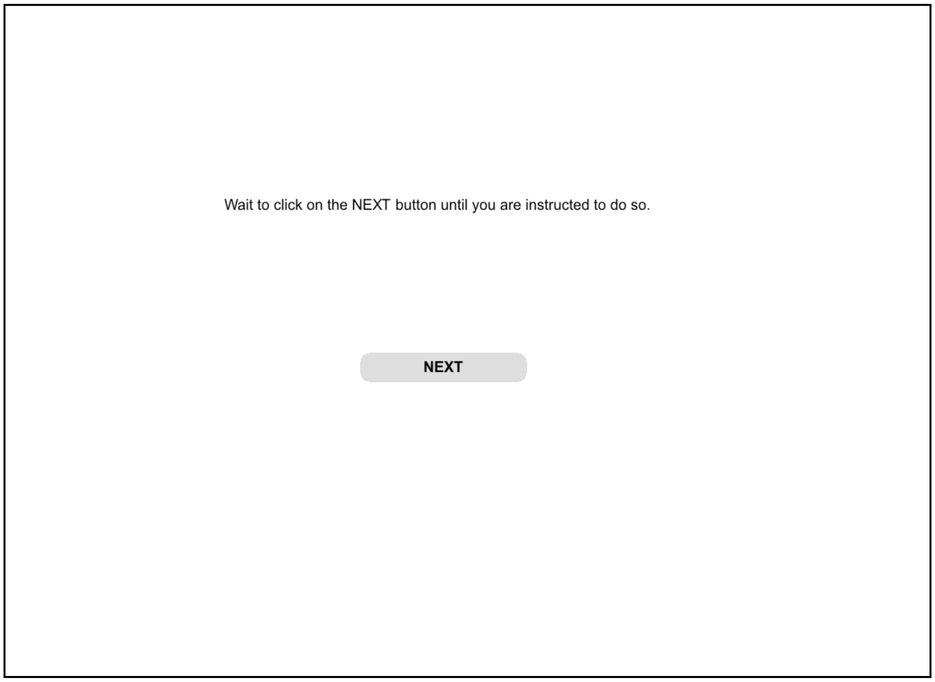


Figure 20 Screen Showing the NEXT Button

Upon clicking the NEXT button the experimental questions were presented. The study included three factual questions (level 1), three comparison questions (level 2) and three open ended questions (level 3) as seen in Figures 18, 19, 20. Level 1 questions required participants to click on a data point in a specific location in the display, level 2 required comparisons to be made before identification of data points, and level 3 questions needed high level scans across the display for global comparisons to be made. Level 3 questions were open ended and required participants to write their answers along with supporting data and a rationale for their choices.

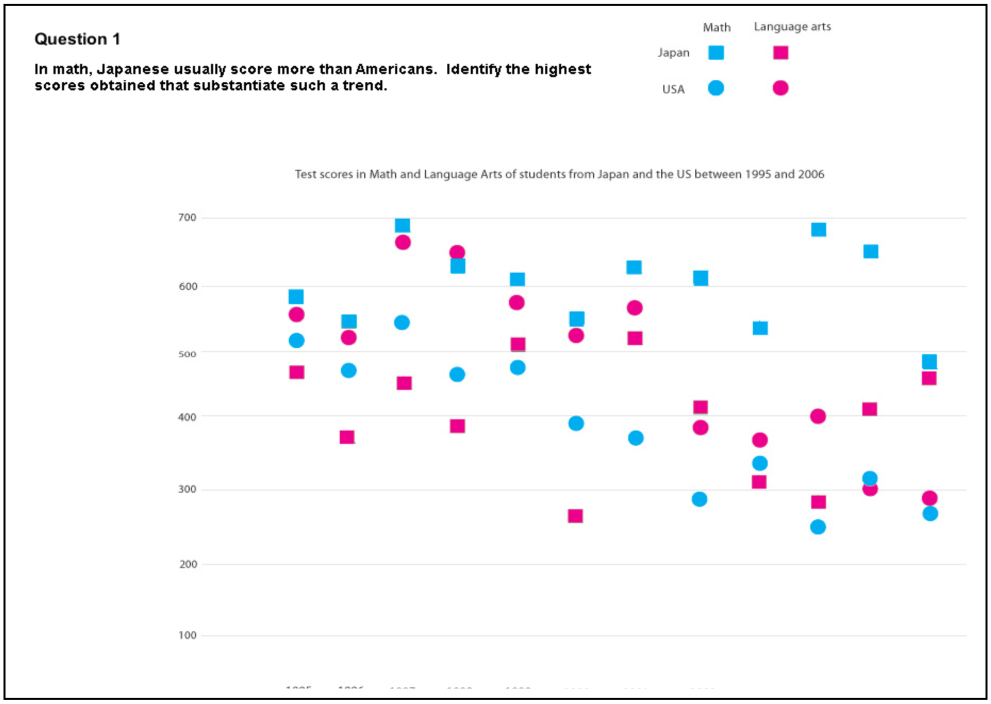


Figure 21 Question Level 1

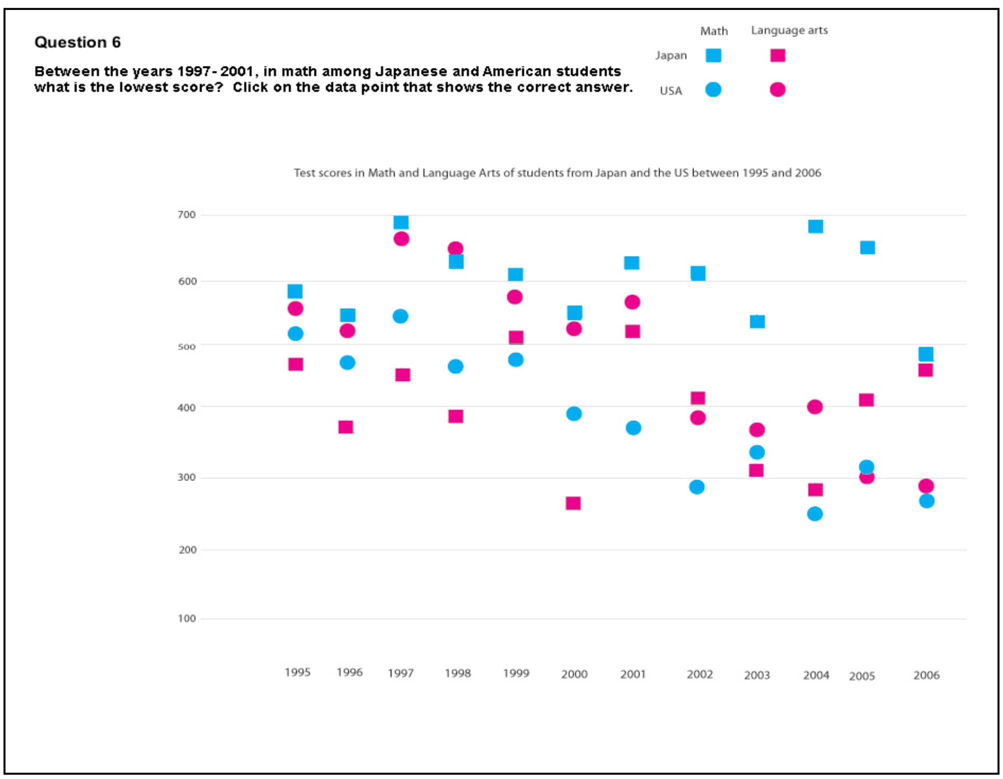


Figure 22 Question Level 2

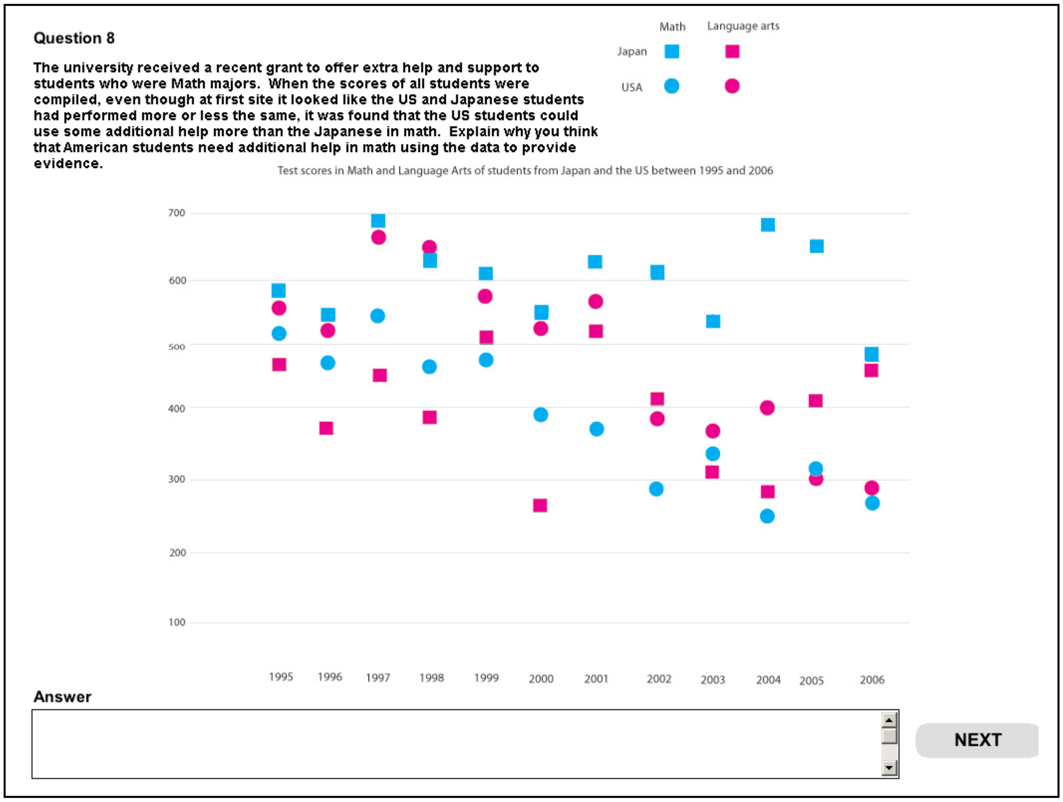


Figure 23 Question Level 3

After answering the nine questions displayed, the participants were asked to rate the usefulness of the display. Figure 21 shows the screen where the participants rated the usefulness of the display and shared any mental strategies they used to accomplish the task. The researcher provided examples of mental strategies that included thoughts that occurred in participants' minds as they tried to solve the problem on hand or any tricks or patterns they followed or used to respond to the questions.

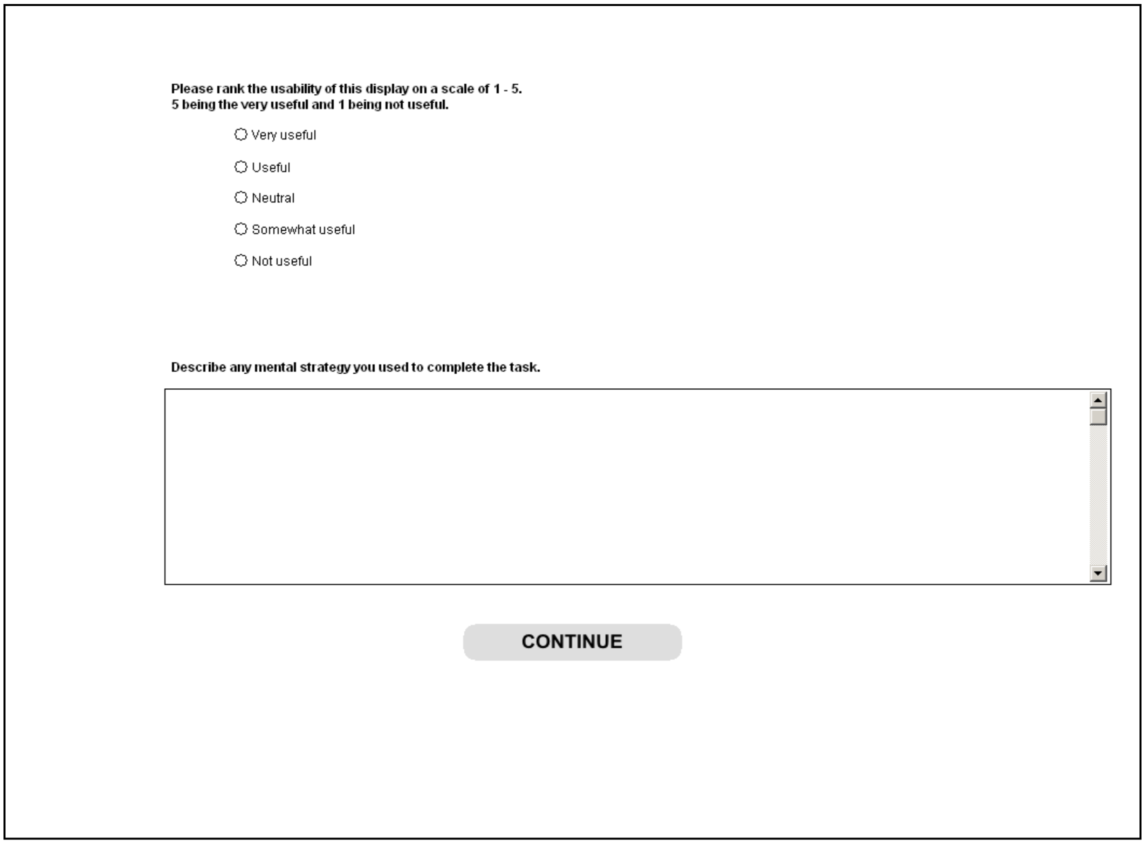


Figure 24 Screen Showing Questions Related to Usability Rating and Mental Strategies

After completion of the tasks, the participants were thanked for their time. The study did not involve deception of any kind. The questions for all the three treatment groups were the same. The only difference was the combination of retinal variables used in conjunction, color/value, value/shape, and shape/color, which made up the three treatments. Each participant's performance scores were measured as the number of correct responses. The time taken to answer each of the differnet levels of questions and the overall total time on task, were automatically recorded by the interactive program into a text file on the desktop. Scores of each of individual participants were saved in a flash drive and data was transferred from the desktop computer to the flash drive through a USB port. The usefulness of the display as indicated by the rating, descriptions of mental strategies written in response to the question presented on the screen to the participants were also recorded

Eye Tracker System

Twelve participants from the larger group of N = 135 volunteered to complete the task while their eye movements were recorded by the eye tracker. The eye tracker system was set up in the back of the classroom. Once the chin rest was adjusted to fit with the height of each participant, the researcher explained the details of the eye tracker system and the scores recorded by it. The stimulus was set to the interactive display on the monitor and the participant was asked to start the task. The displays with the regions of interests (ROI) defined were loaded to the system to record the eye gaze in each of the regions. The data for each participant was stored as a separate file to be analyzed by the View Point software. The eye tracker recorded the saccade and the fixation scores, the regions of interests that were visited and the fixation scores in each of the regions.

Instruments used in the study

Rubric

The research study included the use of rubric as seen in Appendix E to score the level 3 open ended responses in each of the three treatment groups. Two individuals rated the responses independent of each other. Frequently, after rating twenty of the 405 responses from the 135 participants, the raters checked in with each other and evaluated their rating approaches. If the scoring approaches for any of the items were found to be different for the two raters, and the inter rater reliability was less than 95%, then the rating was re-evaluated. Raters identified the differences in the rating, discussed their approaches and rationale and made revisions as needed based on the consensus they reached. This was done to ensure the inter-rater reliability of the rating by the two individuals. Inter-rater reliability was 99% when calculated after all the items were scored. Results are shown in Table 2.

Scoring rubrics are descriptive scoring schemes that are developed by teachers or other evaluators to guide the analysis of the products or processes of students' efforts (Brookhart, 1999). Rubrics are used when the quality of a diverse range of tasks need to be assessed. In this study one of the main criteria for the use of the rubric was to assess the participants' persuasiveness of the argument towards their choice and decisions. Accurate responses followed by data and reasoning to substantiate their responses was what required to be assessed. Every participant who was able to make the argument and provide sufficient evidence to support it would indicate understanding of the data interactions and relationships in the data display. In the rubric each score category describes the characteristics of the responses that would receive that score.

Table 1 Inter-Rater Reliability Results

Inter-Item Correlation Matrix		
	Rater A	Rater B
Rater A	1.000	.994
Rater B	.994	1.000

Writing samples have been evaluated using scoring rubrics. Scoring rubrics have also been used to evaluate group activities, extended projects and oral presentations (e.g., Chicago Public Schools, 1999; Danielson, 1997a; 1997b; Schrock, 2000; Moskal, 2000). They are equally appropriate to the English, Mathematics and Science classrooms (e.g., Chicago Public Schools, 1999; State of Colorado, 1999; Danielson, 1997a; 1997b; Danielson & Marquez, 1998; Schrock, 2000). Both pre-college and college instructors use scoring rubrics for classroom evaluation purposes (e.g., State of Colorado, 1999; Schrock, 2000; Moskal, 2000; Knecht, Moskal & Pavelich, 2000). Where and when a scoring rubric is used does not depend on the grade level or subject, but rather on the purpose of the assessment. The rubric used in this study was evaluated and approved by the researcher's major professor with many years of research experience in the area of graphical displays.

Scores from the Interactive Displays for Level 1 and Level 2 Questions

The interactive computer program developed using Macromedia Director automatically scored the participants as they answered the level 1 and 2 questions. Correct responses for each answer were awarded 4 points and incorrect answer zero points. The program also recorded the scores for each question. In order to evaluate the dependability of these scores, reliability coefficients for each of the levels of questions were calculated. The three different treatment groups displayed the same questions in each level. In a period of a week, data was collected from participants in five different sections of a 2000 level course. The screen resolutions of the display monitors of the desktop computers used for the study were checked before the study began. The screen resolution was maintained at 800X600. Each monitor was also calibrated to ensure that the settings for color display, brightness and contrasts were consistent. The logistics of collecting the data from a pool of 135 participants, differences in the motivation levels of participants to complete the study, the small number of items in each of the levels may contribute to the variations in the scores. Reliability analysis of the scores at each of the question levels is shown in Table 3 and 4.

Table 2 Reliability Coefficients for Questions at Level 1 for the Three Treatment Groups

Inter-Item Correlation Matrix

	Color/Shape	Color/Value	Shape/Value
Color/Shape	1.000	.117	.155
Color/Value	.117	1.000	-.185
Shape/Value	.155	-.185	1.000

Table 3 Reliability Coefficients for Questions at Level 2 for the Three Treatment Groups

Inter-Item Correlation Matrix

	Color/Shape	Color/Value	Shape/Value
Color/Shape	1.000	-.033	-.132
Color/Value	-.033	1.000	-.202
Shape/Value	-.132	-.202	1.000

Reliability analysis indicated inconsistencies in participant responses to the questions presented at each of the levels. There are several factors that may have contributed to such variations. As mentioned earlier, only three questions were presented in each level. The participant responses produced scores that were homogenous. Differences in skill levels, experiences and perceptions might have contributed to the inconsistencies in the scores. The same questions were presented to all participants at different times during data collection. However the treatment groups and hence the displays were different. These differences and interaction of participants with the display provided might have also caused inconsistencies. The motivation levels of the participants contribute to the performance research studies. Motivational factors may have also contributed to the inconsistencies in this study as well.

Reliability however, is a property of the scores obtained by the specific population of participants used in this research study (Feldt & Brennan, 1989) and does not indicate a generalizability of the coefficients across all populations. Dependability of the automatic scoring by the program and hence the scores are discussed in the results chapter as well.

Usefulness Rating and Mental Strategies

Rating scales have been used in several studies to get an indication of the user's perspective. Ratings are used in a variety of ways in social sciences, education and in behavioral psychology. Rating scales are used quite frequently in survey research and there are many different kinds of rating scales. A typical rating scale asks subjects to choose one response category from several arranged in hierarchical order. Either each response category is labeled or else only the two endpoints of the scale are "anchored" (Freidman & Amoo, 1999). The rating scale used in the study is a likert scale, which is also a nominal scale. The question and the scale were designed by researcher, reviewed by the major professor. Participants did not have any difficulties understanding the usefulness rating question in the pilot study. Hence no changes were made to the question and the scale used. The participants were asked to also describe any strategies they used to answer the questions.

