

Utah State University

DigitalCommons@USU

All Graduate Theses and Dissertations

Graduate Studies

5-1999

The Perception and Analysis of Authentic Graphical Elements : An Empirical Study of Perceptual Skills and Analytical Tasks That Affect Graphicacy

Derek G. Borman
Utah State University

Follow this and additional works at: <https://digitalcommons.usu.edu/etd>

 Part of the [Psychology Commons](#)

Recommended Citation

Borman, Derek G., "The Perception and Analysis of Authentic Graphical Elements : An Empirical Study of Perceptual Skills and Analytical Tasks That Affect Graphicacy" (1999). *All Graduate Theses and Dissertations*. 6140.

<https://digitalcommons.usu.edu/etd/6140>

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



THE PERCEPTION AND ANALYSIS OF AUTHENTIC GRAPHS: AN EMPIRICAL STUDY OF
PERCEPTUAL SKILLS AND ANALYTICAL TASKS THAT AFFECT GRAPHICACY

by

Derek G. Borman

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Psychology

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1999

Copyright © Derek G. Borman 1999

All Rights Reserved

ABSTRACT

The Perception and Analysis of Authentic Graphical Elements: An Empirical Study of
Perceptual Skills and Analytical Tasks That Affect Graphicacy

by

Derek G. Borman, Doctor of Philosophy

Utah State University, 1999

Major Professor: Dr. Lani Van Dusen
Department: Psychology

In this study, the idea that authenticity should be integral to graphicacy research was advanced. That is, graphicacy researchers should use graphical stimuli that most closely approximate graphs as they might be encountered in the real world (i.e., in text books, newspapers, journals, etc.). It was contended that because of the lack of task authenticity and experimental control inherent in past studies of the analytical tasks and perceptual skills underlying graph reading, there was a need for further study of these issues. To this end, a 24-item graphicacy test was devised, such that key graphical elements and specifiers were more tightly controlled across test items and more closely approximated graphs as they might appear in a real-world setting.

An analysis of data revealed strong support for the independence of analytical tasks and basic perceptual skills, when single test items were considered. However, when the data from basic perceptual skills were collapsed across analytical tasks, there was moderate performance overlap among the different perceptual skills. When analytical tasks were collapsed across perceptual skills, there was little performance overlap among analytical tasks.

The other critical issue that was studied was the ranking of basic perceptual skills and analytical tasks according to the judgment accuracy associated with them. When all factors are taken into account, this study's ranking of basic perceptual skills was inconsistent with the predictions of the basic perceptual

skills model. Conversely, this study's ranking of analytical tasks was moderately supportive of the analytical tasks model.

In addition to (and in light of) other analyses performed and explanations rendered, alternative, more compact conceptions of analytical tasks and perceptual skills were advanced as well as the conclusion that when the levels of authenticity and experimental control are increased, the basic perceptual skills model may not predict graph reading in a satisfactory way.

(170 pages)

ACKNOWLEDGMENTS

There were numerous individuals who gave their time and talent to the completion of this project. I wish to express much needed thanks to the many individuals who helped to support me.

First, to Dr. Lani Van Dusen, I am unendingly grateful for the effort that she has invested in me. She has spent tireless hours molding me into a professional in the field of psychology. Her expertise and direction have served to make me a more conscientious researcher and teacher. I am most appreciative of her abilities as a mentor. I could not have asked for a kinder, more sensitive person to guide me through my doctoral studies. I have often thought of how I will always remember her as one of the truly great people that I have met in my life. Lani is an amazing instructor. She loves to laugh. She works harder than anyone else that I know, and she has a wonderful ability to cajole excellence out of those who might otherwise think that they are not up to the task. What I would really like to do is put the greatness of this woman in perspective. I think there are only three names—Dr. Albert Schweitzer, probably Mother Theresa of Calcutta, and Dr. Lani Van Dusen.

I am especially fortunate to have had such superlative individuals as my committee members. To Professors Susan Crowley, Xitao Fan, Tamara Ferguson, and Andrew Gibbons, I give sincere thanks for their valuable time, support, dedication, and expertise.

I would like to thank my wife and family for their support. Julie (my wife) has never stopped encouraging me throughout this whole process. I could not have finished this degree without her efforts to make me a better person. She is the most giving person that I know. My children (Allison and Christopher) have put a smile on my face, late at night when I could not find the energy to smile. My parents (Eugene and Marlene) have similarly been encouraging, always offering a kind word and wisdom when I asked for it. My brothers and sisters (Bret, Greg, Dawn, and Nicole) are scattered throughout the western United States, but I know who they are and they know who I am, every day. Finally, to my in-laws, I offer thanks for helping me to take life a little less seriously. This is what I have needed most over the past few years.

There are many other individuals who helped my family and me along the way. These individuals have been close friends and even business acquaintances. There are too many to name, but I am most appreciative of their efforts, nonetheless.

Most importantly, I am grateful to my Heavenly Father for giving me the strength and ability to complete a task such as this one.

Derek G. Borman

PREFACE

The boom in computer graphics capabilities has brought with it a line of research focusing on the construction of effective graphs. Although this vein of research is very much in its infancy (the vast majority of empirical work in this area has emerged only during the last one and a half decades), it is possible to identify prominent models attempting to account for *graphicacy*.

Inherent to the interpretation of any graph are two processes: perception and analysis (Carswell, 1992). Cleveland and McGill (1984) have proposed a *basic perceptual tasks* model that appears to be a robust predictor of graphicacy. However, this taxonomy is founded on a conceptualization that does not reflect authentic graph reading situations. In short, this model's predictive value is restricted to graphicacy tasks that are rarely encountered in any medium.

In a thorough review of the literature, Carswell (1992) identified four *analytical tasks* implicated in graph reading. However, there may be some conceptual and methodological problems underlying Carswell's taxonomy. In her review, Carswell compared and contrasted the results from a wide variety of studies and discussed her findings in terms of whether the general graphicacy literature supports the perceptual and/or analytical tasks models. The caveat herein is that her assertions failed to take into account the distinct possibility that an individual's ability to accurately make judgments about graphs appears to change from one setting to another—even when the difference between such settings seems trivial. Therefore, comparing and contrasting task performance across graphicacy experiments, as Carswell did, may have led to spurious support for her model, as well as Cleveland and McGill's (1984) model.

To their credit, Carswell's (1992) analytical tasks model and Cleveland and McGill's (1984) basic perceptual skills model have identified what seem to be the fundamental cognitive issues related to graph reading, and these models provide a common language through which this somewhat fractionated domain may burgeon and evolve. Therefore, the basic perceptual skills and analytical tasks models should continue to be investigated.

This research has been conducted as an attempt to further validate prominent models within the graphicacy literature. Further, the graphicacy literature is somewhat fractionated in its various approaches to graph reading. The graphicacy literature is replete with experimental designs that do not seem to build on one another. This paper provides for a consolidation of graphicacy theories and direction for future studies in this field of interest.

CONTENTS

	Page
ABSTRACT	iii
ACKNOWLEDGMENTS	v
PREFACE	vii
LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF DEFINITIONS	xiv
REVIEW OF THE LITERATURE	1
Introduction to the Literature	1
The Birth of the Graphics Juggernaut	1
The Need for Graphs	2
Measurement of Graphicacy	3
The Place of Taxonomies in Graphicacy Research	4
Task Dependence: The Importance of Universal Taxonomies for Graphicacy Research	4
The Elusive Data-Ink Ratio: Arguments Against the Continued Use of Tufte's Data-Ink Model	5
A Taxonomy for Basic Perceptual Skills	6
Authenticity and the Basic Perceptual Skills Model	9
Analytical Tasks and Graph Reading	11
Shortcomings in Carswell's Metanalysis	14
Summary	16
THE STUDY	17
Purpose of the Study	17
Experimental Design and Methodology	17
Participants	17
Design	18
Materials	18
Procedure	20
Pilot Testing	22
Data Entry and Analysis	22

	x
RESULTS.....	23
Exclusion of Data Related to the Area Task Involving Point Reading.....	23
Intercorrelations.....	24
Analysis Rationale.....	24
Analysis of Individual Items.....	24
Basic Perceptual Skills Collapsed Across Analytical Tasks.....	25
Analytical Tasks Collapsed Across Perceptual Skills.....	27
Rank Ordering of the Different Types of Tasks.....	28
Analysis Rationale.....	28
Perceptual Skills.....	29
Perceptual skills collapsed across analytical tasks.....	29
Perceptual skills ranked within each type of analytical task.....	30
Analytical Tasks.....	32
Explication of the difficulty associated with different types of analysis.....	32
Analytical tasks collapsed across perceptual skills.....	33
Analytical tasks ranked within each perceptual skill.....	34
Over- and Under-100% Estimation.....	35
Gender-Based Performance Differences.....	35
Basic Perceptual Skills Collapsed Across Analytical Tasks.....	35
Analytical Tasks Collapsed Across Basic Perceptual Skills.....	38
ACT Composite, GPA, and Performance.....	40
Test-Taking Strategy for Synthesis Tasks.....	41
DISCUSSION.....	43
Intercorrelations.....	43
Uniqueness of Individual Items and Underlying Cognitive Processes.....	43
Perceptual Skills Collapsed Across Analytical Tasks: Overlapping Perceptual Processes.....	43
Analytical Tasks Collapsed Across Basic Perceptual Skills: Evidence of Unique Analyses.....	46
Rank Ordering of Basic Perceptual Skills: Mixed Findings.....	48
Collapsing Data Across Analytical Tasks: Support for the Basic Perceptual Skills Model.....	48
Ranking of Perceptual Skills Within Analytical Tasks: Failure to Support the Perceptual Skills Model.....	50
Rank Ordering of Analytical Tasks: Mixed Support for Carswell's Taxonomy.....	54

Judgments of Proportions Above and Below 100%	57
Explication of Male and Female Graph Reading Equivalencies	58
General ACT Scores, GPA, and Graph Reading	62
CONCLUSION	64
Summary	64
Limitations of the Study	65
Implications of This Study	67
Directions for Future Research	70
A CAUTIONARY FABLE FOR USERS OF GRAPHS	72
REFERENCES	75
APPENDICES	81
Appendix A. Statement of Informed Consent	82
Appendix B. Experiment Script	84
Appendix C. Answer Sheet and Pilot Checklists	96
Appendix D. Answers to Test Items and Test Items	100
Appendix E. Statistical Analyses	137
Appendix F. Pilot Testing: Methodology and Findings	145
CURRICULUM VITAE	150

LIST OF TABLES

Table	Page
1 Basic Perceptual Skills Grouped Within Each Analytical Task	19
2 Intercorrelations for GPA, ACT, and Basic Perceptual Skills Collapsed Across Analytical Tasks	26
3 Intercorrelations for GPA, ACT, and Analytical Tasks Collapsed Across Basic Perceptual Skills	28
4 Mean Absolute Error Scores for Various Tasks	31
5 Analytical Tasks Ranked for Difficulty Within Each Perceptual Skill	34
6 Descriptive Statistics for Mean Absolute Error for Basic Perceptual Skills Broken Down by Gender	36
7 Descriptive Statistics for Mean Absolute Error for Analytical Tasks Broken Down by Gender	38
8 Correlations Between Original Synthesis Tasks and Test-Taking Strategy Employed for Such Items	42
9 Mean Absolute Error Scores for Perceptual Skills Requiring Judgments in Excess of 100%	52
10 Mean Absolute Error Scores for Perceptual Skills Requiring Judgments of Under 100%	53
11 Mean Absolute Error Scores for Analytical Tasks Requiring Judgments in Excess of 100%	54
12 Mean Absolute Error Scores for Analytical Tasks Requiring Judgments of Under 100%	55
E.1 Descriptive Statistics for Absolute Error Ranked in Accordance with the Judgment Accuracy Associated with Each Item	138
E.2 Straight-Error Descriptive Statistics for Test Items	140
E.3 Intercorrelations Between Test Items	142

LIST OF FIGURES

Figure		Page
1	Sample graphs used in Cleveland and McGill's (1986) experiment	8
2	Line and bar graphs depicting trend data relating to a hypothetical intervention designed to improve fifth graders' reading.....	13
3	Mean absolute error for perceptual skills collapsed across analytical tasks	30
4	Mean absolute error scores for analytical tasks collapsed across perceptual skills	33
5	Mean absolute error for basic perceptual skills broken down by gender.....	37
6	Mean absolute error for analytical tasks broken down by gender.....	39
7	Depiction of more accurate slope line	46

LIST OF DEFINITIONS

1. *Graphicacy*: The ability to understand and interpret graphical representations of quantitative data.
This is a general definition of graphicacy. As will be discussed in the body of this paper, graphicacy can be interpreted as pertaining to a range of tasks. In her meta-analysis of graphicacy studies, Carswell (1992) included a broad range of tasks, many of which seemed to hinge more on decision making than the perceptual assessment of visual spatial stimuli. This paper does not hold tasks in which there are diverse decision-making and nongraphical elements to be graph reading tasks. An example of such a nongraphicacy task can be found in Barnett and Wickens' (1988) study in which subjects were presented with information such as fuel level, engine temperature, oil pressure, enemy intent, and pilot fatigue and then were asked to make decisions about whether to continue a mission in an aircraft based on the value of these different pieces of information. Although many of these data were depicted spatially, many of the data were not. This type of task is distinguished from a graphicacy task and is referred to as a multicue, decision-making task. Additionally, the issue of authenticity is an issue that figures into the definition of graphicacy advanced in this paper. The author maintains that in the creation of a graph reading task, it is not enough to construct visual stimuli that merely contain some of the components that might be found in a traditional graph reading task. Instead, graphicacy studies should be born in real-world settings. That is, graphical stimuli, presented in experiments, should closely approximate appropriate elements of graph reading as it might occur in a student's text book, manager's report, or daily newspaper. Taken as a whole, all of the foregoing ideas indicate that the definition of graphicacy advanced in this paper is perhaps narrower and somewhat different from those definitions that have been advanced in important work by researchers such as Carswell (1992) and Cleveland and McGill (1984, 1986).
2. *Specifiers*: The parts of a graph that strictly convey the quantitative aspects of a graph. Specifiers may be bars, lines, angles, slopes, and so forth.

3. *Elements*: All parts of a graph other than specifiers. Examples of elements include axes, axis labels, titles, legends, and so forth.
4. *Basic perceptual skills model*: Cleveland and McGill's (1984) model ranks graphical elements such as length, area, angle, and slope in terms of how accurately individuals can make perceptual judgments about such specifiers. This model draws its support from psychophysics literature.
5. *Analytical tasks model*: Carswell's (1992) conception of analytical tasks includes different levels of analysis, including: point reading, local comparisons, global comparisons, and synthesis. These different types of analytical tasks can be distinguished on the basis of the number of specifiers that must be attended to and whether actually presented and/or cognitively imaged standards must be judged.
6. *Data-ink ratio*: Tufte's (1983) ideas about how to construct a good graph are embodied within the notion that a graph's effectiveness can be inferred from the ratio of ink that actually conveys quantitative information to ink that conveys no such information. A graph has a favorable data-ink ratio when data (i.e., specifier) ink is more plentiful than superfluous (i.e., element) ink. An unfavorable data-ink ratio leads to clutter and inaccurate graphical judgments.