Coding and thematic analyses of information provided by participants is often used to quantify qualitative data. Blasé (1986) used coding in a qualitative study of teachers' perception of work related stress. Coding strategies have been identified with one shot interviewing (Craswell, 1997). In the qualitative study used to examine the ways in which African American women use religion/spirituality to cope and to construct meaning in times of adversity, Mattis (2002) through coding showed that African American women used religion and spirituality to cope with reality, to gain courage, and recognize purpose and destiny. Spiggle (1994), explored the use of thematic analysis and coding in research related to consumer research. Gerbert et.al (1999) used coding to investigate how physicians with expertise in domestic violence identify victims. Data analysis is the most difficult and most crucial aspect of qualitative research. Coding is one of the significant steps taken during analysis to organize and make sense of textual data (Basit, 2003).

Data Collection and Reporting

Upon completion of the research, all of the scores recorded automatically by the interactive computer program as text files were saved to a flash drive from each of the desktop computers. The data from the eye tracker was stored in the Gateway™ Multimedia PC tower for access. The data files from the eye tracker were analyzed using the View Point Data Analysis software and output as a numeric text file. The data stored in the interactive software, was imported into an excel spreadsheet, correct responses were given a score of 4 and incorrect responses were given a score of 0 for question levels one and two. The responses for level 3 open ended questions were also exported to a separate excel spreadsheet and made available for rating. The usefulness rating and the descriptions of the mental strategies were also tabulated in excel.

In Chapter IV, the results of the data analysis are discussed and supporting displays and graphs are provided.

Pilot Study

To test the displays and the reliability of the tool developed, a pilot test with three participants for each of the three treatment groups was performed. This enabled the researcher to understand the needs and requirements of the participants and make changes to the final study to better accommodate the needs. The pilot study also helped test the efficiency and the usefulness of the instructions available to the participants. The efficiency of the multimedia program in collecting the data was also assessed. The eye tracker system was evaluated and any modifications that need to be made to the system, its use and configuration before the actual study were noted. The scores obtained from the eye tracker helped identify the data that provide the most pertinent, useful information and those that were not required.

The participants consisted of individuals with basic technology skills, who used computers for some of their everyday tasks. None of the participants had prior experience in statistics, engineering or 3D visualization. The pilot study followed all the procedures discussed. The pilot study was successful and did not require any changes to the instruments used and the procedures. However, the sample size in each treatment group was increased to 40 to investigate interaction effects within groups. The logistical challenges of setting up the eye tracker equipment and the extensive amounts of time involved in recording the eye movements for all 120 participants (n=40) were identified. The number of participants whose eye movements were recorded as they answered the questions was reduced to twelve, four in each treatment group.

The pilot study showed that the treatment group using shape and value in conjunction was more effective than others in the time taken to complete the task. Color and shape in conjunction increased the time on task. There was also a significant difference in time on task for the different question levels between groups. However, participants performed better (higher scores) when color and value were used as retinal variables as in treatment group 2. Details of the data analysis and the results are discussed in Appendix M. The data for each of the treatment methods were imported into MS Excel and the analysis was completed using SPSS. The performance scores (number correct) and the time taken for each of the different levels of questions were recorded. The usefulness ratings of the displays and the mental strategies used to complete the tasks were also noted. The data was collected from the three treatment groups and nine participants is shown in, appendix M.

Chapter IV

Results

The experimental study was conducted in spring 2009 with 135 participants from an undergraduate course in education technology at a major university in Southeastern United States. Scores from all participants were included in the data analysis. Out of the 135, 12 participants completed the task while being recorded by the eye tracker and 123 participants completed the task using a desktop computer without the eye tracker. The participants were randomly assigned to each treatment group. None of the participants indicated any prior experience in computer science, engineering or statistical analysis in the sign in sheet. All participants had basic computer skills and were proficient in using input devices such as the mouse and the keyboard. The results from the study addressed the research questions and the hypotheses listed in the methods section of the dissertation. This chapter includes the data analysis results and describes how each analysis method answered the different questions. The first part of the section discusses results from the interactive displays. The second part discusses results from analysis of scores from the eye tracker.

Dependability analyses indicated variations in the individual responses to the questions at levels 1 and 2. The inconsistencies in the responses may have been caused by several factors such as the instrument used, the motivation levels, and individual differences among participants. Recommendations to resolve these inconsistencies in future studies using such interactive displays are discussed later in chapter 5.

$R_{1:}$ Is there a significant difference in performance measured as number of correct responses based on the retinal variables (color, value, and shape) used in conjunction in the display?

$R_{2:}$ Is there a significant difference in performance measured as number of correct responses based on the complexity level (location, local relationships and global trends) of questions answered using the different retinal variables?

Research questions 1 and 2 evaluate the differences in performance scores in the three treatment groups. Performance scores were measured as the number of correct responses to questions in the three levels of compleixity, in the three different groups. In order to answer this question, the performance scores were compared across the three treatment groups. Table 5 shows mean performance scores by groups and the descriptive statistics of performance scores for each treatment groups are shown in Tables 6, 7 and 8

Table 4 Mean Performance Scores by Groups

Group	Level 1	Level 2	Level 3	Mean
Color/Shape	5.853	8.097	5.268	6.406
Color/Value	4.585	7.902	3.780	5.442
Value/Shape	5.561	6.926	2.414	4.967

N=123

Table 5 Descriptive Statistics for Performance Scores for Color/Shape Treatment Group

Question level	Mean	S.D	Min	Max
Level 1	5.853	2.697	0.0	8.0
Level 2	8.097	4.048	0.0	12.0
Level 3	5.268	2.480	1.0	12.0

N = 41

Table 6 Descriptive Statistics for Performance Scores for Color/Value Treatment Group

Question level	Mean	S.D	Min	Max
Level 1	4.585	3.041	0.0	8.0
Level 2	7.902	2.896	0.0	12.0
Level 3	3.780	1.650	1.0	7.0

N=41

Table 7 Descriptive Statistics for Performance Scores for Shape/Value Treatment Group

Question level	Mean	S.D	Min	Max
Level 1	5.561	3.082	0.0	8.0
Level 2	6.926	4.291	0.0	12.0
Level 3	2.414	1.499	0.0	7.0

N=41

Overall mean performance scores for all three levels of questions was the highest in group 1(color/shape). Group 3 (shape/value) had the lowest mean performance scores. The lowest mean score was in group 3 for the level 3 questions. For all treatment groups, differences in mean performance scores for question levels one and two are narrow while those for level 3 shows some differences. Overall mean scores are higher for level 2 questions in all the three groups. The distances from the mean, the standard deviation in group 2 for level 2 questions' performance scores (SD = 2.896) is less when compared to standard deviations for level 2 performance scores in group 1 and 3. (SD= 4.048, 4.291). Differences in mean scores between group 1 and group 3 for question level 3 are the highest. The range of scores is the lowest for level 3 questions, specifically in group 2 (Range= 6).

In addition to descriptive statistics, a repeated measures analysis with the three treatment groups as between subject factors and question levels as the within subjects variables was conducted. The repeated measures were done to identify significant main or interaction effects between the factors. The results of the repeated measures analysis are shown in Table 9

Table 8 Repeated Measures - Tests of Within Subjects and Between Subjects Effects

Source	df	Sum-of-Squares	Mean-Squares	*F*	*p*
Within Groups					
Scores	1.915	910.981	475.765	54.806	.000
Scores*Group	3.380	102.385	26.736	3.080	.018
Error	229.773	1994.634	8.681		
Between Groups					
Group	2	133.079	66.539	6.547	.002
Error	122	1352.64	11.087		

N=123

Tests of Within Subjects Effects from the repeated measures analysis showed a significant scores main effect (F= 54.806, P<.05) and scores by group interaction effects (F=3.080, P<.05).

Figure 25 shows a graphical representation of the distribution of the mean performance scores by groups by question levels. While for questions in levels 1 and 2, the means scores did not show vast differences, those in level three showed differences among the groups. Level two questions scores the most number of right answers. The between subjects factor treatment group had a significant effect on the performance scores in the different levels. (F=6.547, P<.05). Results from the Tukey post hoc analysis are shown in Table 10.

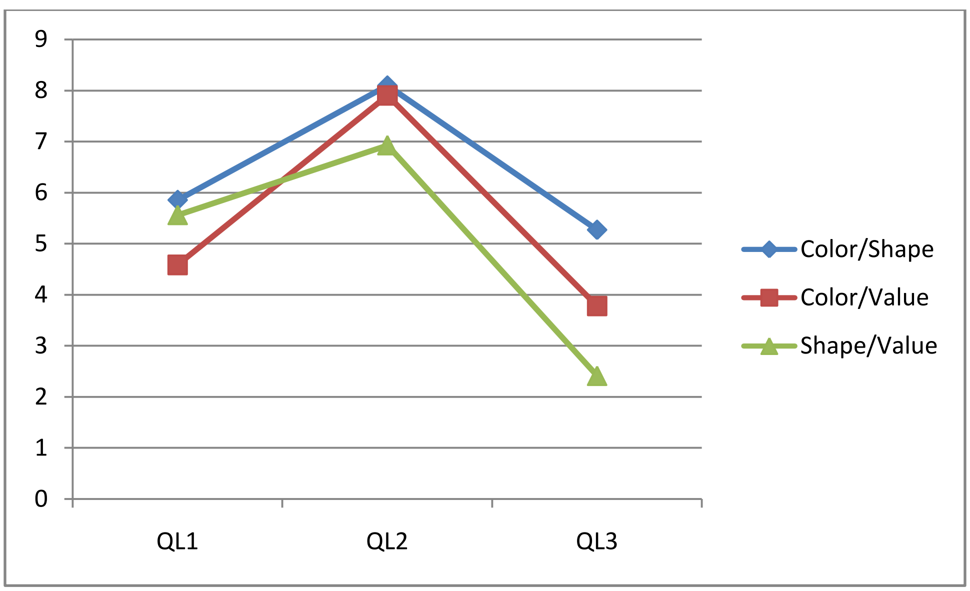


Figure 25 Display Showing Differences in Average Performance Scores by Group by Question Levels

Table 9 Tukey's Post hoc Analysis of Comparison of Means

Treatment Groups	Group	Mean Differences	Std. Errors	P
Color/Shape	Color/Value	0.9837	0.406	.045
	Shape/Value	1.4390	0.406	.002
Color/Value	Color/Shape	0.9837	0.406	.045
	Shape/Value	0.4553	0.406	.504
Shape/Value	Color/Shape	1.4390	0.406	.002
	Color/Value	-0.4553	0.406	.504

N = 123

Tukey's post hoc comparison of means shown in Table 10 indicated significant differences between color/shape and color/value (F= .045, P<.05), color/shape and shape/value (F= .002, P<.05). Based on the overall means by treatment groups, participants in treatment group 1 that consisted of color and shape as the retinal variables in conjunction performed significantly better than participants in treatment groups 2 and 3 that used color/value and shape/value as retinal variables in conjunction. Participants in group 1 overall as seen in Table 3 had the highest performance scores. There were no significant differences in performance scores between groups 2 and 3.

In order to identify significant differences at the question levels, that might have contributed to the overall groups differences, additional multiple comparisons of the groups means for the each of the question levels were analyzed and they showed results as seen in Table 11.

Table 10 Multiple Comparisons of Group Mean Performance Scores for the Different Question Levels Among Groups

Question Level	Group	Mean Differences	Std. Errors	P
Level 1				
Color/Shape	Color/Value	1.268	0.650	.129
	Shape/Value	0.292	0.650	.895
Color/Value	Color/Shape	1.268	0.650	.129
	Shape/Value	-0.975	0.650	.295
Shape/Value	Color/Shape	-0.292	0.650	.895
	Color/Value	0.975	0.650	.295
Level 2				
Color/Shape	Color/Value	0.195	0.838	.971
	Shape/Value	1.170	0.838	.346
Color/Value	Color/Shape	-.195	0.838	.971
	Shape/Value	0.975	0.838	.477
Shape/Value	Color/Shape	-1.170	0.838	.346
	Color/Value	-0.975	0.838	.477
Level 3				
Color/Shape	Color/Value	1.487	0.425	.002
	Shape/Value	2.853	0.425	.000
Color/Value	Color/Shape	-1.487	0.425	.002
	Shape/Value	1.365	0.425	.005
Shape/Value	Color/Shape	-2.853	0.425	.000
	Color/Value	-1.365	0.425	.005

N=123

From the Tukey comparison of means shown in table 11, level 3 questions contributed to the significant differences in the group performances. Participants in treatment group 1 that uses color and shape as the retinal variables in conjunction performed significantly better than those in the other two treatment groups when answering level 3 questions. Participants in treatment group 2, that uses color value in conjunction also performed significantly better than those in treatment group 3 that uses shape and value when answering level 3 questions. There were no significant differences in performance among groups for question levels 1 and 2. While all treatment groups (color/shape, color/value and color/shape) did not show significant differences in performance when participants identified data locations or compared trends to identify relationships, treatment group 1 that uses color and shape in conjunction showed significant difference in performance when participants answered level 3 questions that involved analysis global trends and making decisions.

R_3: Is there a significant difference in total time taken to answer questions based on the retinal variables (color, shape, value) used in conjunction in the displays?

R_4: Is there a significant difference in total time taken to answer questions based on the level of complexity (location, local relationships and global trends) of questions presented using the different retinal variables?

Research questions 3 and 4 evaluate the total time on task scores for each of the three treatment groups and the time on task scores for each level of questions answered by the participant. The time to answer each question was automatically recorded by the interactive displays developed by the researcher. From these values the total time on task for each of the question levels for the three treatment groups was calculated. Table 12 shows the total time on task scores for each of the question levels by group.

Table 11 Mean Time on Task Scores by Group by Levels

Time on Task by level	Color/Shape	Color/Value	Shape/Value	Mean Scores
Level 1	102.151	89.523	82.602	91.425
Level 2	70.619	59.642	65.363	65.208
Level 3	343.433	277.595	271.913	297.647

N=123

Table 12 Descriptive Statistics of Time on Task Scores for Color/Shape Treatment Group

Time on Task by level	Mean	SD	Min	Max
Level 1	102.151	35.316	40.50	185.79
Level 2	70.619	24.843	34.93	152.58
Level 3	343.433	190.279	150.79	1287.48

N=41

Table 13 Descriptive Statistics of Time on Task Scores for Color/Value Treatment Group

Time on Task by level	Mean	SD	Min	Max
Level 1	89.523	26.790	22.28	175.83
Level 2	59.642	21.756	17.77	133.13
Level 3	277.595	103.801	114.20	653.06

Table 14 Descriptive Statistics of Time on Task Scores for Shape/Value Treatment Group

Time on Task by level	Mean	SD	Min	Max
Level 1	82.602	34.350	47.95	217.60
Level 2	65.363	17.776	39.44	102.47
Level 3	271.913	79.907	102.03	530.50

N=41

Tables 13,14 and 15 show descriptive statistics for time on task scores. Mean time on task scores by treatment groups by levels revealed that participants took the longest time to answer level 3 questions in all groups. Participants in group 2 took the least amount of time than those in other treatment groups for all the three levels of questions. It must be noted however, that there are significant variations in the range of time on task scores in all the three levels of questions in all the three treatment groups. Treatment group 3 (shape/value) showed the lowest range of time on task scores for question level 2.(Range= 63.03) Treatment group 1(color/shape) showed the highest range of time on task scores for question level 3 (Range = 1136.69). Deviations from the mean were highest for question level 3 in all the three treatment groups (SD = 79.907).

Repeated measures analysis with treatment groups as between subject variables and time on task scores for the different levels of questions as within subject variables revealed effects as seen in Table 16.

Table 15 Repeated Measures Tests of Within Subjects Effects

Source	df	Sum-of-Squares	Mean-Squares	*F*	*p*
Within Groups					
TOT	1.095	3986968.049	3574299.638	342.973	.000
TOT*Group	2.231	86.897	27426.894	2.632	.071
Error	131.350	133.855	104221.530		
Between Groups					
Group	2	78927.635	66.539	6.547	.002
Error	120	921745.097	7681.209		

N=123

Tests of within subjects effects showed a significant time main effect (F=342.973, P<.05) and no significant time by group interaction effects. The between subjects variable treatment group showed significant main effects (F=5.138, P>.05) as seen in Table 16.

Figure 26 displays the differences in the mean time on task scores by group by question levels. From the display differences in mean time on task scores are highly varied for questions in level 3. Group 3 shows the longest time taken to complete the questions at all three levels. Tukey's post hoc analysis comparing the means showed results displayed in Table 17.

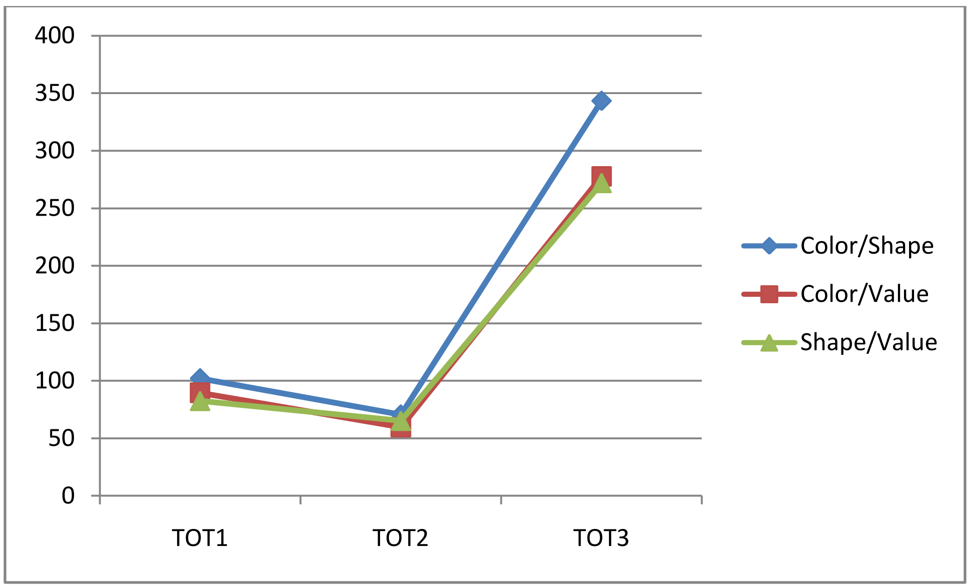


Figure 26 Display Showing Differences in Average Time on Task Scores by Group by Question Levels

Table 16 Tukey's Post Hoc Analysis for Comparison of Means

Treatment Groups	Group	Mean Differences	Std. Errors	P
Color/Shape	Color/Value	29.813	11.175	.023
	Shape/Value	32.108	11.175	.013
Color/Value	Color/Shape	-29.813	11.175	.023
	Shape/Value	2.294	11.175	.977
Shape/Value	Color/Shape	-32.108	11.175	.013
	Color/Value	-2.294	11.175	.977

N=123

Tukey's post hoc analysis showed significant differences in time on task scores between treatment groups 1(color/shape) and 2 (color/value) and treatment groups 1(color/shape) and 3 (shape/value). These significant differences between these groups contributed to the overall effects seen from the repeated measures analysis. There were no significant differences between treatment groups 2(color/value) and 3 (color/shape).

In order to identify which of the within subjects variables contributed to the significant differences, multiple comparison of mean performance scores at the different question levels by groups was studied. The results from the multiple comparisons are shown in Table 18.