REVIEW OF THE LITERATURE

Introduction to the Literature

The Birth of the Graphics Juggernaut

The computer revolution has had a mammoth impact on virtually every aspect of everyday life. From the way that tasks are performed to the quickness with which such tasks can be performed, computer technology continues to enhance our abilities. In particular, the graphics boom that has accompanied dramatic increases in silicon computing speed and power (among other things) has led to an outpouring of printed images, the diversity of which appears to be limited only by the limits of creativity.

In this day and age, the power is in the people. The Orwellian notion that "Big Brother" (Orwell, 1949), or a powerful elite, would possess a monopolizing omnipresence in our modern society was apparently a bit off the mark. It seems that information is controlled and processed not by a handful of individuals, but by anyone who vigilantly observes the world about or maintains the resolve to make her voice heard.

Regarding the dissemination of information, researchers and lay persons alike are exposed to numerous forms of graphs on a daily basis. Further, virtually anyone in a modernized society can complement her presentations, publications, and so forth with graphs tailored to appeal to her own personal preferences and intuitive proclivities (e.g., color, type of graph, line thickness, addition of a third dimension). But which of these penchants makes for the most efficient interpretation and understanding of graphs? This question is only now being formulated; the answer to it lies somewhere in the distance. It seems that once again, the growth of a technological innovation has outdistanced society's ability to comprehend it.

The current wave of technological gadgetry has heightened researchers' awareness of the need to empirically investigate the most efficient means by which quantitative information can be visually displayed. However, the history of graphs did not emanate from the garage of a computer prodigy two and a half decades ago. Instead, graphs were first introduced by William Playfair back in the late 1700s (as

cited in Cleveland & McGill, 1984). His use of graphs was apparently the first serious attempt to provide a visual supplement to quantitative information.

It has taken nearly two centuries for researchers to recognize the importance of studying graphical communication. Even at that, empirical studies have often been overshadowed by prominent but intuitively based texts (i.e., Enrick, 1972; Kosslyn, 1994; Schmid, 1983; Tufte, 1983; Tukey, 1977). The result of this opening is a research domain that finds itself in ever increasing states of fractionation. All is not lost, however, as there have been attempts to develop universal taxonomies that can be applied to the entire spectrum of graph reading tasks.

Before wading into the literature concerned with graphicacy taxonomies, however, there are several issues that must first be broached—the first of which is an answer to the question: Why do we need graphs?

The Need for Graphs

As Tukey (1990) contended, graphs are intended to be spatial appendages to quantitative information. The whole appeal of graphs lies in their ability to depict quantitative information in a spatial format. This is particularly attractive to human beings for whom immediate perception and memory are enhanced by the presence of integrated verbal and spatial stimuli.

Researchers (i.e., Feliciano, Powers, & Kearn, 1963; Legge, Gu, & Luebker, 1989; Lewendowsky & Spence, 1990; Sparrow, 1989; Washburne, 1927) are in general agreement about the benefits of supplementing tabular data presentations (e.g., numerical tables like spreadsheets) with corresponding spatial information. It is generally held that data in tabular presentations are judged with more accuracy (than are quantitative data presented with a spatial display of one form or another) when such judgments involve simple judgments (i.e., judgments that do not involve comparison or extrapolation) concerning few data points. However, the addition of spatial stimuli statistically significantly improves performance for more complex analytical tasks involving judgments made about greater numbers of data points. These performance trends have been demonstrated when performance with tabular formats has been compared to

performance with bar charts, line charts, pie charts, and scatterplots, in conjunction with estimations of percentages, slopes, and areas. Inasmuch as the addition of spatial data displays to tabular displays improves performance in more complex graphicacy tasks and does not significantly degrade performance in relatively simpler tasks, the value of graphs cannot be ignored.

The importance of graphs can also be discussed in terms of their effects on encoding and recall. Graphs provide visual information to supplement quantitative information, and these two forms of information seem to be distinct from one another. This assertion can be made in light of Paivio's (1975) *dual code* theory, a major tenet of which holds that visual images and verbal information are processed and stored via different cognitive codes. A fundamental outgrowth of this seemingly natural process is that utilizing more than one code to portray the same information increases the likelihood of recalling such information (see Paivio, 1986, for a review of the literature in this area). This is important because a graph's efficiency should not only be measured by how well it is initially perceived and judged, but also by how accurately its contents can be conveyed from one individual to another, once the graph is no longer accessible.

For many individuals, it is simply not very informative to know, for example, that 28% of children who have learned to read through a phonetics-based approach and 31% of children who have learned to read through a whole-word approach are reading at or above grade level. But when such information is accompanied by bars that proportionally represent these values, these data may become more comprehensible and even more memorable.

Measurement of Graphicacy

In graphicacy research, judgment error is the dependent variable most often studied. That is, researchers are primarily concerned with the difference between an individual's perception and the actual, spatial arrangement of graphical stimuli. For example, in a typical experiment a subject might be asked to estimate the difference between the lengths of two bars. If the first bar is two inches in length and the second bar is four, then the second bar is twice the length of the first. However, if a subject estimates that

the length of the second bar is 225% of the length of the first bar, judgment error has entered into the process. Commonly, a discrepancy such as this would translate into a datum value of +25. That is, the judgment of 225% overestimates the actual proportional relationship between the bars by 25 percentage points. If absolute error is the primary concern, positive and negative weightings are eliminated.

The Place of Taxonomies in Graphicacy Research

Task Dependence: The Importance of Universal Taxonomies for Graphicacy Research

Croxton and Stryker's (1927) replication of Eells' (1926) study added the most to our current understanding of the task-dependent nature of graphicacy.¹ Croxton and Stryker found that the efficacy of graphs seems to be dependent on a number of factors, the most compelling of which are related to the values of the graphs themselves. In their study, subjects simply looked at bar and pie graphs and estimated the disparity between lengths or proportions, respectively. Croxton and Stryker found that when two values were compared within the same graph, a 12%-88% split (between the values depicted in the graph) led to an average estimate error of 4.2% for bar graphs and 5.2% for pie charts. A 25%-75% split led to average error estimates of 5.3% for bars and .6% for pies.

This error estimate was for a pie chart that was divided by lines at 0 and 90 degrees. Perhaps, more importantly, this error estimate for pies increased to 3.5% when the same split was used, but the chart was divided by lines at 135 and 225 degrees (see Croxton and Stryker [1927] and Kruskal [1982] for more discussion). It seems that fine shifts in the nature of the pie-reading task, will dramatically affect the precision with which the graphs could be interpreted. Croxton and Stryker's work served as an important reminder that graphicacy is a task-dependent phenomenon.

¹ Eells (1926) provided the first substantive machinations related to graphic efficiency in studies focusing exclusively on the benefits of pie and bar charts. Initially, Eells contended that the pie chart was superior to the horizontal bar chart. This position was subsequently challenged by Croxton (1927), Croxton and Stryker (1927) and von Huhn (1927). This issue was later revisited by other researchers (e.g., Croxton & Stein, 1932; Culbertson & Powers, 1959; Peterson & Schramm, 1955). Although, as Spence and Lewandowsky (1991) claimed, the above research may have done little to help clarify the debate over circles and bars, the issue of judgment sensitivity to task variations seemed to be the important factor underlying the entire controversy.

Still other examples serve to further complicate the observations made by Croxton and Stryker (1927). Spence (1990) found that bars, boxes, pies, and cylinders are all judged with relatively equal accuracy except when demanding time constraints or stressful components were introduced as part of the task. Simkin and Hastie (1987) have discussed the 0, 90, 180, and 270 degree positions of a circle as natural, perceptual anchoring points and that pie charts comprising lines that run close to these positions lead to more accurate graph interpretations. Finally, there are Carter's (1947) findings which demonstrate that as the length of a graph's abscissa increases, judgment accuracy decreases.

In short, graphicacy varies from one task to another and from one graph to another. This makes it extremely difficult to compare findings from different studies and to develop a common framework within which to do so. What is needed are theories that address common aspects of graph reading tasks. In recent years, there have been several notable efforts to develop theories that could provide a common language for the findings from a broad range of graphicacy research efforts. The most influential of these theories are discussed below.