Table 17 Tukey's Multiple Comparisons of Means

Time on Task(TOT)	Group	Mean Differences	Std. Errors	P
Time on Task Level 1				
Color/Shape	Color/Value	0.9837	0.406	.186
	Shape/Value	1.4390	0.406	.020
Color/Value	Color/Shape	0.9837	0.406	.186
	Shape/Value	1.4390	0.406	.598
Shape/Value	Color/Shape	0.9837	0.406	.020
	Color/Value	1.4390	0.406	.598
Time on Task Level 2				
Color/Shape	Color/Value	0.9837	0.406	.060
	Shape/Value	1.4390	0.406	.517
Color/Value	Color/Shape	0.9837	0.406	.060
	Shape/Value	1.4390	0.406	.458
Shape/Value	Color/Shape	0.9837	0.406	.517
	Color/Value	1.4390	0.406	.458
Time on Task Level 3				
Color/Shape	Color/Value	0.9837	0.406	.069
	Shape/Value	1.4390	0.406	.044
Color/Value	Color/Shape	0.9837	0.406	.069
	Shape/Value	1.4390	0.406	.980
Shape/Value	Color/Shape	0.9837	0.406	.044
	Color/Value	1.4390	0.406	.980

N=123

Multiple comparisons of mean time on task scores at the different question levels for each of the treatment groups showed significant differences between group 1 (color/shape) and group 3 (shape/value) for both question levels 1 ($F=7.151$, $P<.05$), and level 3 ($F= 29.457$, $P<.05$). Thus in terms of time taken to complete the task, treatment group 3 that used the retinal variables shape and value in conjunction performed significantly better than treatment group 1 that used retinal variables color and shape in conjunction for questions that required location of data points. For level 3 questions that involved study of global trends and decision making, treatment group 1 that used retinal variables color and shape in conjunction took significantly less time than treatment group 3 that used retinal variables shape and value in conjunction. Treatment group 2 that used color and value in conjunction did not show any significant differences with the treatment groups 1 and 3 for any of the question levels.

Thus in terms of performance scores, significantly better performances were seen in group 1 only for level 3 questions, and in terms of time on task scores, group 3 took significantly less time than group 1 for level 1 questions and group 1 took significantly less time than group 3 for level 3 questions.

Correlational Studies

Besides repeated measures analysis to explore significant differences, the direction and strength of relationships between several variables in the study were investigated. This section discusses results obtained for correlations examined. The correlations were calculated for both time on task, and performance scores by treatment groups. Table 19 shows the strength and relationship between overall mean group performance scores and time on task scores. Questions related to correlational analysis are reported in this section. Appendices I, J and K show scatter plots that display correlations between the mean group performance scores, time on task scores and usefulness ratings.

Question 1. What is the direction and strength of the relationship between performance scores and time on task scores based on the retinal variables used in conjunction in the displays?

Table 18 Strength and Direction of Relationships between Overall Group Performance Scores and Time on Task Scores

	Mean Total Score	Mean Total TOT
Mean Total Score	1.000	.943 .217
Mean Total Time on Task	.943 .217	1.000

Pearson Product Moment Correlation shows strong positive relationships between overall performance scores and time on task scores by group.(r= .943), which indicates that as the performance scores increase, the time on task increases. Participants took more time to answer questions that they scored higher in. However, the correlation was not statistically significant (P>.05)

Question 2. What is the direction and strength of the relationship between performance scores and time on task scores based on the level of complexity (location, local relationships and global trends) of questions presented using the different retinal variables?.

Tables 20, 21 and 22 show correlational results between performance scores and time on task scores for each of the three treatment groups.

Table 19 Strength and Relationships between Performance and Time on Task Scores by Question Levels in Group 1

	Qlevel1	QLevel2	Qlevel3	TOT1	TOT2	TOT3
Qlevel1	1.000	.056	.058	-.102	-.321	.029
		.727	.717	.525	.041	.860
QLevel2	.056	1.000	.087	.089	.026	.033
	.727		.589	.581	.874	.837
Qlevel3	.058	.087	1.000	.300	.410	.518
	.717	.589		.057	.008	.001
TOT1	-.102	.089	.300	1.000	.235	.179
	.525	.581	.057		.139	.263
TOT2	-.321	.026	.410	.235	1.000	.315
	.041	.874	.008	.139		.045
TOT3	.029	.033	.518	.179	.315	1.000
	.860	.837	.001	.263	.045	

N=41

Table 20 Strength and Relationships between Performance and Time on Task Scores by Question Levels in Group 2

	QLevel1	QLevel2	QLevel3	TOT1	TOT2	TOT3
QLevel1	1.000	-.039	-.093	.089	.035	.189
		.810	.562	.580	.830	.236
QLevel2	-.039	1.000	.205	.075	.258	.441
	.810		.200	.639	.104	.004
QLevel3	-.093	.205	1.000	.131	.229	.604
	.562	.200		.413	.150	.000
TOT1	.089	.075	.131	1.000	.536	.234
	.580	.639	.413		.000	.141
TOT2	.035	.258	.229	.536	1.000	.252
	.830	.104	.150	.000		.112
TOT3	.189	.441	.604	.234	.252	1.000
	.236	.004	.000	.141	.112	

N=41

Table 21 Strength and Relationships between Performance and Time on Task Scores by Question Levels in Group 3

	QLevel1	QLevel2	QLevel3	TOT1	TOT2	TOT3
QLevel1	1.000	.281	.008	.014	-.090	-.539
		.075	.961	.932	.575	.000
QLevel2	.281	1.000	-.038	-.058	-.285	.140
	.075		.814	.720	.071	.382
QLevel3	.008	-.038	1.000	.003	.233	.147
	.961	.814		.987	.143	.360
TOT1	.014	-.058	.003	1.000	.283	.242
	.932	.720	.987		.073	.127
TOT2	-.090	-.285	.233	.283	1.000	.242
	.575	.071	.143	.073		.128
TOT3	-.539	.140	.147	.242	.242	1.000
	.000	.382	.360	.127	.128	

N=41

Correlational analyses showed weak to medium strength relationships. Direction of the relationships were both positive and negative. Treatment group 1 showed significant negative weak correlation ($r = -.321$, $P<.05$) between performance scores at question level 1 and time on task for question level 2, and medium positive correlations between question level 3 and time on task scores for question level 2 ($r= .410$, $P<.05$), and question level 3 and time on task scores for question level 3 ($r= .518$, $P<.05$).

Treatment group 2 showed medium positive correlations between time on task scores for question levels 1 and 2 ($r= .536$, $P<.001$), between performance scores at question level 2 and time on task scores for question level 3 ($r=.441$, $P<.05$), and between performance scores at question level 3 and time on task scores for question level 3 ($r=.604$, $p<.001$).

Treatment group 3 showed a significant medium negative correlation between performance score at question level 1 and time on task score at question level 3(r= -.539, P<.001). None of the other correlations were significant.

Question 3. What is the direction and strength of the relationship between usefulness rating and the performance scores measured as the number of correct responses, based on the different retinal variables (color, shape, value) used in conjunction in the displays? The mean usefulness rating by treatment group is shown in Table 23. The correlations between usefulness rating, performance scores and time on task scores by group are shown in Tables 24.

Table 22 Mean Usefulness Ratings by Treatment Group

Treatment Group	Mean Rating
Color/Shape	3.317
Color/Value	3.220
Shape/Value	3.068

N=123

Table 23 Strength and Relationship between usefulness rating and performance scores by group

Group		Level 1 scores	Level 2 Scores	Level 3 Scores
Color/Shape	Rating	.118 .462	-.141 .380	.351 .024
Color/Value	Rating	.350 .025	.123 .443	-.178 .265
Shape/Value	Rating	.226 .156	.273 .084	.017 .917

N=41

As seen in Table 22, the usefulness rating in treatment group 2 (color/value) showed a weak significant positive relationship with the question level 1 performance scores (P<.05). It is worth noting that mean performance scores were the highest for level

1 questions in treatment group 2. Usefulness ratings in treatment group 1 (color/shape) showed a significant weak positive relationship with scores in level 3. Group 1 participants had the highest performance scores for level 3 questions. None of the other relationships were significant. Almost all relationships were weak positive relationships except for usefulness rating in treatment group 1(color/shape) and question level 2 scores, and usefulness rating in treatment group 2 (color/value) and question level 3 scores, which showed weak negative relationships.

Question 4.What is the direction and strength of the relationship between usefulness rating and the time taken to answer questions at different levels of complexity (location, local relationships and global trends) of questions presented using the different retinal variables?

The correlation between participant's time on task scores to answer questions at the three levels of complexity and the usefulness rating are shown in Table 25.

Table 24 Strength and Relationship between Usefulness Rating and Time on Task Scores by Group

Group		TOT1	TOT2	TOT3
Color/Shape	Rating	.146	.047	.233
		.363	.772	.142
Color/Value	Rating	-.008	.046	.069
		.962	.776	.670
Shape/Value	Rating	-.401	-.391	-.171
		.009	.011	.285

Correlations between usefulness ratings and time on task scores for each question levels in all the three treatment groups showed weak relationships as seen in Table 25. The direction of the relationships were varied. Group 3(shape/value) showed significant weak correlations (P<.05) for time on task scores for level 1 and 2 questions, and a weak negative correlation that was not significant for question level 3. Time on task scores were the highest for group 3 Treatment groups 1 (color/shape) and group 2 (color/value) showed weak, positive correlations which were not significant. Thus the relationship between time taken by each participant to complete answering 9 questions in the three different complexity levels and usefulness rating of the displays was weak .

Eye Tracker Data Analysis

In addition to the analysis of scores from the interactive displays used in the study, data was also collected from the ViewPoint eye tracker system and analyzed. 12 out of the 135 participatns completed the task of answering the 9 questions while theie eye movements were recorded. There were thus four participants in each treatment group. The study included several research questions related to the data obtained from the eye movements of the participants. This section of the chapter, discusses the results from the analysis of the eye tracker data. The performance scores and the time on task scores by group and by question levels are included. Saccade and fixation scores from the eye tracker measured as time in seconds, the fixation scores at the different regions of interests defined by the researcher, and the glance frequencies were calculated and used to answer the research questions. Performance scores and time on task scores by question levels by group are reported in Tables 26 and 27. Performance scores and time on task scores by participant by group is shown in Tables 28 and 29.

Table 25 Mean Performance Scores by Question Level by Group

Group	Score Level 1	Score Level 2	Score Level 3	Mean Total
Color/Shape	5.0	5.8	2.5	4.43
Color/Value	7.0	7.4	3.38	5.57
Shape/Value	4.77	6.74	3.5	5.92

N=12

Table 26 Mean Time on Task Scores by Question Level by Group

Group	TOT1	TOT2	TOT3	Mean Total
Color/Shape	81.92	69.46	221.67	124.35
Color/Value	84.78	68.43	350.10	167.77
Shape/Value	81.18	69.67	324.41	158.62

N=12

Table 27 Descriptive Statistics for Color Shape Treatment Group

	Mean	S.D	Min	Max
TOT	124.354	21.968	101.25	153.66
Saccade	0.826	0.977	0.14	2.24
Fixation	0.737	0.483	0.44	1.45

Table 28 Descriptive Statistics for Color Value Treatment Group

	Mean	S.D	Min	Max
TOT	222.048	160.592	108.36	450.77
Saccade	0.599	0.255	0.27	.87
Fixation	0.788	0.788	0.32	1.28

Table 29 Descriptive Statistics for Shape Value Treatment Group

	Mean	S.D	Min	Max
TOT	135.038	24.033	101.28	157.91
Saccade	0.842	0.637	0.29	1.76
Fixation	0.873	0.071	0.79	0.96

Analysis of descriptive statistics for each of the treatment groups indicated that treatment group 3 that used shape and value in conjunction showed the highest mean saccade and fixation scores. It must also be noted that among the 12 participants the 4 participants in treatment group 3 showed the highest mean performance scores. Treatment groups 2 and 3 showed a difference of 26 seconds in their time on task scores.

Q_4. What is the direction and strength of the relationship between eye gaze time, saccade scores and total time taken to complete the task based on the different retinal variables (color, shape, value) used in conjunction in the displays?

The eye tracker data was collected from 12 participants who completed the task while their eye movements were recorded by the system. All data that are reported reflects scores and analyses related to the 12 participants. Saccade scores were measured as time in seconds that reflects the movements of the eye that locate interesting areas in the scene and build a mental map of what is being looked at, before the eye fixates on a region of interest. Saccades are the fastest movements of the eye and are recorded as the participants eye move towards the stimulus or the visual onset. Gaze scores or fixation scores are also measured as time in seconds and reflect the amount of time a participant fixated at a certain region in the display. Overall mean saccade scores by group and fixation scores by group are shown in Tables 31.

Table 30 Mean Saccade Scores and Fixation Scores by Group

Group	Saccade Scores	Fixation Scores
Color/Shape	0.8332	0.7374
Color/Value	0.7027	0.9143
Shape/Value	0.6882	1.164

In group 1(color/shape) participants' eye moved around and studied the display more than the other two groups before they fixated on the areas of interests on the displays. The highest fixation scores were found in group 3. It must be noted that participants in group 3 took the longest time to complete answering the questions. However group 1 (color/shape) participants had the highest performance scores. It appears that they probably studied the graph more before they answered the questions. The average saccade and gaze fixation scores by participant by treatment group are reported in Table 31.

Participant 1 in group 3 (shape/value), and participants 4 in group 1 (color/shape) and 2 (color/value) show the highest saccade scores. Lowest saccade scores were recorded in participant 2 in group 1 (color/shape). The time on task scores and performance scores for these specified participants when observed were very highly variable to report any plausible inferences. Individual differences in eye movements could be the result of strategies adopted by the participants to answer the questions and the interpretations of the displays and the questions presented.

Table 31 Strength and Direction of Relationship between Overall Mean Time on Task Scores, Fixation and Saccade scores

	Time on Task	Saccade	Fixation
Time on Task	1.000	.351 .772	-.822 .385
Saccade	.351 .712	1.000	.244 .843
Fixation	-.822 .385	. 244 .843	1.000

Table 32 Strength and Direction of Relationship between Time on Task Scores, Fixation and Saccade Scores

	TOT	Saccade	Fixation
TOT	1.000	-.186	.248
		.563	.438
Saccade	-.186	1.000	.527
	.563		.079
Fixation	.248	.527	1.000
	.438	.079	

N=12

In order to explore the relationships between the fixation score, saccade scores and the total time on task overall and for each group, Pearson Product Moment Correlations were calculated. Table 32 shows the relationships between the overall mean time on task scores, fixation and saccade scores. While time on task and saccade movement scores showed a negative medium relationship (r=-.485), and time on task scores and fixation scores a positive strong relationship (r=.861), neither of them were significant. As the fixation scores increased, it led to an increase in the total time on task, as fixation represents the duration of time a participant looked at a point of interest.

Table 33 represents the correlation results by treatment group. All correlations were weak and positive except the relationship between total time on task and saccade scores for group 2 (color/value) which showed a negative strong correlation (r=-.825). However none of the relationships were significant. Significant group differences in saccade and fixation scores were also analyzed using the analysis of variance, a 3 (treatment group)x two(eye scores) ANOVA. The results of the ANOVA are shown in Table 34. Analysis of Variance results did not report any F values or significance.

Table 33 3 (treatment groups) X 2 (eye scores) ANOVA Results

	Source	df	Sum-of-Squares	Mean-Squares	*F*	*p*
Between Groups	TOT	2	22972.056	11486.028	1.283	.323
	Saccade	2	.148	.074	.155	.858
	Fixation	2	103552.50	0.019	.143	.869
Within Groups	TOT	9	80550.44	8950.050		
	Saccade	9	4.279	.0475		
	Fixation	9	1.192	0.132		

N=12

In addition to exploring statistically significant differences and correlations, the study also explores the sequence of the eye movements and the fixations scores that relate to each of the regions of interest defined by the researcher in each of the displays.

Thus the study also investigated the eye movements of the participants using fifteen predefined regions of interests. The sequence in which the participants looked at the predefined regions and the duration in seconds they fixated in the regions was looked at. All participants performed the same tasks. The three displays used in the three treatment groups used the same fifteen predefined regions of interests. Hence differences in the eye scan paths and fixation regions could be an indicator of possible strategies and approaches taken by the participant. Moreover the researcher tried to analyze if there were any patterns that would indicate relationships between the eye movements, fixation duration and performance. Data and analysis completed are reported below each question they relate to.

Q_5. Is there a difference in the sequence in which participants look at the regions of interests defined in the displays in the three treatment groups as they perform their tasks?

The View Point eye tracker system allows for 0-99 regions or areas of interests to be defined to track the gaze fixation of each participant in these regions. Several research studies have used regions of interests defined in areas where there is defined interests or solutions to a given task. Observations of patterns in the movement towards the regions of interests and the duration of fixation has provided insight into possible strategies and behaviors towards problem solving, decision making and user interactions. Twelve Regions of Interests (ROI) were defined in this research study. These regions were based on the locations of the data points that the participants should have to click on as answers to the questions presented to them in the display. Table 35 shows regions of interests that needed to be looked at for each of the nine questions presented in the display.

Table 34 ROI and the Location of Data Points that needed to be looked at to Answer the Questions

Question Number	ROI with the answer
1, 2	2
3	9
4	4,5,12
5	3,5,11,12
6	6,7,8,9
7	7,8,9,10,11
8	4,5,10,11,12
9	1,2,3,11,12

The number of regions of interests that participants needed to look at varied based on the complexity level of the questions. Level 1 questions (questions 1,2 3) required them to identify a single data point location, while level 2 questions (questions 4,5,6) required comparisons to be made between data in different locations mentioned in the questions and level 3 (questions 7,8, and 9) required the participants to scan the entire display and identify interactions and relationship to make global comparisons. Table 36 shows the regions in which the regions of interests were fixated upon.

Table 35 Sequence in which Regions of Interests were Fixated Upon

Group	Participant	Regions of Interests Fixation in Sequence
Color/Shape	1	12,11,10,4,11,10,4,3,7,4,3,4,12,10,3,4,10,3,2,4
	2	10,4,10,4,10,12,3,4,10,12,3,3,4,10,12,10,3,4,10,12
	3	6,4,2,4,6,2,3,4,7,3,7,3,4,2,6,2,6,4,6,3,6
	4	11,8,9,11,9,11,9,12,4,14,15,4,0,5,4
Color/Value	1	12, 11,12,11,12,5,8,9,8,5,8,12,8,5,8,12,10,11,5,9,8
	2	9,10,9,10,4,10,10,4,10,9,10,7,4,10,9
	3	6,4,3,6,4,3,6,4,6,3,6,4,7,6,8,6,10,11,3,4
	4	0,9,10,9,10,4,10,9,3,9,10,4
Shape/Value	1	2,7,0,9,4,3,6,7,2,9,7
	2	6,7,6,7,4,0,12,7,4,10,7,8,6,8,9,0,7,10,6,10,7,6
	3	9,4,7,6,4,6,3,4,2,6,7,6,2,7,10,0,9,3,2,4,12
	4	4,0,4,9,10,9,10,11,9,10,9,10,4

Of the twelve regions of interests defined, all regions were looked at by participants at least once. The duration of fixation however varied among all of them. In group 1 (color/shape), regions of interests that were most frequently looked at include 3, 4, 6, 10 and 12. In group 2 (color/value), regions of interests included 3, 4, 6, 8, 9 and 10. Group 3 (shape/value) included regions 4, 6, 7, 9, and 10 as most frequently visited regions. All .groups included regions 4, 6, 9, and 10, which are regions that included the answers to the majority of questions (question 3, 4, 6, 7and 8) as seen in Table 35. The eye scan movements predominantly started from the left of the display for all participants. It must be noted that the legend was on the top left corner of the display which could have caused the participants to start from the left. However, not all participants looked at the region of interest 0 that was defined for the legend. The eye scan movements based on the regions of interests are displayed in Figures 24-35 by participant by group. Participant 1 in group 2 had the highest performance scores and the eye movements recorded for this participant indicates a thorough scan of the different regions of interests in the display. The eye movements shown in Figures 24-35 are based on Table 36 that shows the sequence in which the regions of interest were fixated upon. The start and end points are indicated by blue arrows. Participant 1 in the color/shape treatment group looked at region 12 first and region 4 last. A blue arrow in Figure 27 starting at region 12 and heading the direction of region 11 indicates the beginning of the sequence and a blue arrow from region 2 to region 4 indicates the end. The arrows in red indicate the different regions the participants fixated upon, and the direction of movement is represented by the direction of the arrow head. Thus participant 2 in the color/shape treatment groups as seen in Figure 28 fixated upon region 10 first and region 12 last, participant 3 as seen in Figure 29 fixated upon region 6 first and 6 again when they completed their task, and participant 3 as seen in Figure 30 fixated upon region 11 first and region 4 last. Eye movements for participants in treatment group 2 using color and value are shown similar to treatment group 1 in figures 31-34, and eye movements for participants in treatment group 3 using shape and value as retinal variables are shown in figures 35-38.