The Elusive Data-Ink Ratio: Arguments
Against the Continued Use of Tufte's
Data-Ink Model

Tufte (1983) began his assault on superfluous graphical specifiers (in keeping with the terminology introduced by Kosslyn [1989], the visual elements of a graph will be referred to as *specifiers*) with the following statement: "Occasionally artfulness of design makes a graphic worthy of the Museum of Modern Art, but essentially statistical graphics are instruments to help people reason about quantitative information" (p. 91). This statement conveys the conventional wisdom that an excess of visual properties in any graph is likely to distract attention from and obfuscate the intended message contained within the graph. In other words, any ink (contained in a graph) that does not convey data represents a superfluous use of ink. Tufte contended that such excesses hinder judgment accuracy.

The problem with Tufte's (1983) outwardly intuitive approach is that it does not hold up under close scrutiny. There are numerous studies that call into question the underlying assumptions of the data-

ink principle. For example, the data-ink rule would hold that the border surrounding a graph represents a superfluous use of ink. However, Baird and Noma (1978) and Gregory (1966) have asserted that a frame increases the accuracy with which one is able to judge bar lengths. When a frame is present, the distance between it and the tops of the bars serves as an additional distance cue to improve the accuracy with which the bars are perceived. As appealing as Tufte's (1983) ideas were initially, ultimately they are too imprecise and without empirical merit to provide a useful and comprehensive framework for the continued study of graphicacy.

A Taxonomy for Basic Perceptual Skills

Authors such as Barnett and Wickens (1988), Bertin (1981), Cleveland (1985), Kosslyn (1994), Seidler and Wickens (1992), and Tufte (1983) have done a great deal of debating (about issues like excessive decoration and proximity of text to graph) without the support of empirical evidence. Because of this initial trend toward intuitively based conceptualizations of graphic efficiency, an empirically grounded, shared language is difficult to find in the graphicacy literature.

Cleveland and McGill (1984), reacting to the confusion that had been and in anticipation of battles to come, proposed a concise taxonomy of basic, perceptual, graphical tasks. They proposed that there are 10 elementary, perceptual skills in graph reading. These skills were ranked in accordance with the perceptual accuracy that is associated with them. Certain skills were grouped together because they are associated with a similar level of perceptual accuracy: (a) *position along a common aligned scale*, (b) *positions along identical, nonaligned scales*, (c) *length, direction, angle* (d) *area*, (e) *volume, curvature*, and (f) *shading*, and *color saturation*. Cleveland and McGill (1986) have asserted that most judgments can be accounted for by the following, which are listed in the order in which they are ranked accordingly: position along a common scale, positions along identical nonaligned scales, length, slope, angle, and area.

Baird (1970) gave an excellent review of many experiments, across which, the above rankings for length, area, and volume seem to fit. Generally speaking, length judgments tend to be more accurate than

area judgments that, in turn, tend to be more accurate than volume judgments. The specifiers comprising most commonly used graphs can be thought of in these terms. The rationale for listing these skills in the above order is based on Stevens' (1975)² psychophysics law.

In their experimental task, Cleveland and McGill (1986) exposed subjects to different graphs, each of which comprised four graphical objects (see Figure 1). In each graph, the upper left object was the standard. Subjects attempted to determine what percentage (i.e., size, length, slope, angle) each of the other objects was of the size of the standard. They found that judgments relating to position along a common aligned scale were the least difficult of the perceptual skills. Why should this skill be ranked ahead of length, as it has already been determined that length estimation is relatively free of bias? Put simply, a judgment involving direct point comparisons (see box 1 in Figure 1) is easier because each data point along the common scale functions as a measurement cue that can be utilized to estimate proportions across the scale. Similarly, when one is comparing data points on nonaligned, common metric scales (see box 2 in Figure 1), there are additional visual cues that make this type of judgment easier than determining the length of a line (or even comparing the lengths of different lines) without visual cues such as axes or frames. The preceding rationale provides part of the argument for the ranking of the angle and slope judgments as more difficult than the judgment of length. Slopes were determined to be more difficult to judge, as they involve the estimation of angle, in spite of the absence of a reference cue, like an axis or a frame (see box 4 in Figure 1). Angles, when perceived by themselves, lack the visual, comparison cues of frames or X and Y axes, as well. The fourth box in Figure 1 illustrates this idea via a graph containing nonaligned angles, to which subjects were exposed in Cleveland and McGill's (1986) study. Some, including Stevens (1975), have suggested that angle estimation may be even more complicated than it appears. One may very well judge the angle of two connecting lines by mentally imaging a third line

² In judging physical aspects (e.g., weight, distance, loudness) of perceptions, Stevens' (1975) power law of psychophysics holds that if p is the perceived magnitude, a is the actual magnitude and k is a constant value, then p is related to a by $p = ka^\alpha$. If a_1 and a_2 are two such magnitudes and p_1 and p_2 are corresponding perceived values, then $p_1/p_2 = (a_1/a_2)^\alpha$. So, when $\alpha = 1$, the perceived scale is the same as the actual physical scale. In terms of Cleveland and McGill's (1986) most common perceptual skills, the value of α is nearest to 1 for judgments involving positions along a common scale and greatest for judgments involving area.