Participant 4 in treatment group 1 and participant 4 in group 3 show high performance in those groups. Examination of the eye movements of all these participants indicate that participant 1 in group 2 and participant 4 in group 1 showed similar scanning patterns. The two participants scanned the display and fixated on most regions of interests defined. All 12 participants showed patterns of search, re-search, re-identification and exploration of features in the display. Participants also visited regions where features were in groups rather than those where features were isolated. Close proximity contributes to the perception that those features may be related. While all participants searched and discriminated between the features, there is no clear indication of any patterns in the sequence of eye movements among them. Based on the eye paths and the regions visited, it appears that most participants used the variable year across the x axis as the foundation upon which their searches were built. In other words, a variable not represented by the features in conjunction was a key factor integral to the search.

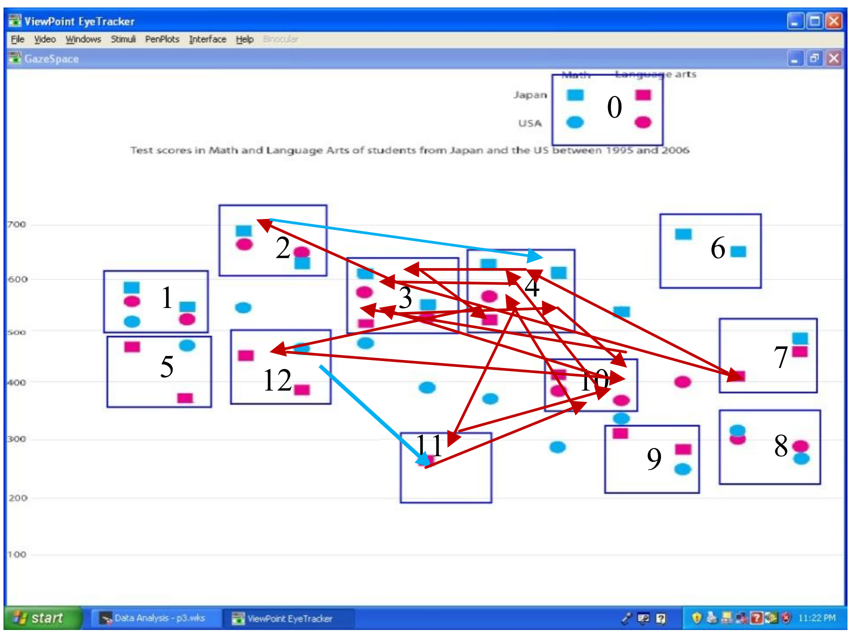


Figure 27 ROI and Eye Movements for Participant 1 in Color Shape Treatment Group

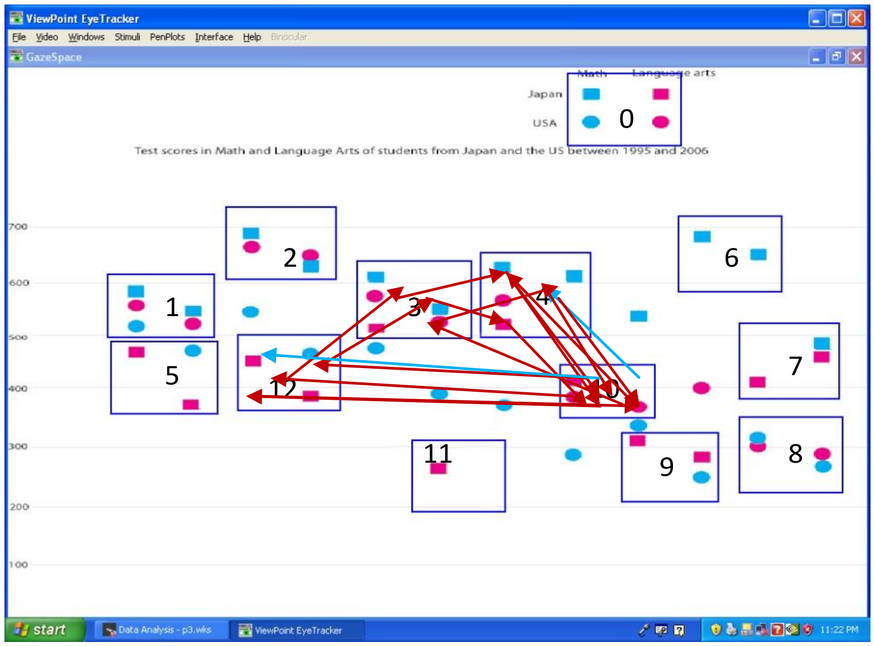


Figure 28 ROI and Eye Movements for Participant 2 in Color Shape Treatment Group

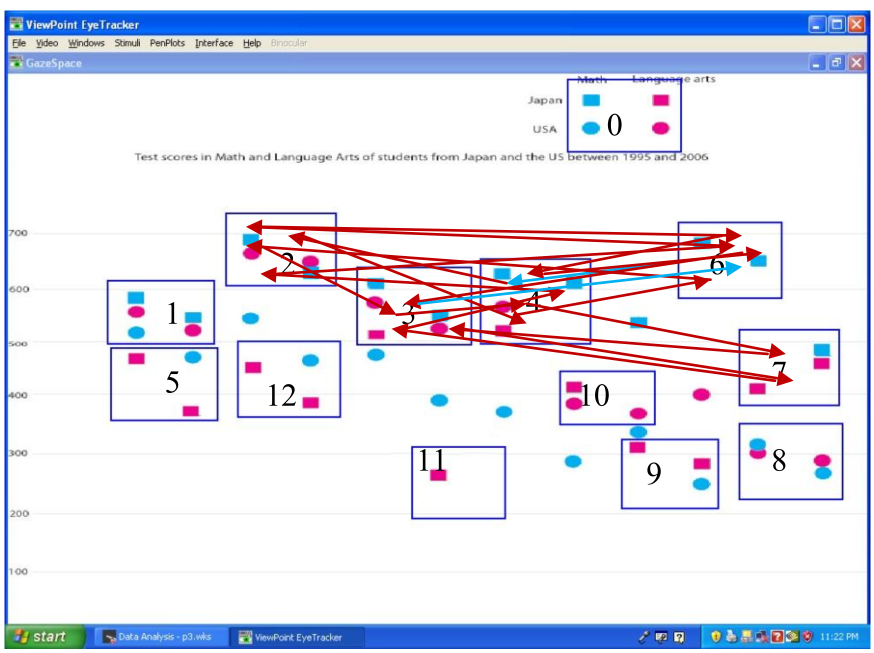


Figure 29 ROI and Eye Movements for Participant 3 in Color Shape Treatment Group

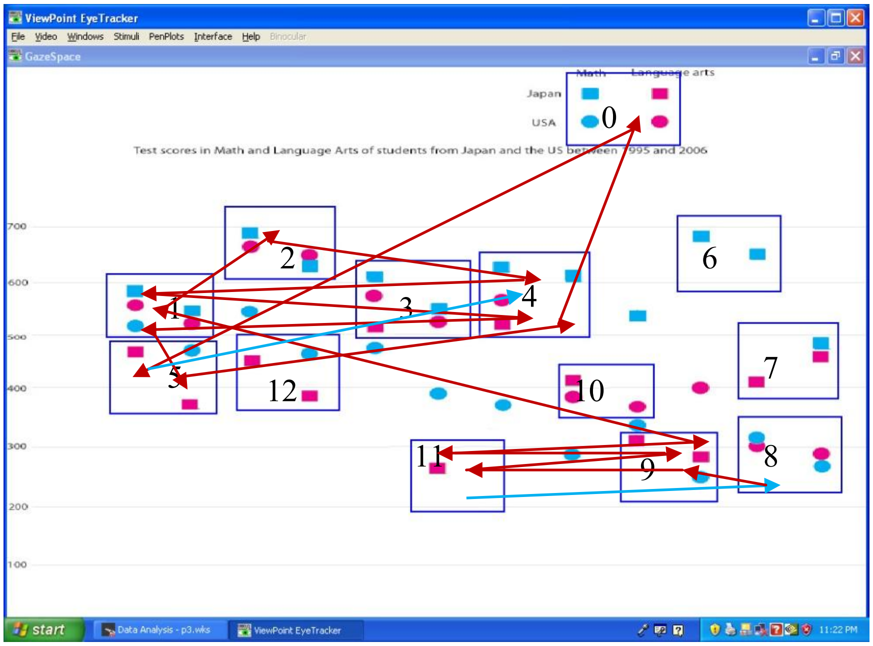


Figure 30 ROI and Eye Movements for Participant 4 in Color Shape Treatment Group

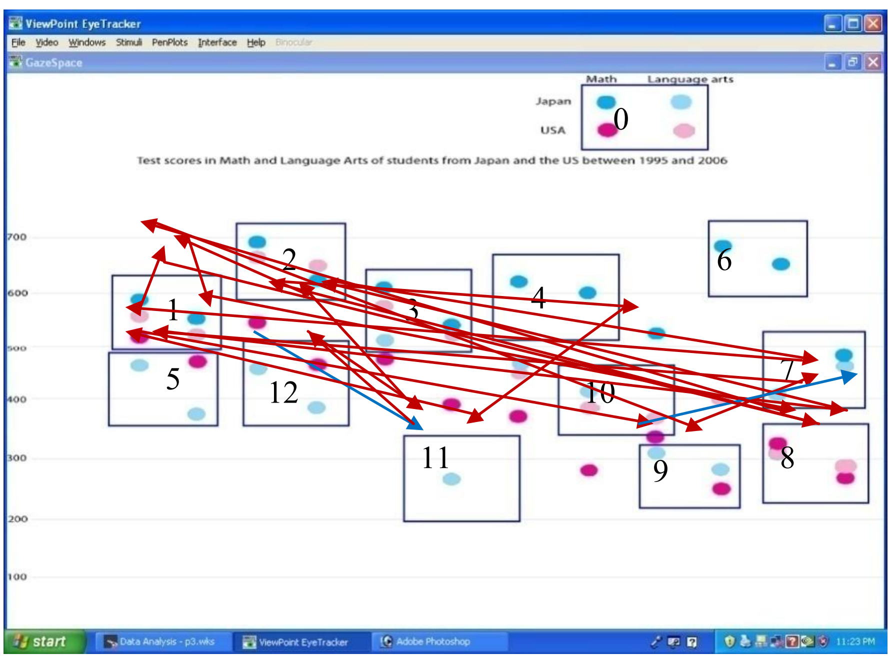


Figure 31 ROI and Eye Movements for Participant 1 in Color Value Treatment Group

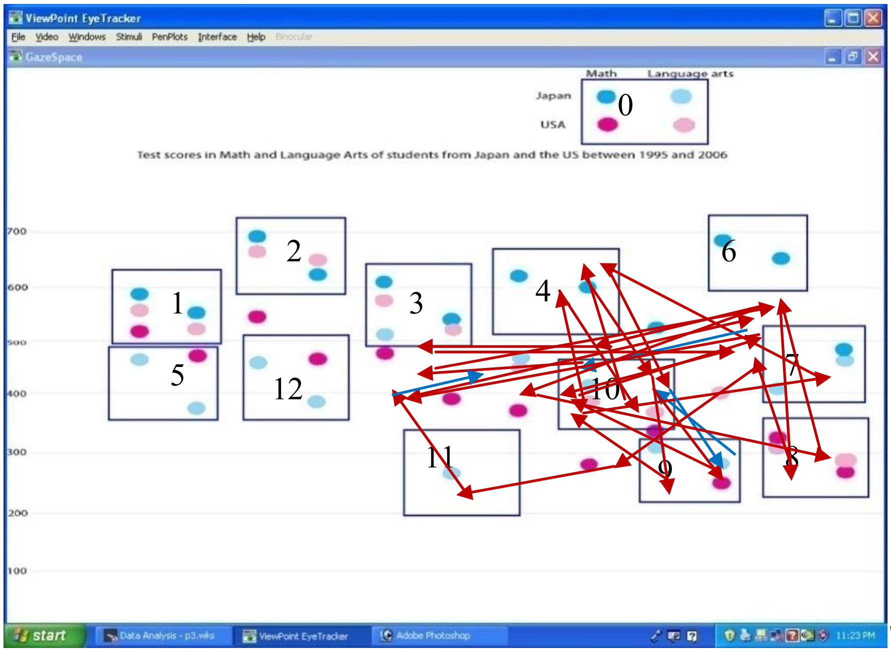


Figure 32 ROI and Eye Movements for Participant 2 in Color Value Treatment Group

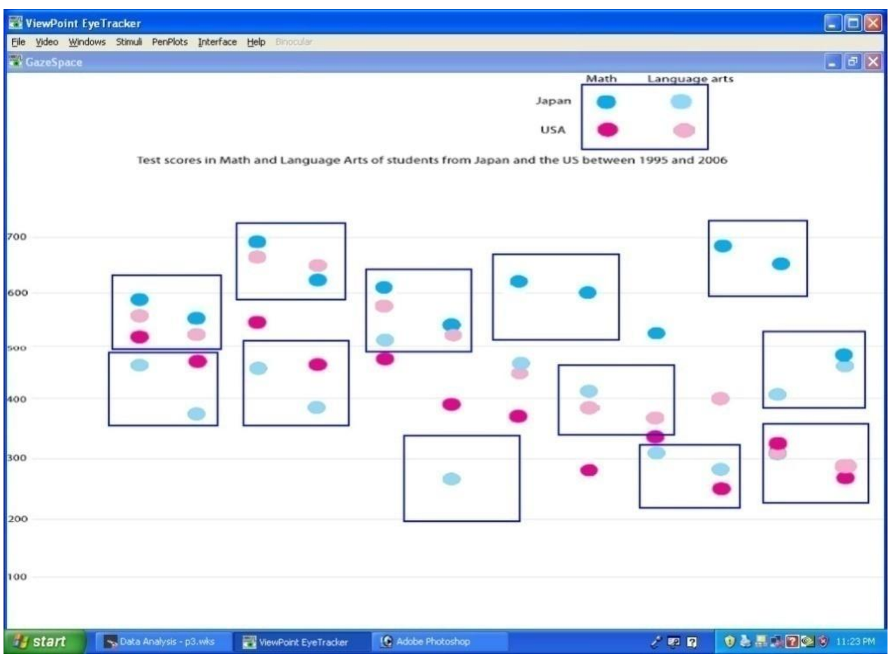


Figure 33 ROI and Eye Movements for Participant 3 in Color Value Treatment Group

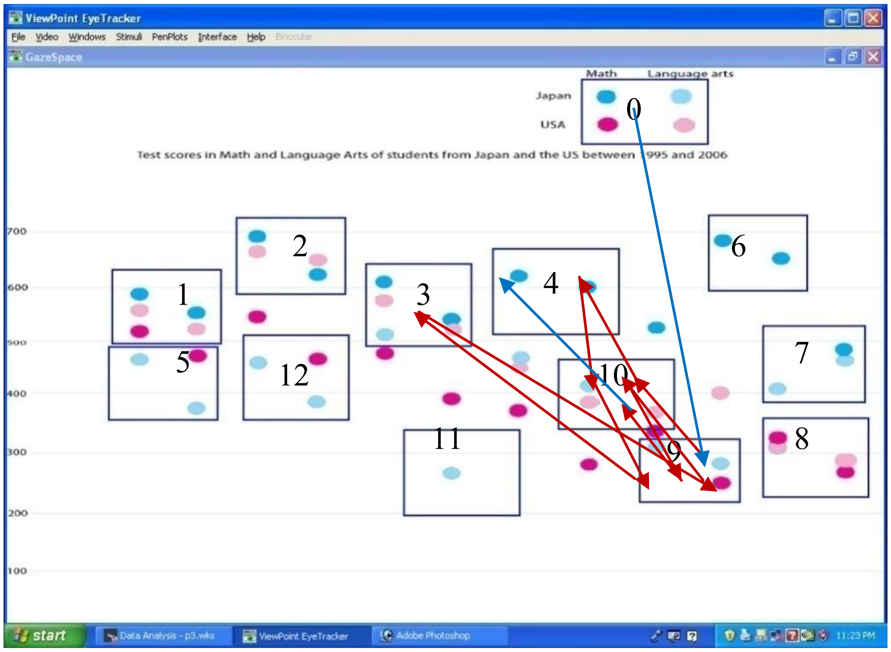


Figure 34 ROI and Eye Movement for Participant 4 in Color Value Treatment Group

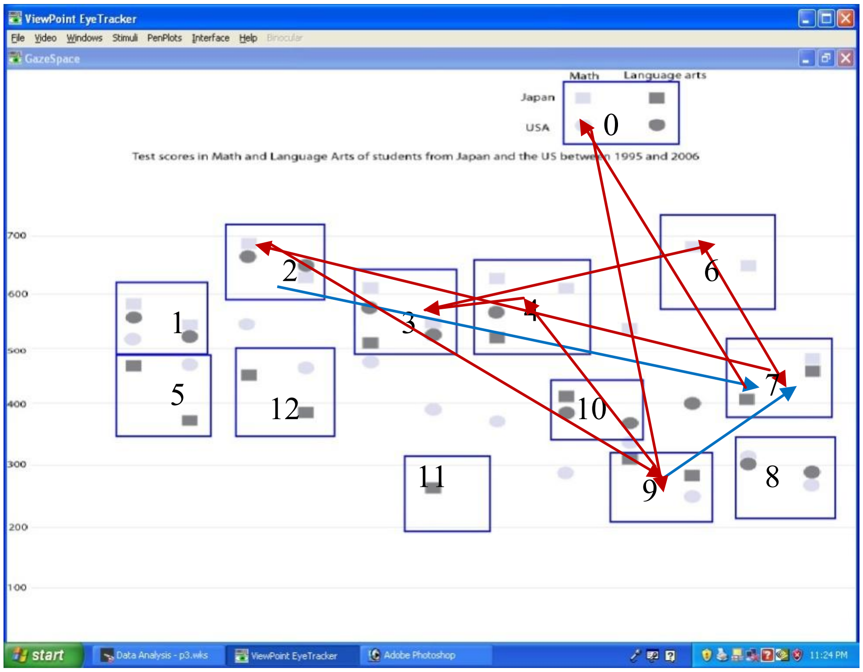


Figure 35 ROI and Eye Movements for Participant 1 in Shape Value Treatment Group

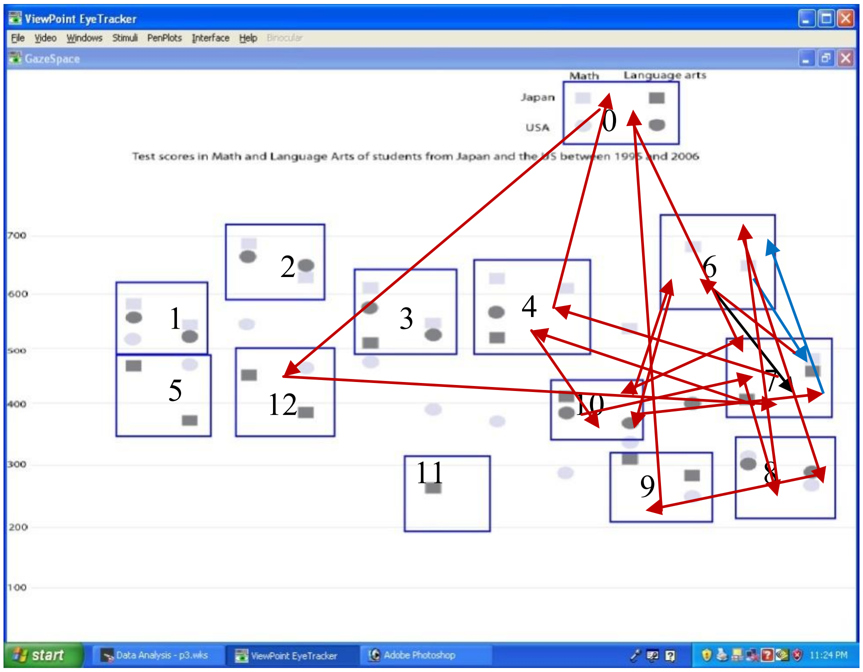


Figure 36 ROI and Eye Movements for Participant 2 in Shape Value Treatment Group

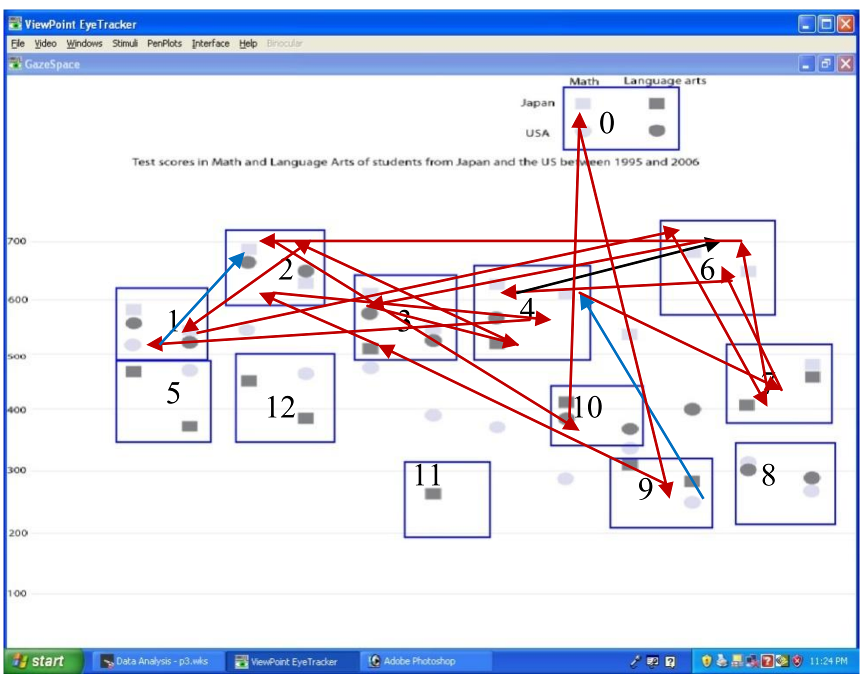


Figure 37 ROI and Eye Movements for Participant 3 in Shape Value Treatment Group

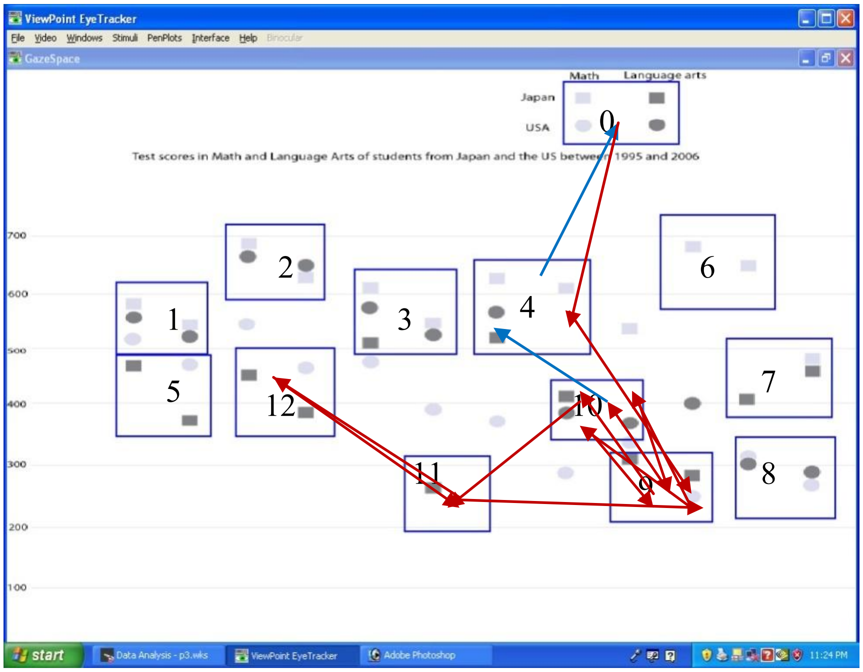


Figure 38 ROI and Eye Movements for Participant 4 in Shape Value Treatment Group

The eye tracker recorded the duration that each participant fixated on the different regions of interests consecutively over time. These are measured as glance frequencies, and provide insights into where the participants looked at and for how long. Table 37 shows the different regions of interests and the glance frequencies for each participant in group 1. Group 2 and 3 glance frequencies are shown in Table 38 and 39

Table 36 Glance Frequencies by Participant in Color Shape Treatment Group

ROI	P1 Frequency	P2 Frequency	P3 Frequency	P4 Frequency
0	0.1164		0.0204	0.0163
1				
2	13.974	0.1036	8.689	
3	1614.52	2187.92	152.059	0.2525
4	0.152	6503.84	201.908	0.4192
5			0.0169	
6	1.582		1817.25	0.1982
7	0.0997		5.52	0.1759
8			0.1995	
9	0.8761		0.3562	23.023
10	9.92	40.09	0.1675	0.09
11	2.749			0.0854
12	26.375	2.863	0.3661	0.0509

Table 37 Glance Frequencies by Participant in Color Value Treatment Group

ROI	P1 Frequency	P2 Frequency	P3 Frequency	P4 Frequency
0				3.5489
1	44.04			
2	5.376		0.7606	
3	20.92	0.3741	3014.06	173.86
4	0.1836	183.83	35.62	
5	457.21	8.5038	0.1332	
6		10.698	8992.51	23.038
7		0.4197	1308.32	
8	99.55	2.744	4477.21	
9	1.539	2868.7	0.0873	499.06
10	3.615	10594.56	2.294	19362.94
11	75.686	0.0673	0.0495	0.1843
12	2094.44	0.675	0.1474	

Table 38 Glance Frequencies by Participant in Shape Value Treatment Group

ROI	P1 Frequency	P2 Frequency	P3 Frequency	P4 Frequency
0		0.6354	0.1534	0.0166
1		0.0168	3.8411	
2	0.0329		3.7218	
3	0.018	0.925	3.03	
4	0.6034	86.383	21.831	0.1028
5		0.0167		
6	52.787	768.59	9101.85	
7	494.916	127.201	407.89	
8				
9	0.9069	47.83	1.0773	762.381
10	0.0314	47.36	0.0832	3119.21
11	0.1687	0.0842	0.1665	1.0478
12			0.0168	

Glance frequencies are defined as the duration in seconds that a participant fixated upon a region consecutively. These frequencies would then indicate the duration of attention to a specific region. The aim of the study was to find out if there were any patterns that participants displayed in the areas they looked at and in the duration of the fixation. Overall, while there were some common areas that all participants looked at, there were great variations in the patterns and the gaze frequencies between groups and within each group. This indicates that each participant adopted their own strategies based on what they understood the task and the data interactions to be.

On the other hand, there may exist, a pattern in the glance frequencies that is related to the performance. When each participant's performance scores and time on task scores along with the glance frequencies were explored, the performance scores varied and possibly resulted based on the regions that were looked at. Participant 1 in group 1 had the highest glance frequencies in regions 3 followed by 12, 2, 10 and 1. They looked at other regions but for considerably less amounts of time. Performance scores for level 1 questions were low. They answered only one question right. Questions in level 1 required participants to look at regions 1, 2 and 9. Again for level 2 questions which required them to look at regions 3, 4, 11, they obtained only one question right. Their glance frequencies were the highest for region 3. For level 3 questions which required them to look at all the regions and study the interactions to make decisions, they obtained a low rating score of 2. Such patterns were observed in all participants. Table 40 shows the highest glance frequency regions and the corresponding performance scores for each participant in group 1. Table 41 and 42 show the highest glance frequency regions and the corresponding performance scores for each participant in group 2 and 3.

Table 39 Highest Glance Frequency Regions and the Corresponding Performance Scores in Color Shape Treatment Group

Participant	ROI(highest Frequency regions first)	# correct Level1	#correct Level2	L3 Ratings
1	3,12	1	1	2
2	4,3	1	2	2
3	6,4,3	1	2	2
4	9	2	2	4

Table 40 Highest Glance Frequency Regions and the Corresponding Performance Scores in Color Value Treatment Group

Participant	ROI(highest Frequency regions first)	# correct L1	#correct L2	L3 Rating
1	12,5,8,11	2	2	11
2	10,9,4	1	2	2
3	6,8,3	2	1	1
4	10,9,3	2	3	4

Table 41 Highest Glance Frequency Regions and the Corresponding Performance Scores in Shape Value Treatment Group

Participant	ROI(highest frequency regions first)	# correct L1	#correct L2	L3 Rating
1	7,6	0	1	1
2	6,7,4,9,10	1	0	2
3	6,7,4	0	1	3
4	10,9	2	3	12

As they began their tasks, participants' eye movements show that they immediately fixated on the location of the data points following bottoms up approach to problem solving. Besides the glance frequencies listed in Tables 35, 36 and 37, participants gaze also includes other regions with less glance frequencies. Areas that recorded high glance frequencies indicate that these data locations were points that participants came back to when studying interactions and relationships. Besides these regions, other regions fixated upon helped them find answers to the questions asked.

Time on task scores varied with the number of regions that were looked at. It did not indicate any patterns that could be established. Participants who took more time to complete each of the question levels and overall task did not show higher performance scores.

Mental Strategies Reported

Appendix I shows the different strategies reported by the participants as they completed the task for the research study. Strategies varied among participants between the three treatment groups and within each group. The strategies described were coded and grouped based on the following criteria, the patterns or approaches followed by participants as they completed the task, any biases or perception describing the images, the questions, the data displayed and the displays themselves. There were a total of 98 mental strategies recorded. Ten of them were from the participants whose eye movements were recorded as they completed the task. Table 43 shows the different codes and the percentages of participants under each code.

Table 42 Coding of Mental Strategies Reported

Category or Code	Description	Number of participant's responses	Percentages in each group
Patterns or strategies used	Distinguished between colors first and understand what they are.	72	73%
	Read and re-read the questions	58	59%
Biases or perceptions	Used the knowledge of displays to solve the problem	11	11.2%
Challenges	Display confusing, overwhelming, difficult to read	8	8.1%

Some of them used the legend frequently to identify the different data displayed, but some of them went directly to the data location or to the highest scores first. 73% of participants indicated that they had to first distinguish between the features used in the display (color, shape, value) and constantly keep in mind what the colors and shapes meant and refer back to the legend if necessary to answer the questions presented. 59% of those who indicated having to distinguish between the features also indicated that they had to read and re-read the questions several times and study the data points that answered the questions. 11.2% of the participants reported that they did not use any specific strategies but mainly followed their knowledge of graphs and what they knew about displays. .8.1% found the color and shape display to be very overwhelming and confusing.

In summary, while each participant might have brought along with them their perceptions and bias, their strategies for problem solving were driven by the task on hand, and the complexity of the question presented. Many of them however, followed a pattern once they identified the data locations. They fixated on certain areas where they found the data locations, moved back and forth to the area after visiting other regions and studied the display for sometime before they made a decision.

Summary of Findings

The purpose of the research was to explore the effectiveness of retinal variables (color, value and shape) in facilitating interpretation of information when used in conjunction for the display of multidimensional data and their relationships. This section provides a summary of the results of the analysis for each of the research questions.

$R_{1:}$ Is there a significant difference in performance measured as number of correct responses based on the retinal variables (color, value, and shape) used in conjunction in the display?

$R_{2:}$ Is there a significant difference in performance measured as number of correct responses based on the complexity level (location, local relationships and global trends) of questions answered using the different retinal variables?

Overall mean performance scores by treatment groups and mean performance scores by group by question levels were compared. Overall mean performance scores were highest in group 1(color/shape) and lowest in group 3(shape/value). A majority of the participants had high scores for level 2 questions in all groups. Data from a repeated measures ANOVA using performance scores in the three question levels as within subjects variable and treatment group as between subjects' variable indicated a significant score main effect and significant score by group interaction effects. A Tukey post hoc analysis showed significant differences in performance scores between treatment groups 1 and 2 and 1 and 3. However all of these differences are contributed by level 3 questions. Treatment group 1 performed significantly better than treatment groups 2 and 3 for question level 3. Treatment group 2 performed better than treatment group 3 for question level 3. Questions levels 1 and 2 did not contribute to the effect. Thus performance is affected by the treatment group used and the level of complexity of the question.

R_3: Is there a significant difference in total time taken to answer questions based on the retinal variables (color, shape, value) used in conjunction in the displays?

R_4: Is there a significant difference in total time taken to answer questions based on the level of complexity (location, local relationships and global trends) of questions presented using the different retinal variables?

Mean time on task scores indicated that group 1 had the highest mean time on task scores and group 3 had the lowest scores. Group 3 participants showed low time on task scores for all three levels of questions. Level 1 question for all treatment groups showed the highest time on task scores followed by levels 2 and 3. This could indicate that as the participants got used to the displays, they performed better. Repeated measures ANOVA with time on task scores as within subjects' variables indicated a significant time on task main effect. The between subjects effects also showed significant group effects. A Tukey post hoc analysis indicated significant differences in time between groups 1 and 2 and groups 1 and 3. Groups 2 and 3 did not show any significant differences. Multiple comparisons analyzed showed that significant effects were due to differences between treatment groups 1 and 3 for question levels 1 and level 3. In terms of time, treatment group 3 that used shape and value as retinal variables in conjunction performed better than treatment group 1 that used color and shape as retinal variables in conjunction for level 1 questions. For level 3 questions treatment group 1 took significantly less time than treatment group 3. Thus question levels thus affect the time on task scores.

Within Groups Differences

Performance scores and time on task scores also varied within groups for the different levels of questions. Treatment group 1 that uses color and shape as retinal variables in conjunction, showed high performance scores and time on task scores. Questions in level 1 and 3 showed similar performance scores. Just as in other groups, level 2 questions yielded the largest performance scores. In terms of time on task, group 1 also had the highest scores. Question level 3 had the highest time on task scores followed by group 1 and group 2. Mental strategies reported by participants indicated that shape and color used in conjunction was overwhelming and made it challenging to identify the features and distinguish between them. While group 1 produced more accurate scores, participants took more time to complete the tasks.

Group 2 that uses color and value produced results that showed the highest performance scores for level 2 questions. Performance was significantly low for level 3 questions than in group 1. Time taken to complete task was low for all question levels when compared to other groups. However, except for questions in level 3, the time differences did not indicate any significant differences with other groups. Group 2 performed the highest for level 2 questions. However it is debatable as to whether these results are due to the retinal variables used or the task that was presented.

Group 3 produced results that contradicted each other. While performance scores were low for all question levels, participants also took the least amount of time to complete the tasks. While the performance scores were significantly lower than group 1, time on task scores were significantly lower than both groups 1 and 2 for question level 3. Time on task scores for questions in level 1 was significantly lower as well. More studies need to be conducted to provide more substantive evidence whether the time on task scores were due to the effectiveness of the retinal variables or due to lack of motivation of participants since performance in this group was low.

Besides the repeated measures analysis several correlation questions were also answered.

Question 1. What is the direction and strength of the relationship between performance scores and time on task scores based on the retinal variables used in conjunction in the displays?

Question 2. What is the direction and strength of the relationship between performance scores and time on task scores based on the level of complexity (location, local relationships and global trends) of questions presented using the different retinal variables?

Correlations between performance scores and time on tasks score in the three different groups were varied and ranged from weak to medium strength relationships. Directions were both positive and negative. The correlations were widely varied, and did not supplement any of the results obtained from the repeated analysis an comparison of means.

Question 3. What is the direction and strength of the relationship between usefulness rating and the performance scores measured as the number of correct responses, based on the different retinal variables (color, shape, value) used in conjunction in the displays?

The usefulness rating in treatment group 2 (color/value) showed a weak significant positive relationship with the question level 1 performance scores. It is worth noting that mean performance scores were the highest for level 1 questions in treatment group 2.

In addition to exploring statistically significant differences and correlations, the study also explores the sequence of the eye movements and the fixations scores that relate to each of the regions of interest defined by the researcher in each of the displays. Thus the study also investigates the following questions

Q_4. What is the direction and strength of the relationship between eye gaze time, saccade scores and total time taken to complete the task based on the different retinal variables (color, shape, value) used in conjunction in the displays?

Q_5. Is there a difference in the sequence in which participants look at the regions of interests defined in the displays in the three treatment groups as they perform their tasks?

An eye tracker was also used in the study to record the sequence of eye movements, the saccade and the fixation scores. Saccade scores had a weak negative correlation and fixation scores had a strong positive correlation, neither of which was significant. The researcher also defined 12 regions of interests that were defined to include data points that provided the answers to the questions presented. The fixation scores and the glance frequencies were noted. Each region of interest was visited at least once. While the eye movement patterns were varied, it must be noted that most participants started from the left lower corner of the display and their movements were focused on areas close to each other. It appeared that the participants adopted bottoms up approach and tried to identify the data points first and did not take a holistic view of the display before they looked at its parts. Glance frequencies were also highly varied. Regions 3, 4, 6, 9, and 10 had the highest glance frequency scores. Glance frequencies and performance scores when compared did not indicate any patterns.

Participants also recorded their mental strategies they adopted to complete the given task. Strategies described are well aligned with the eye movements. 73% of participants who reported mental strategies indicated they first tried to distinguish the features and understand what they represented. . . Finding the data locations were reported as most challenging. While there was certainly one's own perception and bias that was brought into accomplishing the task (11% of participants reported this), most participants indicated trying to repeatedly read and re-read the questions before answering the questions (53%)

Chapter Five discusses these results in more detail. It also provides supporting research evidence and compares results of this study with some of the research studies conducted in the past in the area of data visualization and multivariate displays.

Chapter Five

Discussion

This research study explored the effectiveness of retinal variables in facilitating interpretation of the interactions and relationships among data in a multivariate display. It was hypothesized that the differences in the retinal variables used (color/shape, color/value, value/shape) affects the performance and time on task at different levels of task complexity. This chapter discusses the results of the data analysis and the inferences that can be made from them. Each research question, with the supporting hypotheses and a summary of the results supporting literature is also presented. Future areas of research and conclusions are also included.

Implications of Research Questions 1and 2

$R_{1:}$ Is there a significant difference in performance measured as number of correct responses based on the retinal variables (color, value, and shape) used in conjunction in the display?

H_1: Different retinal variables used in conjunction (color/shape, color/value and shape/value) significantly affect user performance measures as number of correct responses.

$R_{2:}$ Is there a significant difference in performance measured as number of correct responses based on the complexity level (location, local relationships and global trends) of questions answered using the different retinal variables?

H_2: Differences in the levels of complexity of questions answered using a display created with retinal variables used in conjunction (color/shape, color/value and shape/value) significantly affects user performance measures as number of correct responses.

Mean performance scores by treatment groups and by question levels indicated differences. Furthermore, repeated measures analysis with question levels as within subjects variable and treatment groups as a between subjects variable showed significant score main effects and score by group interaction effects. The performance scores of participants showed significant mean differences at the different question levels in the different treatment groups. The differences were further examined through a Tukey post hoc analysis and it was found that the significant differences occurred between groups 1 and 2 and 1 and 3. However, multiple comparisons of means showed that only level three questions contributed to the significant differences. Thus the hypotheses related to question 1 was retained. However the hypotheses held true only for question level 3. Treatament group 1 color/shape used in conjunction appears to be a good design approach when representing data that involve questions that require higher order thinking skills and decision making. In other words, color/shape as retinal variables are more effective than both color/value and shape/value as retinal variables when representing level 3 questions. For levels that involve locating data points and comparisons, none of the groups showed any significant differences. All three treatment groups performed as good as each other in facilitating interpretation of the data trends.

Conclusions derived from the analysis based on performance scores and treatment groups include

1. Group 1 (color/shape) has the highest performance scores overall.

2. Participants in group 1 (color/shape) performed better than others when answering questions in level 3. There is a significant difference in the scores in these levels between groups 1 (color/shape) and 2(color/value) and groups 1 (color/shape) and 3 (shape/value)

3. There were no significant differences between groups 2 and 3.

Participants performed better when answering the level 2 questions in all treatment groups. Level 2 questions may have allowed participants to be more focused on a certain range of years in the display since it involved comparisons. This may have helped participants easily focus their attention on certain areas of the graphs and make decisions.

Implications for Research Question 3 and 4

R_3: Is there a significant difference in total time taken to answer questions based on the retinal variables (color, shape, value) used in conjunction in the displays?