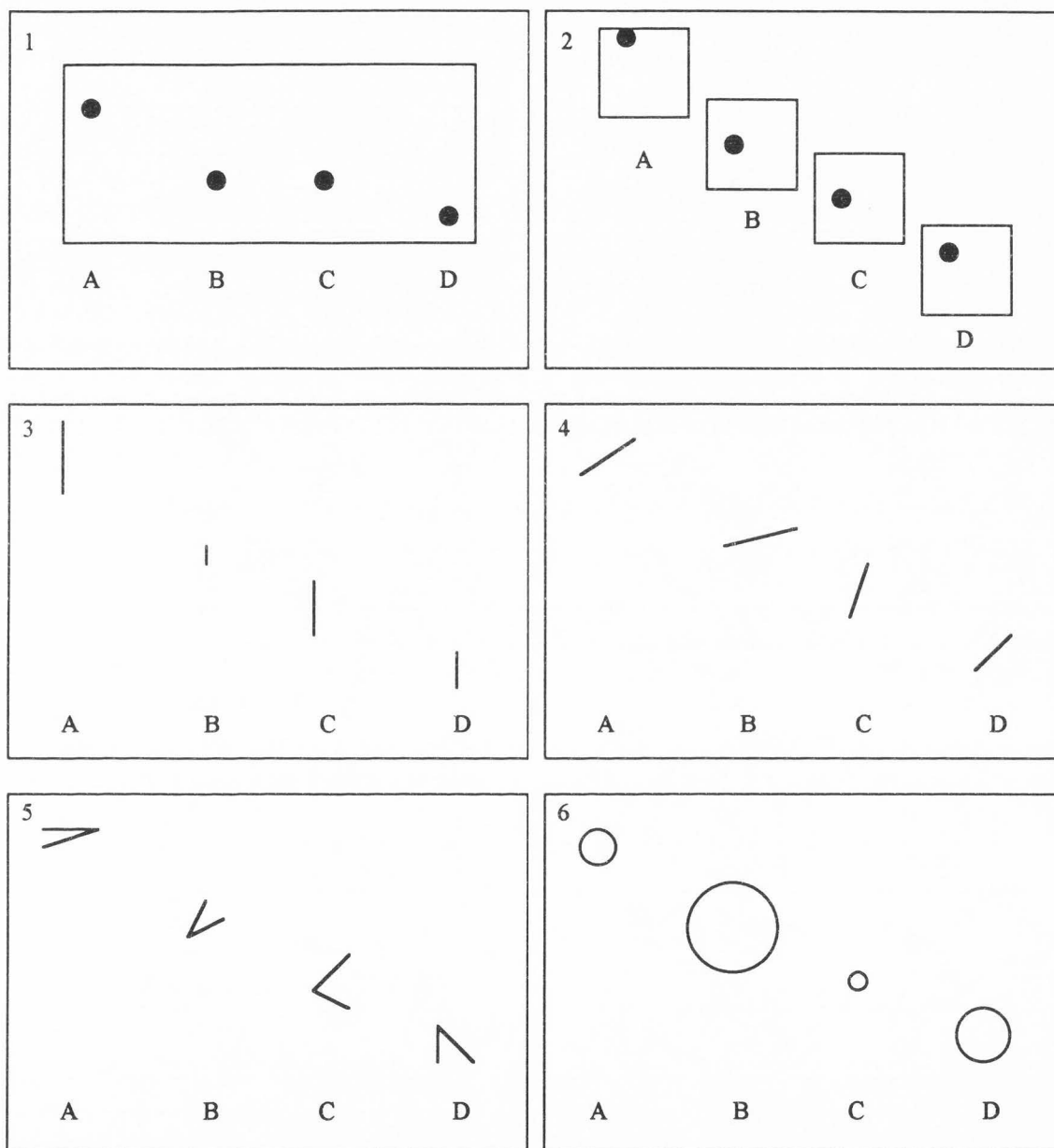


Figure 1. Sample graphs used in Cleveland and McGill's (1986) experiment.

Note. These graphs are designed to test what Cleveland and McGill (1986) asserted are the most prevalent perceptual abilities required in traditional graph reading situations. From left to right and top to bottom, these abilities or skills are: position along a common scale, position along identical nonaligned scales, length, slope, angle, and area.

joining the first two lines. In this way, a triangle is fabricated, yielding two more comparison cues (i.e., the additional imaged angles) to improve one's accuracy in judging the size of the original angle. This process sounds very similar to that of area estimation. If such operations are performed during angle judgment, then it is no surprise that angle estimation is ranked just ahead of area estimation, in terms of judgment accuracy.

Authenticity and the Basic Perceptual Skills Model

There is little to argue about concerning the psychophysical basis for Cleveland and McGill's (1984, 1986) rankings. However, the graphical manifestation of their conceptualization is hardly a template for day-to-day graphical analyses. Although Cleveland and McGill contend that they have identified the six perceptual skills that are the most used in graph reading, the graphs in Figure 1, for the most part, do not resemble graphs that one would encounter ordinarily in one's reading. To Cleveland and McGill's credit, their depiction of aligned and nonaligned points (see boxes 1 and 2 in Figure 1) was a good attempt to eliminate the confound of length, which is ranked further down the basic skills list (Cleveland and McGill refer to such graphs as *dot charts*). In fact, for judging line graphs (like the one depicted in box 1) people tend to rely less on the abscissa than on the ordinate axis and the disparity amongst the points (Teghtsoonian, 1965), which is a possible indication that vertical lengths are not being estimated.

The problems with Cleveland and McGill's (1986) study begin with their conceptualization of nonaligned scales and their use of frames. In box 2 of Figure 1, nonaligned graphs are presented diagonally. This presents perceptual judgment problems for using axes or frames (these two specifiers are one and the same in box 2) as reference cues, because the frames overlap. This might be an acceptable way to test for this perceptual skill if this was how such graphs commonly appeared in text. But this is not how they appear in text. Almost invariably, nonaligned graphs of any type appear one on top of another. Nothing in Cleveland and McGill's (1984) theory precludes placing one nonaligned graph on top of another as such placement preserves the nonalignment of different abscissas. Further, this is how

nonaligned graphs are displayed in text. It is possible that such positioning would lead to better judgment accuracy for this perceptual skill as it has been conceptualized.

Another question relating to authenticity is this: Why were the individual graphs (in boxes 3, 4, and 5 of Figure 1) not enclosed in individual frames? The answer is that inasmuch as Cleveland and McGill (1986) were attempting to isolate perceptual skills, it was imperative that frames and axes be absent from these particular graphs, as frames and axes provide additional perceptual cues for judging distances and angles. However, given Cleveland and McGill's claim that their basic perceptual skills model consists of those perceptual judgments that are most prevalent in common graph reading, the lack of frames and axes is inconsistent with their theoretical premise. This omission would be appropriate if such specifiers typically were absent from graphs one might encounter. But this is highly atypical. Frames and axes are virtually always a part of bar or line graphs.

If the basic perceptual skills taxonomy is truly going to be a model for the perception of authentic graphs (and not just perception of visual stimuli), then frames and axes must be integrated appropriately with the visual stimuli created by Cleveland and McGill (1984). For example, in boxes 3, 4, and 5 of Figure 1, each line or slope or angle should be surrounded by a frame and/or supported by axes. The most noteworthy issue in this alteration of Cleveland and McGill's ideas is that such a change is likely to diminish judgment differences between the different type of perceptual skills. This is a modest theoretical departure from Cleveland and McGill's original conceptualization. However, it is the premise of this dissertation that this amendment represents a departure in the direction of authenticity and a more practical taxonomy.

One last issue remains. In light of the issue of anchoring, studied by Croxton and Stryker (1927) and Simkin and Hastie (1987), the indiscriminate rotation of the angles in box 5 makes it impossible to draw any cohesive conclusions. As Simkin and Hastie noted, angles that are rotated to a certain station (like the 3 or 6 o'clock positions of a circle) are easier to judge. It is possible that angle judgment is ranked as moderately difficult because Cleveland and McGill (1986) rotated them variably around the helpful anchoring positions. Again, if this was how angles appeared in text, this approach would be

adequate. But when angles (i.e., angles aside from those contained in a pie chart, where area becomes another perceptual issue) are portrayed in text, typically the lower line in the angle extends along a 90 degree plane. In light of this reality and anchoring theories, angles being compared should have a common rotation. So long as the angles are not aligned on the same abscissa, such an approach is only a small departure (in the direction of authenticity) from Cleveland and McGill's original conceptualization.

These amendments to the basic perceptual skills model may make it more authentic and reflexively, better suited to predict performance in authentic graph reading situations. This is an important issue in need of testing.

Analytical Tasks and Graph Reading

Aside from perceptual stimuli, there are several other important issues that immediately confront the user of a graph. One of these issues pertains to the conclusions to be drawn from the graph. The complexity of the information portrayed in a graph can range from simple to extremely complex. Similarly, the corresponding range of assessment tasks can vary significantly. Several authors (e.g., Bertin, 1983; Lohse, Walker, Biolsi, & Rueter, 1991; MacDonald-Ross, 1977; Washburne, 1927; Wickens, 1989) have attempted to provide more systematic classifications of the various tasks, but the resulting taxonomies have been either too narrow or vague to be of practical use. For example, Bertin discussed only issues related to the identification of single points and simple comparisons. Lohse et al. developed a taxonomy of graphics categories consisting of icons, graphs and tables, maps, and network/flow charts. Although this study was compelling, from an analytical standpoint, such a classification scheme is guilty of comprising categories that are difficult to distinguish from one another, as they all consist of overlapping visual specifiers. Conversely, Carswell (1992) constructed a four-way classification of analytical tasks that seems to integrate clearly and distinctly the most important aspects of previous categorization schemes.

Suppose you have a bar graph composed of two bars depicting two mean scores. The first bar represents fifth graders' reading skill before some sort of educational intervention and the second bar

represents reading skill after 2 months of this intervention. The values ascribed to these bars are not important to the present discussion. What is important is the relative simplicity of the described graph. How many different and useful assessments can be made from such a graph? For all practical purposes, only three—identifications of the values of each bar and a scaled comparison between the two bars are depicted. In Carswell's (1992) terms, these tasks would be referred to as *point reading* and *local comparisons*. Such tasks require that attention be focused on one or two data points that are actually on the graph being judged.