H_3: Different retinal variables used in conjunction (color/shape, color/value and shape/value) significantly affect time taken to complete the given task.

R_4: Is there a significant difference in total time taken to answer questions based on the level of complexity (location, local relationships and global trends) of questions presented using the different retinal variables?

H_4: Differences in the levels of complexity of questions answered using a display created with retinal variables used in conjunction (color/shape, color/value and shape/value) significantly affects time taken to complete the given task.

Treatment group 1(color/shape) had the highest mean time on task scores among all groups. Level 1 questions took the longest time to answer in all the three treatment groups. Group 3 (shape/value) had the lowest time on task scores for all levels of questions followed by group 2 (color/shape). Level 2 questions showed the least time on task scores among all groups. Repeated measures analysis showed a time main effect and no significant interaction effects. Post hoc analysis indicated significant differences in time on task scores between groups 1 and 2 and 1 and 3. Thus the hypothesis that the level of complexity of questions affected the time on task was also retained. Multiple comparisons of means indicated significant differences in time taken to complete the tasks between treatment groups 1 and 3. For question level 1 which involves locating data points and question level 3 which involves decision making, shape and value used as retinal variables is better choice than color and value as retinal variables if just time on task were considered. There were no significant differences between other groups. Conclusions derived from the observations and the data analysis include

1. Participants who had the highest performance scores (group 1) had the highest mean time on task scores.

2. Level 3 questions that require global comparisons and studying of interactions across the display took the longest time to complete the task.

3. Open ended responses in level 3 that require participants to describe what they interpreted from the displays and provide supporting data for their decisions require more time than level 1 and 2 questions.

4. For identifying data points in a display, shape/value performed better in terms of time on task, than color/shape. Color/value as retinal variables used in conjunction performed as well as shape/value. There was no significant differences between group 2 (color/value) and 3 (shape/value) and group 1 (color/shape) and group 2 (color/value).

5. Questions in level 2 that allow for a more focussed study of the data interactions and relationships had the lowest time on task scores.

6. Shape and value as retinal variables facilitate a faster identification of data points and comparison of global trends and studying interactions

7. Color/value did not show any significant difference from any of the other treatment group at any question level.

Bertin's concept of retinal variables has been recently used in research studies involving instructional displays (Griffin & Robinson, 2000; Verde & Kulhavy, 2002), thematic maps (Rittschoff, Griffin & Custer, 1998), and information graphs (Shah & Hoffner, 2002). However, Bertin's theory and the use of retinal variables lack any link to cognitive information processing theories. The concept of retinal variables is based on the study of data visualization with a goal of enhancing the kinds of decisions that one could make from a given data set. (Kealy & Subramaniam, 2005). This study provides empirical evidence of the use of retinal variables for this data visualization and decision making perspective.

Salomon (1994) considered the retinal variables as components of a symbol system to understand their psychological role. In visually based symbol systems, the visible marks are symbolic elements that correlate to quantities or qualities that exists in some field of reference. This study used the concept of symbol systems where retinal variables were used to depict quantities and interactions in reference to the scores obtained by a certain population of students in math and language arts. Furthermore, interpretations of displays involve both ―bottom up and top down" processes. Salomon (1994) through his research studies showed that a top down process is driven by the expected task and the type of information. The bottom up processes on the other hand are mental operations evoked by the symbol system and its capacity for making discriminations, comparisons and inferences. Healy, Booth and Enns (1996) demonstrated how a data array using features that “pop-up” such as color and direction facilitated high speed estimation of salmon migration patterns. In this sense “bottom-up” processes are pre-attentive in nature, involving the immediate organization of a visual field that precedes a consciously driven search for information. (Kealy & Subramaniam, 2005). In this study, retinal variables color and shape used in treatment group 1 proved to be effective in facilitating study of global trends, data interactions and decision making. Color is a “pop-out” feature and hence pre-attentive. Any color with a distinct hue such as those (cyan and magenta) used in this study are selective variables (Filippakopoulou, et.al, 2005). The eye can isolate such variables and disregard any other information and isolate symbols in this category. While color and shape were effective for tasks that involved higher order thinking and decision making, color/value and shape/value were as effective as color and shape for tasks that involved location of data points and analyzing comparisons. Shape is a selective feature and value in different gradations removes the challenge of competing dominance of colors on the screen and makes identification of data points easy.

William Winn (1994) explains that components of maps and diagrams also vary in their relative dominance. Some components dominate others by virtue of their size, isolation and color. They are accorded more importance by the user. Precedence determines whether a component is looked at early or late in a visual. This provides a basis for why participants in treatment group 1 scored higher than others. However, in terms on time on task, participants in treatment group 1 had higher values than those in group 2 and group 3. Both shape and color can exhibit relative dominance. Color has precedence and is detected first, but the relative dominance of shape and color make it challenging for participants to differentiate features and study interactions. Thus while in terms of performance scores, shape and value used in conjunction were less effective than color and shape for questions that involved decision making, in terms of time on task scores, shape and value in conjunction were more effective that color and shape for both level 1 questions that identify data points and level 3 that required decision making.

Symbols in a graphic have to be detected by the perceptual system (Winn, 1994). Detection of symbols and their discrimination from one another are keys to the comprehension of graphical displays. 73% of participants who reported mental strategies indicated that they first discriminated between the features and tried to identify what they represented. Color and shape being dominant features may have been easier to detect and differentiate than the others. Hence there are significant differences in the mean performance scores between group 1 and group 2 and between group 1 and group 3. Mean differences between groups 1 and 3 were larger than those for group 1 and 2. On the other hand, it could be argued that increased performance required participants to take longer time to read the display and study the interactions. In addition, mental strategies reported by participants indicated that the two colors and shapes used were overwhelming and made differentiation of the features challenging.

Performance scores in group 2 (color/value) are less than those in group 1(color/shape) but higher than group 3 (shape/value). Treatment group 2 (color/value) however, did not show any significant differences in performance nor time on task scores except for question level 3, where treatment group 1 (color/shape) performed better. According to Filippakopoulou (2005), value, texture, hue and size are selective variables. The eye can isolate all symbols in this category immediately. Value and size are also ordered variables that lead to pre-attentive appreciation of underlying order. Since math and language arts scores used in the displays are quantitative variables, users can easily perceive and locate differences. While lack of color and the use of different gray scale levels excluded the issue of relative dominance, and made it easy to discriminate between the features and identify them, it made it challenging for participants to study relationships and global trends.

Filippakopoulou (2005) in his studies that evaluated Bertin's image theory also found that the selectivity of value depended on the gradation of the value effects used in the design of the display. Exponential rather than linear increases in the gradation help in differentiating the variables and studying their relationships. Values used in this study showed more of a linear gradation than exponential. This could have contributed to the decreased performance scores. 14% of participants in group 2 when asked to describe mental strategies used to answer questions indicated that since the only shape included in the display was a circle, it was challenging to differentiate the elements based just on value. Also, if the gradation of the values is not significantly different, it makes the task more complicated. This has valid implications to the design considerations for display of multidimensional data. Thus while treatment group 2 (color/value) allowed for the immediate visual encoding of the elements and discriminations between the various data points and their locations, studying interactions and relationships was a challenge.

Treatment group 3 had the lowest mean performance scores and the lowest time on task scores. The performance scores in group 3 showed a significant difference from those in group 1 for level 3 questions. There were no significant differences in question levels 1 and 2. According to Filippakopoulou (2005), shape is a selective feature. Coupled with lack of color that could contribute to relative dominance among the retinal variables, treatment group 2 made it easy for participants to identify and locate data points. Thus treatment group 3 participants took less time to identify the data point locations and compare them. However, since the performance scores recorded as the number of correct responses, were the lowest among all groups, the results need to be explored further. If the hypotheses that higher performance scores resulted due to more time taken to complete the task, decreased time on task scores seen in treatment group 3 (shape/value) could be due to low performance scores. Lower performances could have contributed to decreased motivation and hence completion of tasks at a faster rate than other groups. Treatment group 3 (shape/value) should be explored further. In terms of performance scores, the group was significantly less effective than group 1 (color/shape)for level 3 questions, and in terms of time on task scores group 3 did better than group 1 for level 1 and level 3 questions. The shapes, circle and squares used in the treatment both showed gradations of grayscale. Gradations used in this study, were not exponential but more linear. The possibility of effectiveness of value as a variable used in the study being a function of the gradation scale and the degree of differences in the values used should be further explored.

Levels of Complexity of Questions Used in the Study

Bertin (1983) explained information processing in displays as: 1) a combination of the variables presented, 2) the method used to represent them and 3) how the interactions are shown in the display. Wainer (1992) proposed that graphical efficacy or the efficiency of reading from graphs and displays is task based and consists of three levels of information processing: 1) retrieval of facts, 2) identification of local trends and 3) study of global trends. This study explored whether the retinal variables influence interpretation of data regardless of the levels at which the data were displayed. For all treatment groups, the time taken to answer the questions increased as the number of variables provided to aid in data search decreased. Level 2 questions showed both increased performance scores and the lowest time on task scores. Level 2 questions provided more variables in the questions than all other levels that helped participants to focus on specific areas on the display and study the interactions. Level 1 questions followed level 2 questions in terms of time on task scores. Even though questions in level 1 were the lowest in complexity requiring them to identify data location points for retrieval of facts, lack of understanding of how the displays work, and the learning curve that may have affected the time on task i.e., as participants became more comfortable with the displays by using it more, they performed better. While there is no empirical data from the study to support this rationale, observations made from the comments shared by the participants when they described their mental strategies are indicators of their performance and time on task being reflective of their comfort levels with the displays. Increase in usage and practice with the displays are really low when participants begin the task by answering the level 1 questions. Moreover, while the questions require participants to identify just data location points, there is no mention of any variables used in the display that would aid in answering the questions. Level 2 questions include the following variables as a part of their stem; 1) Student type (Japanese or American), 2) year, 3) subject area. These variables enabled participants to focus on certain areas of the display. In all other questions, more time was spent on searching the location of data points in order to study their relationships.

Level 3 open ended questions showed the highest time on task scores for all treatment groups. This study showed that the task of solving scenario problems was significantly easier when the displayed used color and shape (treatment group 1) as retinal variables in conjunction. Salomon (1984), suggested that by lowering the perceived cognitive demand associated with differentiating between symbols and components in a display, participants would be willing to invest greater amounts of mental effort to processing information presented. Thus the results of this study have serious implications to the design of displays that involve problem solving. Level 3 questions required participants to answer open ended questions.

As discussed earlier, a continuous piece of prose written in response to a question is commonly intended for higher cognitive levels of learning (Biggs, 1999). Chin, and Brown, (1999) conducted studies to understand differences between deep and surface learning in Science. They found that when students used a deep approach, they expressed their ideas more spontaneously; gave more elaborate explanations that described mechanisms and cause-effect relationships or referred to personal experiences; asked questions that focused on explanations and causes, predictions, or resolving discrepancies in knowledge; and engaged in "theorizing." Participants in all three groups with an exception of 4 participants out of the 123 who used the interactive displays on the desktop computers (3.1% of the total number of participants) obtained a rating of 10 points or more out of total of possible 12 points. While many of the participants in all treatment groups identified the right answer, they had trouble explaining and articulating their thoughts. Thus, while they paid more attention to searching, and identifying the data points, they did not use a deep approach in trying to explore relationships among the data. Surface learning vs. deep learning in a situation like this study, is a result of the effect of several variables such as interest, motivation and level of expertise. Even if participants are able to encode the relevant information accurately, their ability to map between different visual features and the meaning of those features may differ as a function of expertise (Shah & Hoeffner, 2002). Encoding viusal information is influenced in two ways by the visual characteristics of the display. Displays may interact with inherent biases and limitations of our perceptual apparatus to affect both the accuracy and information encoding (Cleveland & McGille, 1984, 1985).

This study provides empirical evidence that indicates that treatment group 1 (color/shape) is best suited for level 3 question types and that as participants begin to have a deeper understanding of relationships and interactions among the data used in the displays, they take more time to explain themselves and are more likely to come up with the accurate answers and effective decisions. For question levels 1 and 2 there were no significant differences among groups. In terms of time on task, participants in treatment group 3 (shape/value) took significantly less time than those in group 1 (color/shape) for both level 1 and level 3 questions. Treatment group 3 (shape/value) also facilitated the answering level 3 questions significantly faster than treatment group 2 (color/value).

However the performance scores in group 3 for level 3 questions was significantly lower than those of group 1. Hence these research results should be studied further.

Strength and Direction of Relationships Explored in the Study

The study included the following question that addressed the correlations between time on task scores and performance scores.

Q1.What is the direction and strength of the relationship between performance scores and time on task scores based on the retinal variables used in conjunction in the displays?

Correlations were explored between time on task scores and performance scores by group and by question levels, and between time on task scores and performance scores and usefulness rating to identify any possible relationships that may exist between these measures. Correlations showed weak to medium strength relationships. The only direct relationships between performance scores and time on task scores was seen in treatment group 2 (color/value), performance scores at question level 3 and time on task scores for question level 3 ($r=.604$, $p<.001$). Results from correlational analyses were varied and hence did not supplement any of the conclusions that were made through the repeated measures analysis and multiple comparisons of means. High performance does not relate to or indicate high or low time on task scores.

Usefulness Ratings

Usefulness ratings collected from participants were also correlated with time on task scores and performance scores by treatment group and by question levels. The questions related to such correlations included:

Q_2. What is the direction and strength of the relationship between usefulness rating and the performance scores measured as the number of correct responses, based on the different retinal variables (color, shape, value) used in conjunction in the displays?

Q_3. What is the direction and strength of the relationship between usefulness rating and the time taken to answer questions at different levels of complexity (location, local relationships and global trends) of questions presented using the different retinal variables?

Overall usefulness rating was the highest for treatment group 1 (3.317) and lowest for treatment group 3 (3.068), indicating that perceived usefulness of the displayed was based on a participant's performance. It must be noted though that the mean differences between the usefulness ratings in the three treatment groups were not high. Pearson product moment correlations showed strong positive correlations (r= .969) between performance scores and usefulness ratings. However, they were not significant.

Mental Strategies

Mental strategies described by the participants included tasks they performed to come up with the responses to each of the questions. Strategies varied from memorizing the two features in conjunction and what they meant, to re-reading, re-identifying and re-encoding of visual features several times, to blocking out the rest of the displays once the features and their meanings were identified, to just using general knowledge about graphical displays. These descriptions provide evidence to some of the already established information process related inferences made by several research studies that explored the graph comprehension (Shah & Hoeffner, 2002; Cleveland, 1993; Kosslyn, 1989; Winn, 1991). In summary, inferences made from the mental strategies described by participants include:

1. 73% of participants who reported mental strategies indicated identifying and isolating features first. Pop- out features that aid in the encoding of visual features as the participants enter the image (pre –attentive features) are critical to the effectiveness of the data display. Discriminations used as a component of the symbol system as proposed by Winn (1991) should facilitate identification of data location and study of their interactions and relationships.
2. Once the visual features are identified and encoded, the relationships and interactions are explored. 59% of participants who reported mental strategies indicated that once the features were discriminated, they read and re-read the questions to understand relationships. In order to facilitate a participant or user to stay focused on the features once they are identified, it is important to avoid any distracters that would make the visual encoding challenging. Tufte (1992) refers to such distracters as ―Chart Junk”

Eye Tracker Results

The study also included 12 participants randomly assigned to the three treatment groups who completed the task as their eye movements were being recorded. Since the experimental set up for these participants were different from those of the others, the eye tracker participants were considered as a separate group. The goal of this portion of the study was to identify patterns in eye movements and the regions of interests that participants fixated upon while completing the task and make inferences on the way in which participants search, identify and encode visual features.

The eye tracker data answered questions related to eye movements and their correlations to the time on task and performance scores and provided insight into the approach the participants took to solve the questions presented to them in the study. The research questions include:

Q_4. What is the direction and strength of the relationship between eye gaze time, saccade scores and total time taken to complete the task based on the different retinal variables (color, shape, value) used in conjunction in the displays?

Overall gaze scores and saccade scores were correlated with the overall time on task scores and these scores in the three treatment groups were also analyzed. For the overall mean scores there were no significant relationships. Time on task and saccade scores showed a positive weak correlation (r=.485) which was not significant and time on task and fixation scores showed a positive strong relationship (r=.861) which was also not significant.

Correlations between time on task scores, and fixation and saccade scores by group showed strengths ranging from weak to strong. Except for time on task scores and saccade scores in group 2, which showed a strong negative correlation (r=-.825), all other relationships were weak and positive. Thus time on task scores did not have any significant correlations with saccade and fixation scores.

Q_5. Is there a difference in the sequence in which participants look at the regions of interests defined in the displays in the three treatment groups as they perform their tasks?

Interpretations of graphical displays involve both bottom- up and top down approaches (Salomon, 1994; Schnotz, 2002). While the top down approaches are driven by the task on hand and the type of information presented the bottom up processes are mental operations evoked by emergent characteristics of a symbol system including but not limited to, its capacity for making discriminations, comparisons and inferences (Kealy and Subramaniam, 2005). Initially as a novice, when one is trying to learn to read displays and problem solve the bottom up approaches predominate. Over time and use, as the decoding skills improve; mental activities that are top down in nature guided by the task on hand direct the focus of attention, problem solving and knowledge building. This is indicated clearly by the participant's eye movements in the study. Eye scan paths, showing the sequence, regions of interests visited and the glance frequencies related to each of the regions of interests indicate that the participants were immediately engaged in search and identification of the features and discriminating between them. Studying the relationships and the interactions came later as they fixated on the regions of interests. The bottoms up mental operations use the working memory and when challenged, could result in a cognitive load (Krischner, 2002). Thus retinal variables that aid in the bottom up processes and facilitate identifications and discriminations among features are most suitable for use in the design of displays.

All participants started on the left middle section of the display and moved quickly to areas where data points were in close proximity to one another. The general tendency seemed to be to start the search and identification immediately, and focus on areas where features were concentrated. However, an analysis of the glance frequencies and the subsequent eye patterns for each of the participants showed differences. There were individual differences in the way each participant perceives the image they see based on their own perceptions, knowledge and biases, but 73% of the 98 participants who reported mental strategies, including both eye tracker participants and those that used just the interactive displays tried to first locate and discriminate the features. Once this initial phase was complete, there are indications of revisits to the regions, and hence re-identification. If symbols are not identified at a first attempt, other interpretations are tried iteratively until identification is made (Palmer, 1975). It must also be noted that not all participants whose eye movements were recorded used the legend. A few of them first identified data points and then used the legend to study the relationships.

The complexity of questions used in the research may have also contributed to the eye movements. Level 3 questions that required participants to identify global trends and study several interactions and relationships showed several scan paths that moved both horizontally and vertically across the stimulus display.

Conclusion

Comprehension occurs when a symbol system used to convey a message interacts with psychological processes active in the person who receives it (Salomon, 1974). From the view point of the visual attention theory and attention processing, comprehension is the result of the interaction between top down knowledge and task driven processes and bottom up data driven processes (Neisser, 1976). Perceptual organization is affected by the structural characteristics of the graphical display and interpretation of what the characteristics mean and hence the attentive processing also depends on structural characteristics of the display (Winn, 1991). Also, comprehension succeeds or fails to the extent that the information organized by the pre-attentive processes can be assimilated to the existing schema or the schemata can be altered to include the information comprehended (Winn, 1991). According to Shah and Hoeffner (2002), three factors play an important role in the viewer’s interpretation of data: 1) the characteristics of the visual display, 2) the viewer's knowledge about graphs, and 3) the variables and/or content of the display. Data analysis and conclusions from this research study provides further evidence to these theories and research results. Previous studies in graph comprehension also provided empirical evidence to Bertin's proposed usefulness of the retinal variables. Conclusions and inferences made from the research results include

1. Color, shape and value exhibit varying degrees of usefulness as discriminators in the display of multidimensional data.
2. Color, value and shape identified as pre-attentive features (Healy, 2000) facilitate the identification of symbols and their discrimination (level 1 questions). However, if performance is measured as a function of time taken to complete the task rather than just accuracy of responses, shape and value are more effective than color and shape when used in conjunction. Color and value used in conjunction is equally effective as shape and value and color and shape for identification and discrimination. However, the study does not provide any substantive evidence to show that a treatment group is effective both in terms of both performance and time taken to complete the task. Hence further exploration is required to make a determination in this regard.
3. For studying comparisons and relationships among features, color, shape and value used in conjunction are effective. All these features perform equally well as discriminators for questions in level 2.
4. Color and Shape used in conjunction are best suited for identification of data points and retrieval of information, making comparative judgments and

identifying global trends. Thus they are also suitable for higher order thinking as well as knowledge building.

5. Value, is an effective discriminating factor. However optimal gradation scales for values used in a display should be determined before their use. Color and value used in conjunction were as effective as color/shape and color/value in terms of performance scores except for questions that involved decision making and studying global trends. In terms of time on task scores color and value in conjunction were as effective as shape and value for question level 1, and showed significant higher time on task scores than shape and value used in conjunction for question level 3. Color and value had the highest performance scores for questions in level 2, and also did not show significant differences in time on task scores from other groups for this level. Hence color and value may be more suitable for questions in level 2 where participants are required to make comparisons. However more studies need to be conducted to provide substantive evidence to support this conclusion.
6. Shape and value were as effective as color and shape, color and value used in conjunction in the identification of features and comparing them. Moreover, shape and value showed significantly lower time on task scores than color and shape used in conjunction for question level 1. Thus shape and value and color and shape are probably better than color and shape for the identification of data point locations (question level 1). However, shape and value indicated significantly lower performance scores and lower time on task scores for questions that involved higher order thinking skills. This needs to be explored

further since results from this study do not lead to any substantive conclusions.

7. Based on the variability of the results related to performance scores and time on task score differences between color/shape and shape/value, color and value could be considered as the most suitable for identification and discrimination of data (question level 1) and studying comparisons (question level 2). Color and shape used in conjunction are most effective for tasks that involve studying global trends and decision making (question level 3).
8. Providing information in questions that allow the users to focus more on specific areas of the display and study the interactions results in better performance and decreased time on task.
9. 11.2% of the participants who reported mental strategies indicated that they did not follow any sequence or strategies. They used their knowledge of graphs and the lessons learned from reality. 2 participants indicated their concern that language arts scores could be influenced by Japanese lack of knowledge of the English language. The study supports Shah and Hoeffner's (2002) conclusion that interpretation of relationships and interactions are influenced by one's knowledge of graph, their perceived judgments and biases
10. Descriptions of scenarios that allow biases to be removed, and promote an objective approach to looking at the display will further aid in speed and performance. Detailed descriptions and orientation to the goals and task to be completed are necessary as well.

11. The study provides further evidence to the fact that users not familiar with the displays used and the data patterns use most of their working memory resources in search and identification of features, and retaining the search results as they explore relationships.
12. Reducing cognitive load by facilitating easy search and identification of features and symbols to allow for more working memory to be available for study of relationships and their processing should be the goal of design approaches to graphical displays.

In summary, shape/value used in conjunction is better suited for identification and discrimination of data points (question level 1) than color and shape. Color value when used in conjunction was as effective as other groups. All three combinations of retinal variables color/shape, color/value and shape/value were effective in analyzing comparisons and interactions (question level 2). These results could be compounded however by the presentation of the tasks for question level 2. The questions included more than one variable that enabled participants to complete the responses quickly and accurately. Further studies are needed to explore these data trends. In terms of performance scores for tasks that involved studying global trends and decision making, color and shape were significantly more effective than color and value and value and shape. However, if high performance is a function of both accuracy and speed, then further studies are needed, since shape and value which had significantly low performance scores, showed significantly lower time on task scores for tasks that involved identification of data points and decision making. (Question levels 1 and 3).

Color and value did not show any significantly lower time on task scores for question level 1 but were significantly higher than shape and value for question level 3.

~~"~~Scanning of successive locations serially and focal attention is the means by which the correct integration of features into multidimensional percepts are ensured" (Treisman, 1980, p106). This study further supports the feature integration theory in that features when used in conjunction are treated as a single unit, searched and identified first. Such searches deal with a central scan with focused attention and deals with each single item. Hence allowing for discriminations and searches to be more effective through design is critically important. Treisman and Gelade (1988) also believed that the amount of differentiation between the target and the distracters for a given feature affects search time. The results from this study support the findings. Increased time on task scores may be the result of the time taken to differentiate between the features and study global trends. The results also support several findings that relate to analysis of cognitive processes involved in the interpretation and information processing when using graphical displays. Studies include Winn (1991), Healy, Booth and Enns (1996), Salomon (1979), Shah and Hoeffner (2002), all of which deal with the notational symbol system, their ability to inform the relationships, the importance of their locations, organization and presentation. Comprehension occurs when a symbol system used to convey a message interacts with psychological processes active in the person who receives them (Winn 1994, p2). This study showed that features used in the display play an important role in the identification of data locations and their relationships. The results also support the studies that demonstrated that reasoning with graphs involves interactions between perceptual and cognitive abilities that vary with specific task requirements (Peeble &

Cheng, 2002). Both performance and time on task scores varied with the different question levels. Wainer (1994), explored the graphical efficacy in humans and discussed the three levels of information processing. The study results showed that questions in level three which required global comparisons and a thorough understanding of the data relationships were the most challenging and took the longest time to complete. These findings support Wainer's theory of graphical efficacy.

The study explored Bertin's theory on the effectiveness of retinal variables in the display of multivariate data. The results support the theory only based on the three variables that were used, color, value and shape.

Recommendations for Future Studies

There are several recommendations for future research as a result of this study. Analyses conducted to evaluate dependability of scores recorded automatically by the interactive computer program showed inconsistencies in responses in levels 1 and 2. The research study needs to be conducted accounting for the factors that contributed to such variations and thereby increasing the reliability scores. Reliability scores from questions to be included should be evaluated and appropriate revisions should be recommended before the questions are included in a future study. The study needs to also include a more heterogonous group of participants, more than three questions at each level, and a thorough examination of the interactive computer program to identify any inconsistencies contributed by the interactive displays in the three treatment groups. Colors magenta and cyan and circles and squares as shapes were used in this study. Further exploration of the differences in effectiveness of the different colors and shapes need to be explored. Most optimal colors and shapes need to be identified to provide specific information on the types of shapes and colors that facilitate information processing effectively at different levels of complexity of tasks to be performed. Gradations in value when used as a discriminating feature that would be most suitable for display of multivariate data need to be investigated. Value, second to color could prove to be a valuable discriminator if ways in which we can make different values pop up and be distinguished from each other can be defined. Color, shape and value were the only retinal variables tested. Other variables as proposed by Bertin need to be studied. The mental strategies shared by participants in the study indicate that the features are searched first and their relationships are identified. Hence emphasis on the design of discriminations (Winn, 1994) is critical to establishing design guidelines.

If effectiveness is defined as a function of accuracy and speed, color and shape and shape and value should be studied further. While shape and value used in conjunction showed low performance scores, it also showed low time on task scores. Further explorations that focus on whether increased performance increases time on task, or if poor performance decreases the motivation to complete the task with a thorough analysis of the data display are required.

Features are identified and searched first and participants followed bottoms- up approach. Identifying specific differences in the ways in which novices and experts look at the same display and perform the same task would help recommend ways to help novices and learners improve their graphical efficacy. ―Think aloud Protocols" such as those used in visual cognition research, can be used to record the various strategies and steps taken by both novices and experts and their patterns can be studied. The results could provide recommendations to novices on how to improve their skills in reading graphs. There are several differences in the cognitive processes and strategies undertaken by experts who to problem solve that have been established by several research studies (Huang & Eades, 2005; Shah & Hoeffner, 2002; Gobet & Simmons, 2006). Findings from these studies suggest that understanding the way experts recognize patterns based on accumulated knowledge and their analysis by search will help discover the balance of recognition and search in tasks involving medical diagnosis, engineering design, critical thinking and logic. Using specific groups of participants with prior experience in reading data from several disciplines of medical sciences, engineering, environmental science, psychology and other areas where data visuals are a critical part of such participants' everyday tasks would provide interesting insights into how participants view displays and make inferences based on their experiences

Eye tracker data in this study included results from only 12 participants. To provide further evidence, such data should be collected from a larger participant sample.

Eye scan path were recorded for the entire duration of the task that was performed. It would be beneficial to record eye data separately for each of the three different levels of complexity of questions used. Such patterns will then indicate specific eye movements related to the levels of the tasks to be accomplished. Based on the what and where aspects of the visual attention theory, the proposed bottom up and top down approaches to search and pattern recognition, identification of differences in eye movements based on tasks at hand or complexity levels of problem presented would provide more substantial information on how people search and make connections at different levels of complexity and for different tasks.

The format of the display is an important factor that influences how participants respond to graphical displays (Shah & Hoeffner, 1992). A future study that compares the eye movements and glance frequencies by defining regions of interests in bar graphs and other formats displaying multivariate data would provide evidence as to whether eye movements, searching and identifications vary with the different formats.

References

Abel, R.R., & Kulhavy, R.W 1986. Maps, mode of text presentation, and children's prose learning. *American Educational Research Journal*, 23, (pp. 263 – 274).

Anderson, J.R. (1990). *Cognitive Psychology and its implications.* San Francisco: W.H Freeman, (3rd edition).

Anzai, Y. (1991) Learning and use of representations for physics expertise. *Towards a general theory of expertise*. (pp. 64 – 92). Cambridge University Press.

Ausubel, D 1968. Educational Psychology: A cognitive view. New York: Holt, Reinhart & Winston.

Baek, Y., & Layne, B.(1988). Color, graphics and animation in a computer assisted learning tutorial lesson. *Journal of Computer-based Instruction*, 15 (4), (pp.131 – 135).

Banathy, B. 1987. Instructional systems design. In R.Gagne(Ed.) , *Instructional Technology :Foundations* (pp. 85 - 112). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bertin, J 1983. Semiology of Graphics: Diagrams, Networks, Maps. Madison, WI: University of Wisconsin.

Bertin, J.1978. Theory and Communication of ―The Graphic". *International Yearbook of Cartography* 18, p. 119.

Bertin, J 1981. Graphics and graphic information processing. New York: Translated by William J. Berg and Paul Scott Berlin and Walter de Gruyter.

Bruner, J 1986. Actual minds, possible worlds Cambridge MA: Harvard University Press.

Buswell, G.T. (1935). How people look at pictures. Chicago: University of Chicago Press. Carpenter, P.A, & Shah, P., (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4, (pp. 75 – 100).

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction,* 8(4), (pp. 293-332).

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology,* 62, (pp. 233 – 246).

Chi,M., Feltovich, P., & Galser , R.(1981) Categorization and representation of physics problems by experts and novices. *Cognitive Science,* 5, (pp. 121 – 153).

Chun, M.M., & Wolfe, J.M. (2001). Visual Attention- Blackwell Handbook of Perception. Editor E.B. Goldstein. Blackwell Publishing, (pp. 272 – 310).

Cleveland, W.S. 1993. A model for studying display methods of statistical graphics. *Journal of Computational and Graphical Statistics*, 2, 323-343.

Duncan, J. & Humphreys, G.W (1989). Visual Search and Stimulus Similarity. Psychological Review 96.

Eastman, J.R. 1985. Graphic organization and memory structures for map learning. *Cartographica*, 22, (pp. 1 – 20).

Enns, J. T. (1987). A developmental look at pattern symmetry in perception and memory Developmental *Psychology*, 23(6), (pp. 839 – 850).

Enns, J.T. Seeing Textons in context. Perception & Psychophysics 39 (1986), (pp.143 – 147).

File, E, Portia, 1979. Visual Memory for Color. *Acta Psychologica* 46, (pp.103 – 114).

Finke, R. A., & Pinker, S. (1983). Directional scanning of remembered visual patterns. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(3), (pp. 398 – 410).

Fleming, M. L. (1978). Instructional message design: principles from the behavioral sciences. Englewood Cliffs, NJ: Educational Technology Publications.

Goldwaith, C. (1993). Relation of eye movements to visual imagery. *American Journal of Psychology.*

Green, E, Mark, 1991. Toward a Perceptual Science of Multidimensional Data Visualization:Bertin and Beyond. *Journal of Human Factors Science,* 34, (pp. 617 – 635).

Hayhoe, M.M, Besinger, D.G., & Ballard, D.H. (1998). Task constraints in visual working memory. *Vision Research,* 38(1), (pp. 125 – 137).

Herrmann D., Kerwin J., Pickle L.W., Croner C., White A., Jones G., Jobe J. 1994. Cognitive processes in statistical map reading. *Proceedings of the Statistical Graphics Section*, American Statistical Association 1993 Meeting, San Francisco.

Healey, C, Kellog.B, & Enns, J (1996) High Speed visual estimation using preattentive processing. *ACM Publication.*

Healey, C & Enns, T. (2002) Perception and Painting: A search for effective, engaging visualizations. *IEEE Computer Graphics and Applications*, (pp. 10 – 15).

Healey, C (2006). Perception in Visualization Obtained from the world wide web at http://www.csc.ncsu.edu/faculty/healy/PP/index.html

Hoffman, J.E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception and Psychophysics*, 57, (pp. 787 - 795).

Huang, W & Eades, P., (2005). How do people read graphs. Proceedings of the Asia Pacific Symposium on Information Visualization, Vol 45.

Johnson, J. T., Verdi, M. P., Kealy, W. A., Stock, W. A., & Haygood, R. C. (1995). Map perspective and the learning of text. *Contemporary Educational Psychology,* 20, (pp. 457 - 463).

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology,* 13, (pp. 351 - 371).

Kealy, W. A. & Sullivan, H. J. (1991). Question density and processing attention in computer-based instruction. *British Journal of Educational Psychology*. 61, (pp. 230 – 232).

Kealy, W. A. & Webb, J.M. (1995). Verbal Learning with maps and diagrams. *Contemporary Educational Psychology*, 20, (pp. 340 – 358).

Kosslyn, S.M. 1985. Graphics and human information processing. *Journal of American Statistical Association*, 80, (pp. 499 - 512).

Kosslyn, S. M. (1994). Image and brain: The resolution of the imagery debate. Cambridge, MA: MIT Press.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research,* 61(2), (pp. 179 – 211).

Kulhavy, R. W., Stock, W. A. & Kealy, W. A. (1993). How geographic maps increase recall of instructional text. *Educational Technology, Research and Development*, 41(4), (pp. 47 – 62).

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, (pp. 65 – 99).

Lohse, G.L (1997). Consumer eye movement patterns on yellow pages advertising. *Journal of Advertising*, 26 (1), (pp. 61 – 73).

Mayer E.R., 1992. Designing Instruction for Constructivist Learning. In Charles M Reigeluth, *Instructional Design Theories and Models,* vol 2, (pp. 141 - 156), Hillsdale, NJ: Lawrence Erlbaum Associates.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist,* 32(1), (pp. 1 – 19).

Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), (pp. 3 – 25).

Paivio, A. (1991). Dual coding theory: retrospect and current status. *Canadian Journal of Psychology,* 45 (September), (pp. 255 – 87).

Reiber, P, Lloyd, 2000. Computers, Graphics and Learning. Brown and Benchmark Publishers, Wisconsin (pp. 35 – 52).

Reiber L.P, 1995. A Historical overview of visualization in human cognition. *Educational Technology Research and Development*, vol 1 (pp. 45 – 56).

Rittschof,A,K., Griffin M. M., & Custer, W.L., 1998. Learner differences affecting schematic for thematic. *International Journal of Instructional Media.* vol 25, 2, p. 179.

Robinson, H.D., Robinson, S., & Katayama, A 1999. When words are represented in memory like pictures: Evidence for spatial encoding of study materials. *Contemporary Educational Psychology* 24, (pp. 38 – 54).

Robinson, D. H., & Schraw, G. (1994). Computational efficiency through visual argument: do graphic organizers communicate relations in text too effectively? *Contemporary Educational Psychology,* 19(4), (pp. 399 – 415).

Shah, P & Hoeffner, J 2002. Review of Graph Comprehension Research: Implications for Instruction. *Educational Psychology Review* vol 14, No 1.

Scaife, M., & Rogers, Y. (1994) External Cognition: How do graphical representations work?. University of Sussex, UK.

Spivey, M.J., Richardson D.C., & Young, E. (2000). Eye movements during comprehension of spoken scene descriptions. Paper presented at the 22nd Annual conference of cognitive science society, Philadelphia.

Tufte, E.R. 1983. The visual display of Quantitative Information. Chesire, CT Graphics Press.

Tufte, E. R. (1990). Envisioning information. Cheshire, CN: Graphics Press.

Tukey, J.W. 1979. Statistical mapping: what should not be plotted. *Proceedings of the 1976 workshop on Automated Cartography and Epidemiology*, Hyatsville, MD, DHEW Publications No. (PHS) (pp. 79 – 1254, 18 -26).

Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology* 12, (pp. 97 – 136).

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics and Image Processing* 31, (pp. 156 – 177).

Wainer, H., Francolini,C.M 1980. An empirical inquiry concerning human understanding of two variable color maps. *American Statistician*, 34, (pp. 81 – 93).

Wainer, H 1991. Understanding Graphs and Tables. *Educational Researcher*. vol 21, no 1, (pp. 14 – 23).

Winn, W. D. (1991). Learning from maps and diagrams. *Educational Psychology Review,* 3(3), (pp. 211 – 247).

Winn, W. (1993). The effect of the spatial arrangement of simple diagrams on the interpretation of English and nonsense sentences. *Educational Technology, Research and Development,* 41(1), (pp. 29 – 41).

Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher*, 21(1), (pp. 14 – 23).

Yarbus, A.L (1965). Role of eye movements in the visual process. Oxford, England: Nauka.

Yarbus, A.L., (1967): Eye movements and vision. New York, Plenum Press.

Appendices

Appendix A: Participant Sign-Up Sheet

Participant Name	Major area of study	Time slot	Class visited

Appendix B: Study Descriptor

Thank you for your time.

The study explores different design approaches to display data and their interactions using variables such as color, value and shape in conjunction.

For the research study, you will be looking at a display that involves the variables color, value and shape used to display math and language arts scores obtained by Japanese and American students, from the years 1995-2006. You will be asked to answer ten questions related to the displays. The questions will be presented in no particular order. The display and the question will appear in the same screen one after the other. Some of them need to be answered by clicking on a data point on the actual display and some of them are open-ended questions that required you write short answers in the text field provided. You may proceed from one question to another once you are convinced that you have answered it. There is no feedback. The program will automatically score the right answers and will also time your responses. At the end of ten questions, you will be asked to rate the usefulness of the displays in helping you answer the questions and also share with the researcher any mental strategies that you employed in answering the questions. You are required to answer the questions as quickly as possible.

Before you begin the actual study, you will be given an opportunity to practice the task on a display that is exactly the same as the one that is going to be used in the study. You will be asked to go through the practice session in 10 minutes.

If at any time you have any questions regarding the task, you may feel free to talk to researcher. Also you can leave the room at anytime with the task unfinished if you choose to do so. Once you have completed the ten questions, please be sure to pick up a business card from the researcher. This will ensure you receiving extra credit for your time and support for the study.

Appendix C: Scores of Japanese and American Students in Math and Language Arts between the Years 1995 - 2006

Year	**Math**		**Language Arts**	
	Japan	US	Japan	US
1995	596	516	464	551
1996	542	480	386	532
1997	683	534	450	669
1998	620	452	396	640
1999	606	491	507	572
2000	542	399	262	515
2001	620	372	438	420
2002	601	283	401	390
2003	529	330	312	363
2004	680	263	290	404
2005	657	313	404	302
2006	493	270	482	293

Appendix D: Questions Used in the Displays

Question for sample display

Which year did Europe have the largest number of female visitors. Click on the data point that indicates the largest number of visitors against the year.