Point reading and local comparisons are tasks that are not restricted to simple graphs. Such tasks may certainly be part of reading more complex graphs. Figure 2 shows a line graph and a bar graph depicting the same quantitative information. Even when each graph contains more than two data points, one can perform elemental point reading and local comparison tasks within these graphs. For instance, a local comparison question concerning the above graphs might be, "How much greater is the value at 4 months than the value at 3 months?" A point reading question might be, "What is the value at 2 months?"

Of course, simple judgments about a complex graph are not made with as much accuracy as are simple judgments about graphs containing fewer data points. Washburne (1927) and Croxton and Stryker (1927) were some of the first researchers to note that increasingly complex graphs lead to decreases in judgment accuracy. Similarly, when judgments are being made about complex graphs, it is often necessary to compare data points that are on opposite sides of the graph. This is noteworthy, because greater distance between data points being compared leads to judgments of less disparity between such points (Cleveland & McGill, 1986; Hollands & Spence, 1992). Judgment in such a scenario is complicated even more by the presence of numerous data points in between the two points being compared. Even judgments of pie graphs are affected adversely as the number and proximity of segments is increased (for more complete discussion, see Carswell & Wickens, 1987; Goettl, Kramer, & Wickens, 1991; Wickens & Andre, 1990).

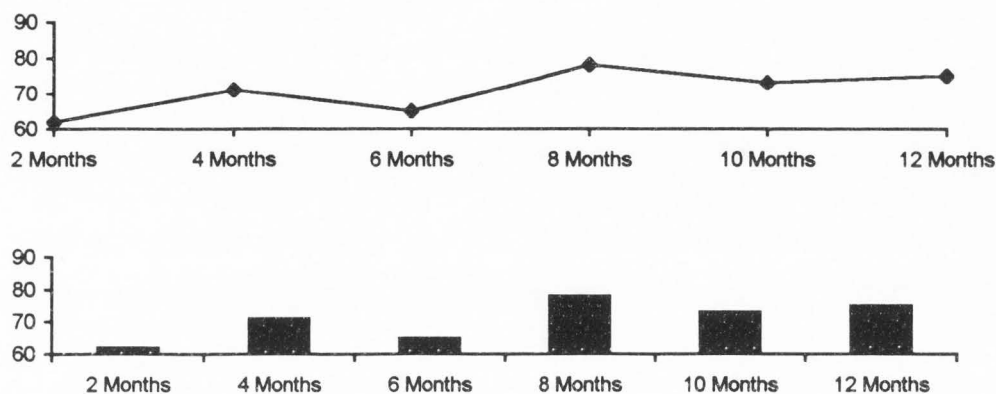


Figure 2. Line and bar graphs depicting trend data relating to a hypothetical intervention designed to improve fifth graders' reading.

Note. The above data reflect sample mean scores over the course of a year's intervention.

In addition to simple tasks, a graph depicting a greater number of data points allows for the performance of more complex tasks. In light of the data in Figure 2, it would be fair to ask, "Is the average for months 10 and 12 greater than the value for month 8?" or "What value would you predict for month 14?" Carswell (1992) would label the first question as a *global comparison* question, because it involves a single comparison between more than two data points. Another example of a global comparison task would be to determine whether the sum of the proportional relationship between the sum of months 10 and 12 and month 8. The latter question relates to a task involving *synthesis*. That is, such a task requires a judgment based on a subjective cognitive or mental standard, extrapolated from the perception of relationships and values that are actually on a graph. This type of analytical task is characteristic of trend prediction.

Just as Cleveland and McGill's (1984) taxonomy offers a construction philosophy for the perceptual specifiers of a graph, so too does Carswell's (1992) model provide a comprehensive, common language for discussing the analytical, judgment tasks associated with typical graph reading. In spite of the promise that these models offer, there are issues that have yet to be resolved.

Shortcomings in Carswell's Metanalysis

Carswell (1992) examined 39 experiments to determine whether Tufte's (1983) data-ink principle or Cleveland and McGill's (1984) basic perceptual skills model provides the better explanation of graph reading. Carswell found interactions between type of analytical task and type of basic perceptual skill; she concluded that the "basic [perceptual] tasks model is most successful at predicting performance in local comparison and point reading tasks" (p. 550), and it is least successful for predicting performance in synthesis tasks. However, before these findings are embraced wholeheartedly, the research from which they gain support must be critiqued further. There were several conceptual and methodological issues in Carswell's study that must be improved upon before the basic perceptual skills and analytical tasks models should be considered as the most useful tools for predicting graphicacy in authentic graph reading tasks.

First, in Carswell's (1992) study there was the issue of the variability in the distribution of the effect sizes of the experiments analyzed. Carswell acknowledged the statistically significant heterogeneity of this distribution but did not go into any detail about how this could have been alleviated or how it affected her findings. This sampling problem is most likely due to the broad range of experiments analyzed and the task-dependent, variability that is inherent to different graphicacy tasks. Much of this problem could be eliminated by a single study that provides for common graphical specifiers across different perceptual skills and analytical tasks. For example, if the number of bars or segments in a series of graphs was held constant, this could help to eliminate some of the variability that almost certainly resulted from uncontrolled factors such as the number of value specifiers comprising the graphs.

Second, Cleveland and McGill's (1984) taxonomy was developed to predict behavior in graphicacy tasks. But 31% (i.e., Barnett & Wickens, 1988; Coury, Boulette, & Smith, 1989; Goettl et al., 1991; Goldsmith & Schvaneveldt, 1984; Sanderson, Flach, Buttigieg, & Casey, 1989; Wickens & Andre, 1990) of the experiments utilized in Carswell's (1992) meta-analysis were not graph reading tasks. These experiments were more along the lines of multicue, decision making tasks. In such tasks, subjects are presented with a number of graphical displays. However, the task at hand is not to simply identify or extrapolate spatial relationships. The task in such a diagnostic task is more of a problem-solving task,

wherein subjects are required to integrate numerous quantitatively and qualitatively different stimuli. Such tasks typically involve a great deal of probability estimation. For example, in Barnett and Wicken's (1988) study, subjects were asked to make decisions about whether to continue a mission in an aircraft. This decision was based on assessment of the reliability, diagnosticity, and information worth of numerous variables like fuel level, engine temperature, oil pressure, enemy intent, and pilot fatigue. Although many of these data were depicted spatially, some of the data were not. And more importantly, the processing involved in such a task goes well beyond the processing required in the type of graph reading task that has previously been defined in this paper.

The above criticism is cause for distress when we consider Carswell's (1992) conclusions about synthesis tasks. Carswell indicated that the one weak spot for the basic perceptual skills model was its predictive value for synthesis tasks. In such instances, the data-ink model seemed to provide the better explanation for performance. It is likely that the basic perceptual skills model emerged as less able to predict performance for synthesis tasks because 75% of the experiments labeled as synthesis tasks were those that involved multicue information displays and the problem solving described above. It is safe to conclude that the basic perceptual skills model was not validly applied to true synthesis tasks. Applying this model to more closely controlled synthesis tasks might reveal that it is more predictive of performance in such graphicacy tasks than we have been led to believe.

To remedy the above problems, it will be necessary to develop a series of graphs that comprise all of the basic perceptual and analytical tasks and that hold constant the quality and quantity of as many variables and specifiers as possible. It is anticipated that this will yield a sample of graphs that are conceptually and perceptually similar. After such graphs have been produced, we will be in a better position to determine whether and how the basic perceptual skills and analytical tasks models describe graphicacy as it is truly manifest in day-to-day graph reading.

Summary

The commonsense foundation underlying graphicacy theories seems to have impacted the empirical research in a rather divisive way. Numerous researchers have approached the issue of graphicacy from several perspectives that seem to be only tangentially linked to one another. This has made it difficult to discuss groups of studies under the same theoretical rubric.

The lack of a common theoretical umbrella has led to numerous design implementations that have hindered this domain's evolution. This is important because there is tremendous performance variability across tasks. In addition, individual specifiers (e.g., where they are actually located, how they are arranged, how they are sized) within the graph seem to further add to task-dependent fluctuations in performance accuracy. Because a graph's effectiveness varies from one context to another, it is of paramount importance to identify those factors that seem to be pivotal in this variability. This domain has struggled to develop general rules that will provide some common ground for broad discussion.

Fairly recently, efforts have been made by researchers, such as Cleveland and McGill (1984, 1986), Kruskal (1982), Tufte (1983) and Tukey (1990), to introduce more standardized criteria for judging and analyzing graphs. Although several of these efforts were too general or incomplete to provide a clear standard (at least for the time being), Cleveland and McGill's basic perceptual skills model is concisely articulated and has proven to be reasonably predictive of judgment accuracy across a broad range of graph reading tasks, according to Carswell's (1992) meta-analysis.

However, the basic perceptual skills model derives from graphicacy tasks that are not as authentic as they should be, if genuine graph reading is the central concern. Further, this model has not been applied to graphs for which important graph reading specifiers have been well controlled. With only minor amendments, Cleveland and McGill's (1984) basic perceptual skills model and Carswell's (1992) analytical tasks taxonomy may provide a useful foundation for the further development of common graphicacy rules. Now, the question must be asked: How well do these models describe performance in more authentic, well controlled graphicacy tasks?

THE STUDY

Purpose of the Study

The purpose of this study was to test prevalent taxonomies in the graphicacy literature. The preceding arguments identified Cleveland and McGill's (1984) and Carswell's (1992) taxonomies as excellent conceptualizations of the most important tasks that underlie graph reading. These taxonomies were utilized to examine graphicacy as it is manifest in more authentic, well controlled, and consistently defined graph reading tasks.

Because this area of research is relatively undeveloped, this experiment proceeded from an exploratory framework. The following were the primary questions that were focused on:

1. Do any of the basic perceptual skills or analytical tasks share a statistically or practically significant amount of variance?
2. When performances for the basic perceptual skills are collapsed across analytical tasks, are the basic perceptual skills ranked in the order predicted by the basic perceptual skills model?
3. For each type of analytical task, are the basic perceptual skills ranked (for performance difficulty) in the order predicted by the basic perceptual skills model?
4. When performances for the analytical tasks are collapsed across the basic perceptual skills, are the analytical tasks ranked in the order predicted by the analytical tasks model?
5. For each type of basic perceptual skill, are the analytical tasks ranked (for performance difficulty) in the order predicted by the analytical tasks model?

Experimental Design and Methodology

Participants

One hundred twenty-four undergraduate students (84 female and 40 male) from Utah State University served as subjects in this experiment. The majority of these students were taken from introductory, educational, and social psychology classes. All subjects were compensated for their

participation with course credit. Subjects were treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 1992).

Design

The rudimentary independent variables in this experiment were the following: (a) type of analytical task (four levels) and (b) type of basic perceptual skill (six levels). The dependent variable in this experiment was the magnitude of judgment errors committed by subjects.

Given the proposed experimental questions and the fact that the ordering of test items was randomly determined, it was necessary to utilize only a single group of subjects. That is, there was no need to implement an experimental design with a control group and a treatment group. All subjects took the same test.

Materials

Each subject filled out an answer sheet, at the top of which was a brief demographic inventory (see Appendix C).

In the set of graphs, there were essentially four sets of basic perceptual skills—one set for each type of analytical task (see Table 1). In other words, each of the basic perceptual skills provided the perceptual foundation for four different analytical tasks. For example, the basic perceptual skill position on a common scale was integrated with each of the point reading, point comparison, global comparison, and synthesis analytical tasks—in four different graphs, of course.

Graphs were constructed with the Adobe Illustrator (Version 6.0) program and printed with a laser printer producing a print resolution of 800 x 800 dots per square inch.

Furthermore, for all graphs, the number of graphical specifiers displayed was three. Other graphical parameters like axis size and border size (2 and 2.5 inches, respectively) were held constant across all graphs, except those focusing on position along a common scale. Other graphical parameters can be easily discerned by looking at the test in Appendix D.

Table 1

Basic Perceptual Skills Grouped Within Each Analytical Task

	Type of analytical task			
	Point reading	Point comparison	Global comparison	Synthesis
Type of basic perceptual skill	PCS	PCS	PCS	PCS
	PNS	PNS	PNS	PNS
	Length	Length	Length	Length
	Slope	Slope	Slope	Slope
	Angle	Angle	Angle	Angle
	Area	Area	Area	Area

Note. In the above table, position along a common scale is designated by the acronym "PCS," and position along identical nonaligned scales is designated by the acronym "PNS." This table illustrates the manner in which the perceptual skills and analytical tasks are grouped. The table does not depict the presentation order of the stimuli.

The systematicity applied to some factors should be described. For point reading, local comparison, and global comparison tasks, the value of graph/specifier A was a randomly determined number between 30 and 39 (including 30 and 39); the value of graph/specifier B was a randomly determined number between 40 and 49 (including 40 and 49); the value of graph/specifier C was a randomly determined number between 50 and 59 (including 50 and 59).³

A different system was used for synthesis tasks. For synthesis tasks, the value of graph/specifier A was a randomly determined number between 30 and 39 (including 30 and 39). Then this value was

³ The constraint of having particular lengths, angles, diameters or distances for each graph increases judgment reliability among a series of graphs (Cleveland & McGill, 1984, 1986), and, reflexively, the comparability of different perceptual and analytical tasks. In addition, the controls placed on true proportions being judged take into account the work of researchers like Barnett and Wickens (1988), Cleveland and McGill (1984), Croxton (1927), Croxton and Stein (1932), Kruskal (1982), and Simkin and Hastie (1987), who have discussed the issues of anchoring points and/or proportional disparities between graphical elements and how judgment seems to be affected by such factors.

multiplied by a trend factor (which was a randomly determined percentage between 125% and 135%)⁴ to yield the value of graph/specifier B, which was rounded to the nearest one-tenth (for slope, angle, and area tasks) or to the nearest one-hundredth (for distance and length tasks). The value of graph/specifier B was multiplied by that same factor to yield the value of graph/specifier C, which was rounded to the nearest one-tenth or one-hundredth. The value of graph/specifier D was multiplied by that factor to yield the value of extrapolated graph/specifier D, which was rounded to the nearest one-tenth or one-hundredth. The effects of rounding (on value-to-value trend) were negligible. After this rounding procedure, all but one value-to-value trend were within .1 of the trend factor. The one value-to-value comparison that was an exception was still within .2 of the trend factor.

A calculator was used to compute the graph/specifier and trend values for tests in the previously described manner. The calculations were rechecked by an assistant before the test was actually constructed.

Procedure

Upon their arrival at the laboratory, subjects were greeted, seated, and invited to read and sign a statement of voluntary consent (see Appendix A). One subject did not sign the statement (because she was not yet 18 years old), and she was dismissed from participation in the experiment. Subjects proceeded to fill out the pertinent participant and demographic information requested at the top of the answer sheet. Then, the experimenter introduced the experiment and led the subjects through eight practice trials and eventually the test (see Appendix B for details). The duration of each experimental session was about 45 minutes.

Instructions were scripted (see Appendix B) and issued, verbatim, by the experimenter. All subjects were exposed to the original 24 graphs. The presentation order of the graphs was randomly determined using a random number table (see Table 7 of Fisher & Yates, 1974).

⁴ A randomly determined trend factor (for synthesis tasks) between 125% and 135% was chosen, because when graph/element A is a value between 30 and 39, a multiplication factor between 1.25 and 1.35 yields an overall trend that is similar to the general trend exemplified in the other types of analytical tasks.