Questions Level 1

1. In math, Japanese usually score more than Americans. Identify the highest score obtained that substantiates such a trend. Click on the data point that shows the correct answer – year 1997 Japan 683

2. Among both Japanese and Americans, identify the highest score obtained in Language arts. Click on the data point that shows the correct answer. – Year 1997, US 669

3. Are Japanese men the least interested in language arts? Identify the data point that indicates this level of interest. Click on the data point that shows the correct answer. – year 2004, Japan 290

Questions Level 2

4. In the year 2000, both in math and language arts among Japanese and Americans, what was the highest score obtained. Click on the data point that shows the correct answer. – Japan 542, Math

5. Between the years 1997- 2001, in math among Japanese and American students what is the lowest score? Click on the data point that shows the correct answer. – US 372, year 2001

6. Between the years 2004- 2006, what was the highest score the Japanese obtained? Compare both Math and Language scores. Click the data point that shows the answer. – year 2004, 680

Questions Level 3

7. The university received a recent grant to offer extra help and support to students who were Math majors. The grant was reserved for students who could use additional help. When the scores of all students were compiled, even though at first sight it looked like the US and Japanese students had performed more or less the same, it was found that the US students could use some additional help more than the Japanese in math. Explain your answers using the data to provide evidence.

8. The US department of Education is investigating scores of Americans in math to identify the years where the performance declined. Looking at the data display explains when you think the decline happened and which are the years that the performance is low. Provide data to substantiate your results.

9. The university is looking for students who can be spokespersons to explain both their math and language arts programs to incoming new students. In order to be a spokesperson for the university, a student should have received high scores in math and language arts. Based on the scores, which group of students would you select as spokespersons for the two programs? Explain your answers with supporting data.

Appendix E: Rubric for Evaluation and Scoring of Open-Ended Questions

Category	4	3	2	1
Accuracy of facts	All supportive facts and inferences made are reported accurately	Almost all supportive facts and inferences are reported accurately	Most supportive facts and inferences are reported accurately	No facts are reported OR most facts are inaccurately reported
Support for facts	Relevant, telling quality details giving the reader more information beyond what is predictable	Supporting details and the information are relevant, but one of the key concepts are missing	Supporting details and the information are relevant, but many of the key concepts are missing	Supporting information and details are unrelated to the questions being answered
Clarity and Focus	There is a clear focus on the question asked. The main idea stands out and is supported by details	Main idea is clear but the supporting information is very general	Main idea is somewhat clear but there is need for more supporting information	The main idea is not clear. There is random collection of information
Sequencing	Details are presented in an effective logical order	Details are placed in a logical order but the presentation is less effective	Some details are not in logical order and the presentation distracts the reader	Many details are not in logical or expected order

Appendix F: Data Collected From Participants in Color Shape Treatment Group

Group	Score L1	TOT1	Score L2	TOT2	Score L3	TOT3	Rating
1	8	129.52	12	82.94	3	163.41	1
1	4	151.18	12	67.75	8	509.71	4
1	8	103.13	12	114.9	6	161.57	2
1	8	143.9	0	59.44	7	226.5	4
1	4	132.3	12	115.63	7	366.43	3
1	8	153.4	8	68.81	9	290.49	5
1	8	164.14	8	60.8	4	333.9	3
1	4	155.77	12	53.26	5	236.89	4
1	8	91.64	12	62.31	9	539.4	2
1	4	85.49	12	108.71	8	413.06	4
1	8	60.37	8	64.4	12	473.73	4
1	0	106.37	12	53.8	4	353.53	2
1	4	137.4	12	93.26	7	361.9	4
1	8	60.1	8	95.06	7	306.63	4
1	8	71.54	12	81.23	2	368.1	3
1	8	77.81	8	36.07	3	155.17	3
1	0	101.3	12	48.8	4	239.6	2
1	8	97.83	8	64.11	5	446.69	2
1	8	70.06	8	50	4	284.8	4
1	8	84.04	8	34.93	5	299.47	3
1	4	98.83	8	75.44	3	186.13	2
1	8	127.07	8	54.27	4	189.87	4
1	8	85.67	0	79.34	6	373.26	4
1	4	78.73	4	82.4	3	287	3
1	8	55.74	12	76.07	7	374.14	3
1	0	181.67	0	152.58	10	627.55	4
1	4	113.19	8	95.53	5	536.62	4
1	8	88.73	8	59.47	1	456.92	3
1	8	68.22	12	53.35	4	204.52	4
1	0	83.63	12	105.1	5	315.17	3
1	8	87.58	8	56.88	6	362.77	5
1	8	185.79	12	66.29	7	413.02	4
1	8	104.04	8	34.93	6	299.47	3
1	4	90.5	0	37.25	3	272.61	3
1	4	109.93	4	63.69	3	287.36	4
1	4	63.67	8	63.28	3	244.05	5
1	4	62.91	8	49.81	4	212.35	3
1	8	94.35	8	45.16	1	196.1	3
1	8	111.73	8	86.45	9	1287.48	4
1	4	78.42	0	74.63	4	150.79	2
1	4	40.5	0	67.25	3	272.61	3

Appendix G: Data Collected From Participants in Color Value Treatment Group

Group	Score L1	TOT 1	Score L2	TOT 2	Score L3	TOT 3	Rating
2	4	73.33	12	62.9	5	282.23	5
2	0	98.84	8	46.43	2	166.41	1
2	8	114.73	8	46.33	5	263.57	1
2	8	175.83	12	133.13	5	404.24	2
2	4	103.1	8	108.47	6	319.57	3
2	0	77.5	8	59.06	7	290.9	1
2	4	83.96	8	54.96	6	306.13	3
2	8	77.3	8	70.76	2	114.2	5
2	4	145.33	12	59.96	6	379.33	4
2	8	93.17	12	68.77	7	653.06	4
2	8	74.47	8	55.81	5	302.67	4
2	4	73.8	8	37.83	1	167.03	4
2	8	81.88	8	71.24	5	490.39	4
2	0	78.84	8	46.43	4	166.41	1
2	4	104.83	8	49.73	3	292.95	3
2	8	60.41	8	49.16	3	269.93	1
2	0	77.5	8	59.06	6	290.9	1
2	4	76.11	4	64.14	4	253.63	3
2	4	75.1	8	47.66	3	223.14	4
2	8	95.27	8	94.87	3	187.13	4
2	4	93.87	8	81.5	5	266.73	4
2	8	56.41	8	49.16	4	269.93	1
2	0	67.5	8	59.06	5	290.9	1
2	8	101.6	0	44.8	3	214.94	2
2	8	61.03	8	34.6	4	306.6	5
2	8	101.04	8	29.93	3	274.2	4
2	0	73.6	4	47.73	4	270.77	2
2	4	98.25	12	44.6	3	273.53	1
2	4	126.6	4	73.5	4	267.46	4
2	8	97.06	0	44.8	2	214.94	2
2	4	140.53	4	45.39	3	218.27	5
2	8	98.97	12	61.73	2	486.47	5
2	8	95.67	8	66.73	1	117.87	5
2	0	114.97	12	102.29	3	304.61	3
2	0	78.36	8	55.87	2	189.2	3
2	4	68.57	12	62.5	2	303.72	4
2	4	22.28	8	17.77	1	124.08	4
2	4	81.93	8	52.34	5	286.34	5
2	4	85.27	4	83.17	2	168.82	4
2	4	57.67	8	36.05	6	371.53	4
2	0	108	8	65.1	3	336.7	2

Appendix H: Data Collected From Participants in Shape Value Treatment Group

Group	Score L1	TOT1	Score L2	TOT2	Score L3	TOT3	Rating
3	0	108.04	8	65.13	3	336.66	2
3	8	69.24	12	52.11	1	274.09	1
3	8	94.47	8	49.63	3	102.03	3
3	8	59.77	8	84.33	1	236.89	2
3	0	61.67	0	58.5	3	314.99	1
3	0	58.77	0	39.44	1	237.25	3
3	4	72.17	12	68.8	5	326.9	4
3	0	217.6	12	81.36	1	530.5	1
3	4	84.73	0	85.54	0	225.26	1
3	4	144.3	0	90.76	0	292.13	1
3	0	50.44	0	102.47	5	261.53	1
3	8	83.86	12	88.21	1	224.35	3
3	8	71.76	0	84.19	3	164.9	3
3	8	62.16	8	60.09	2	273.17	4
3	8	60.27	8	68.23	4	251.96	4
3	8	145.37	4	102.27	7	348.04	2
3	8	85.87	12	41.67	2	319.43	3
3	8	48.56	12	51.44	1	174.81	4
3	8	124.6	0	56.86	4	198.94	1
3	8	65.66	8	43.2	0	276	4
3	8	60.16	12	62.66	4	332.43	2
3	4	93.96	0	49.14	2	201.45	5
3	8	48.56	12	51.44	2	174.81	4
3	8	90	8	63.43	3	234.73	3
3	8	66.61	8	90.07	2	241.49	3
3	4	61.97	8	43.03	0	276.45	4
3	8	95.76	4	52.27	2	271.5	3
3	8	95.76	4	86.27	4	281.5	3
3	4	61.27	12	61.33	4	343.13	5
3	0	47.95	4	83.18	3	449.65	4
3	4	58.4	12	50.53	3	338.92	3
3	4	63.07	8	48.61	2	205.26	3
3	8	95.76	4	62.27	2	271.5	3
3	8	103.53	8	54.46	3	237.81	2
3	8	102.04	8	52.86	3	250.41	4
3	0	47.95	4	83.18	2	437.89	2
3	4	53.27	8	57.6	3	271.2	2
3	4	109.5	12	56.98	3	303.37	2
3	4	63.07	8	44.61	2	205.26	3
3	8	59.93	8	60.97	2	266.49	3
3	8	138.85	8	90.77	1	183.39	2

Appendix I: Common Sample Mental Strategies Described by Participants in the Three Treatment Groups

Group 1	Group 2	Group 3
I looked at the graph and analyzed it first. Then I concentrated on the obvious trends and patterns throughout the graph. This helped in answering the questions and making decisions	The only time it took was careful differentiating between the pale colors and the darker ones.	I just looked at the key that explained what was what and observed the data looking for the correct answer.
Some questions made the reader make assumptions about the situation. Though overall the division by shape and color was very useful in separating out which variable was relevant to the question. Shape and color were equally useful and provided a consistent method to help in problem solving.	I had to go and reread some of the material and think things through. I used my mouse a lot to make a kind of square around the material in question, so I didn't get confused with the surrounding data. I also read aloud a bit and reread the question before finalizing my answer	I had to read, re-read and read over again to find what the question was asking and look at the graph
There was too many colors and shapes on display. It was hard to focus.	I compared some of the data to reality.	I took the advice I have always learned about interpreting graphs and numbers.
I memorized what two shapes and color belonged to a country and tried to block out the others when looking at the data. It was kind of confusing.	I had to first distinguish who represented what color had the Japanese and the Americans Each had 2 different colors 4 total colors It would have been easier	It was tough at the beginning to distinguish the colors but since the chart stayed the same the shapes and numbers began easier to read.
I always made sure to familiarize myself with the symbols and what they meant before I even looked at the question. I would constantly reference the legend of the symbols throughout answering the question. then I would analyze the graph right after reading the question.	I felt indifferent with the graph. I thought it was quite confusing but I was able to get through it and understand what I thought to be the best of what the Japanese and Americans learned in math and language arts.	Sit and look very hard at the screen to tell the differences between the two different test subjects
I looked at each year separately to get my answers.	I tried to identify the colors with the different countries but found it difficult because they were all circles. I could not really distinguish between the two	I had to look at the shades and rely mostly on the shapes to help find points. When questions asked for data on specific years it was easier to track down on the specific year.
I just looked closely at the display and read each question multiple times to ensure that I gave an appropriate answer	I used process of elimination	Memorize the colors and shapes for the Japanese and the US, for both Language arts and Math
I just used the graph and got my information from that. I was overwhelmed by the fact that the chart was so crowded and the questions were wordy.	Doubling checking the colors with the categories, re-reading the questions	Looking at the information presented on the graph.
I made sure I knew exactly what the key was describing for each graph. Knowing which shape and color represented what was very important to solving the problems.	I would look at each different dot color in the graph and draw a mental line graph through them to visualize the results.	Leaving the mouse arrow on the subject being discussed
More often than not, I simply looked at where the Japanese and American students performed in each subject and simply took approximate averages in my head depending on what I was required to answer.	I first memorized which group of students the dark and the light points represented. Then as I read each question I would identify what subject the question was focused on. From here I would analyze the data.	I would read the question then pick what shapes and values to look for and then from there choose an answer.

Appendix I: (Continued)

Group 1	Group 2	Group 3
I read the question, and then I looked at the key, and looked at the graph from there. I compared whatever the question was asking me to the color keys	When asked to look at data based upon several years, I would mentally block out data from all other years.	I tried to look as carefully at the screen as possible. The colors were a little difficult to read so I stared at the shapes carefully.
I tried to keep to the legend, committing the two color-coded countries to memory.	In the beginning I was confused. I don't think I took enough time to read the key properly and therefore may have answered incorrectly in the first 3 or 4 questions. It got clearer as I went along.	I looked at the shades of the gray and the shapes of the points to figure out the information.

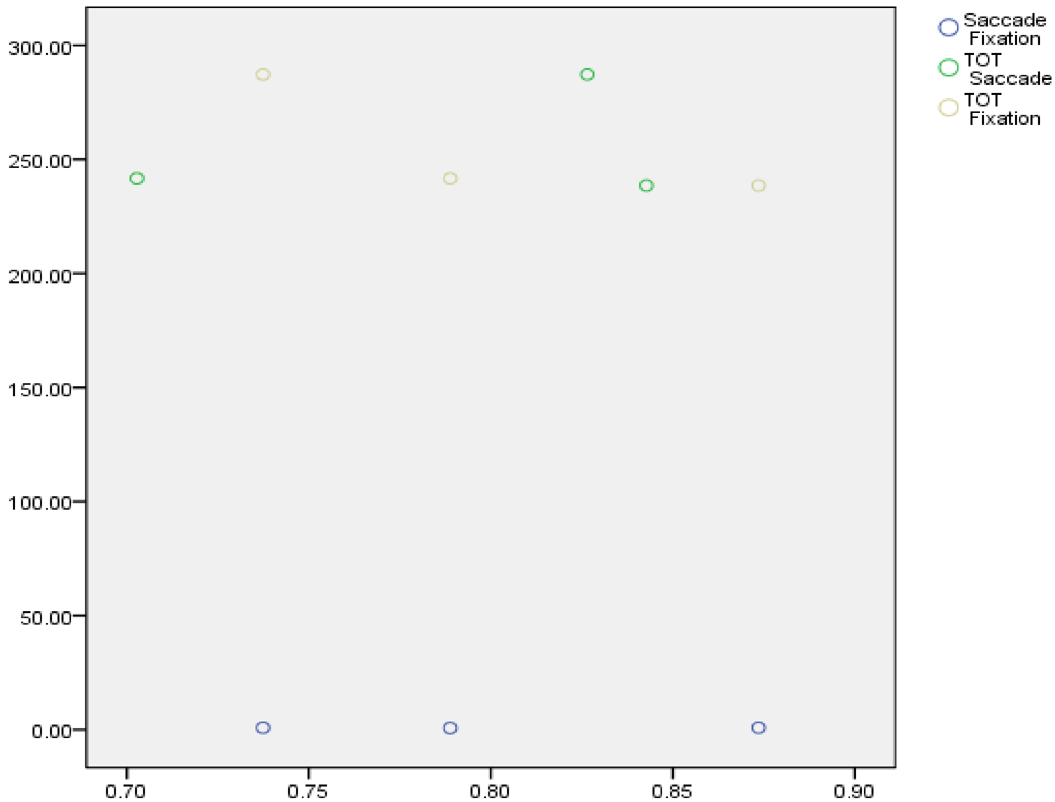
Appendix J: Eye Tracker Participant Data

Group	Participant	L1	L2	L3	TOT 1	TOT 2	TOT 3	Rating
1	1	4	4	2	109.93	63.69	287.36	4
	2	4	8	2	43.67	63.28	244.05	5
	3	4	8	2	70.76	72.27	160.73	4
	4	8	8	4	103.32	78.63	194.56	4
2	1	8	8	11	111.73	86.45	1154.12	4
	2	4	8	2	62.91	49.81	212.35	3
	3	8	4	1	94.35	45.6	196.1	3
	4	8	12	4	84.43	86.72	480.01	5
3	1	0	4	1	108.85	94.53	210.98	3
	2	4	0	2	78.43	74.63	150.79	2
	3	0	4	3	63.51	59.83	305.19	5
	4	4	12	10	37.94	62.08	373.7	4

Appendix K: Saccade and Fixation Scores by Participant by Group

	Group	Participant1	Participant2	Participant3	Participant4
Saccade Scores	1	0.726	0.139	0.203	2.238
	2	0.27	0.872	0.552	1.117
	3	1.76	0.288	0.705	0.618
Fixation Scores	Group	Participant1	Participant2	Participant3	Participant4
	1	0.6057	0.4356	0.4543	1.454
	2	0.8745	0.3245	0.6783	1.278
	3	0.7893	0.8745	0.8654	0.965

Appendix L: Scatter Plot Showing Correlations of Mean Saccade Scores, Fixation Scoresand Time-on-task Scores by Group



Appendix M: Pilot Study Results

Appendix P-1: Pilot Study Performance Scores and Time-on-task Scores by Group by Question Levels

Group	Score level 1	Score level 2	Score level 3	Total Time	TOT 1	TOT 2	TOT 3	Fixation	Saccade
T1	4	12	7	646.20	93.15	83.72	469.33	0.5035	0.9586
T1	4	12	12	562.60	135.77	57.91	368.92	0.5732	0.1941
T1	8	4	8	670.13	99.25	147.55	423.33	0.836	0.4084
T2	8	0	7	506.42	73.81	91.16	341.45	0.4189	1.6651
T2	8	12	12	491.32	56.38	66.02	368.92	0.4895	0.1883
T2	8	8	10	517.93	62.15	81.26	374.52	0.9142	1.3642
T3	8	4	6	364.62	57.87	41.0	265.75	0.3682	0.1821
T3	8	8	11	517.79	40.06	74.4	403.33	0.4775	0.3117
T3	4	8	10	528.87	56.68	93.23	378.56	1.3214	0.9862

Appendix P-2: Mean Performance Scores by Group by Question Levels

	T1	T2	T3
Score L1	5.3	8.0	6.7
Score L2	9.3	6.7	6.7
Score L3	9.0	9.7	9.0

Appendix P-3: Time-on-Task Scores by Group by Question Levels

	T1	T2	T3
TOT1	109.39	64.48	51.53
TOT2	96.39	81.2	69.74
TOT3	420.52	369.67	355.88

Appendix P-4: Pilot Study Repeated Measures Analysis Results

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Score level	30.30	2	15.15	1.57	0.25
S level * Group	23.71	4	5.93	0.61	0.66
TOT	535587.53	2	267793.76	201.01	.000
TOT * Group	1961.70	4	490.42	0.37	0.82

Appendix P-5: Strength and Relationship of Performance Scores and Time-on-task Scores from Pilot Study

	Score1	Score2	Score3	TOT1	TOT2	TOT3
Score1	1.000	.003	.435	.403	.478	.482
Score2	.003	1.000	.902	.917	.880	.878
Score3	.435	.902	1.000	.999*	.999*	.999*
TOT1	.403	.917	.999*	1.000	.996	.996
TOT2	.478	.880	.999*	.996	1.000	1.000**
TOT3	.482	.878	.999*	.996	1.000**	1.000

Appendix P-6: Average Saccade and Fixation Scores by Treatment Groups from Pilot Study

Treatment	Average Saccade Scores	Average Fixation Scores
T1	0.6375	0.3973
T2	0.6075	1.0725
T3	0.7223	0.4933

ABOUT THE AUTHOR

Chitra Subramaniam is a seasoned professional with several years of experience in educational leadership and administration, human performance improvement, distance education, educational measurement and evaluation, instructional design, and new media design**.** She has received numerous awards in education and industry including; Excellence in Academics, Who's Who in the World, Who's Who in America, and Who's Who Among America's College Teachers. Chitra co-owned an educational consulting business working with numerous clients including Pearson Education, Prentice Hall, Cengage Learning, IBM, Center for Medicare and Medicaid, several school districts, colleges and universities. She has co-written five books with topics ranging from specific software skills and technology integration to faculty development and student success strategies. She has presented at numerous educational conferences, written articles and been a guest speaker on education and technology integration at national conferences.

Chitra currently works as the Director of Educational Design and Technology at the American Society of Clinical Pathology.