In each point reading task, subjects identified the value of graph/specifier A or C. Whether subjects were asked to identify the value of A or C was determined in the form *ACCA*.... For each local comparison graph, the comparison was between graphs/specifiers A and C. The direction of this comparison was determined in the form *ACCA*.... For each global comparison task, the comparison was between graphs/specifiers A and B + C or between C and A + B. The type of comparison was determined in the form *ACCA*.... For each synthesis task, the proportional comparisons were between graphs/specifiers B and D or between D and B. Again the direction of the comparison was determined after the form *BDDB*....

An issue that arose during testing was one relating to instrumentation. It was a critical issue and is therefore introduced and explained to some extent at this juncture. For synthesis tasks, some subjects indicated that instead of making comparisons between specifiers/graphs B and D (as they were instructed to do), they instead made comparisons between specifiers/graphs A and C. Given that the proportional rate of increase (i.e., the trend) for graphs A, B, and C was uniform, it was actually not necessary for subjects to extrapolate specifier/graph D and then compare it with B to answer the question. The same answer could have been derived by comparing A and C.

One of the outgrowths of the above revelation was the construction of a new test item. The final test item focused on identifying the test-taking strategy of subjects when they attempted to make judgments about synthesis tasks. For item 25, subjects were asked the following: "For questions in which you were asked to compare graph B with graph D, did you actually do this?" An answer of "yes" was an indication that the subject had followed the instructions, and an answer of "no" was an indication that the subject had used specifiers/graphs A and C to answer the question. The experimenter elaborated upon this topic to the extent that it was clear that everyone understood what was being asked. The question was asked in a matter-of-fact way so as to increase the likelihood that subjects would answer the question with as much honesty as possible. Subjects who were unsure about whether they used one strategy more than the other were told to leave this item blank. Sixty-one subjects responded to this item.

At the conclusion of the experiment, subjects were debriefed as to the general nature of the

experiment. Subjects were informed that they could receive feedback from the experimenter, after the data had been analyzed (see Appendix B for details).

Pilot Testing

Two pilot tests were conducted for the purpose of assessing and amending the comprehensibility of the instruction script. A complete discussion of this process and how the findings translated into alterations to the experimental script is contained in Appendix F.

Data Entry and Analysis

All data were entered using SPSS for Windows and were rechecked for accuracy after the original data entry. To calculate subjects' absolute error scores, the correct response (which was the correct proportion rounded to the nearest one-tenth) for each graph was subtracted from the subject's actual response to yield the magnitude of judgment error. A computer was used to make these calculations. All analyses were performed using SPSS.

RESULTS

Exclusion of Data Related to the Area Skill Involving Point Reading

Prior to the presentation of experimental results, it is necessary to address a problem associated with the interpretation of the area skill involving point reading (i.e., item 5 of the test). As data collection was nearing its end, it was realized that item 5 did not precisely require subjects to make a perceptual judgment of area. To get a clearer understanding of just why this was, look at item 5 of the graphicacy test (see Appendix D). Note the scale of 100 on either axis. Imagine trying to estimate the area of square A within this context. What would you need to do? To answer item 5, you would only need to estimate the length of one side of square A. You could estimate the height or width. It hardly matters which, given that the square in item A is a perfect square. After doing so, you could square that value to derive your answer. This is precisely what subjects had been doing to respond to this question. It was realized (late in the data collection phase) that during such a process, subjects would have been doing as much calculating as perceiving. This calculation or numeracy factor emerges as a confound, given that this type of processing is not required for performance on any other test item. Further, this type of processing is not accounted for by the tenets of the basic perceptual skills or analytical tasks models.

Additionally, even though the calculations for area (i.e., length x width) were explained to subjects and they were told to show all handwritten work for this item, only about 72% of all subjects actually showed their work, and only a few subjects ever used calculators. Of that 72%, about 28% showed work that looked like something other than the expected calculations for the area of a square; in the majority of such calculations, subjects added length and width instead of multiplying length and width. The fact that many subjects who showed their work did not perform the correct calculations is reason to be skeptical about the mental calculations of the 28% of subjects who did not show their work.

In summary, the data for the area judgment involving point reading are spurious for two important reasons. First, in this task subjects were performing certain mental operations that were not performed in other tasks. Second, a significant percentage of subjects was not even performing the correct

calculations for area, as they had been instructed to do so. Therefore, it was deduced that the data yielded from performance on item 5 are extremely misleading and should be omitted from all of the analyses and tables presented.

Intercorrelations

Analysis Rationale

To test the amount of variance shared by each pair of test items and different combinations of test items, the data were subjected to numerous Pearson product-moment correlations. Correlations that were at or lower than an alpha level of .05 (i.e., $p \leq .05$) were considered to be statistically significant. All indicators of statistical significance reflect two-tailed correlations. Further, in accordance with Cohen's (1988) generic guidelines, r values between absolute .1 and .3 were interpreted as weak, and r values between .3 and .5 were interpreted as indicators of a moderate degree of shared variance. Cohen's generic criteria were adopted in light of the fact that neither the relevant perceptual literature nor the graphicacy literature provided any methodical direction for constructing more tailored operational definitions relating to the specific processes which were investigated in this study. The findings are interpreted primarily in terms of shared variance (i.e., r^2) between any two correlated variables (as opposed to the statistical significance of correlations). Because of this dissertation's fundamental focus on correlation coefficients as descriptive measures of the linear association (as opposed to indicators of statistical significance) between variables, it was not necessary to test for the statistical assumptions of normality and homoscedasticity (see Cohen, 1988). Therefore, no data for these assumptions are presented. Mean absolute error scores for each test item are presented in Appendix E (see Tables E.1 and E.2).

Analysis of Individual Items

It would be redundant to present the correlations between items separately for perceptual skills and analytical tasks, inasmuch as they are inextricably linked within each item. Therefore, this section will provide practically significant correlations between test items, and references to identifiable trends for

basic perceptual skills and analytical tasks will be reserved for subsequent sections of this paper. Further, the identification and discussion of weak inter-item correlations (where moderate-to-strong correlations might have been expected) will be reserved for the "Discussion" section of this paper.

Of the 217 correlation coefficients in Table E.3, there were only 13 that were at least .3 in strength. Of these 13 correlations, 10 of them were between items representing the same analytical task but different perceptual skills. Not one of the 13 correlations was between items representing the same perceptual skill but different analytical tasks. Three of the 13 correlations were between items representing different analytical skills and different perceptual skills. Given that only 6% of the inter-item correlations were marginally moderate to moderate in strength, this provides strong support for Cleveland and McGill's (1984, 1986) perceptual skills model as well as Carswell's (1992) analytical tasks taxonomy, because it indicates the autonomy of the individual analytical tasks and perceptual skills. This issue is elaborated upon more fully in the "Discussion" section.

Basic Perceptual Skills Collapsed Across Analytical Tasks

As Table 2 indicates, 10 of the 15 correlations between perceptual skills are statistically significant. The strongest of these correlations are between slope and angle tasks, $r(120) = .38, p < .01$, tasks focusing on the judgment of positions on nonaligned identical scales and position along a common scale, $r(122) = .38, p < .01$, and between position along a common scale and length tasks, $r(123) = .38, p < .01$. Squaring any one of these correlations yields a value of .14, indicating the variables in any one of these variable pairs share 14% variance with one another. In other words, the cognitive processes used to judge graphs in slope and angle tasks (for example) have about a 14% overlap. This should be considered as approaching a moderate amount of overlap and an indication that these pairs of variables were tapping the same constructs, to a moderately significant extent. The other correlations that would be considered as small-to-moderate correlations were between area and angle tasks and between angle and length tasks. The other statistically significant correlations would be considered as small correlations. In summary, several pairs of basic perceptual skills appear to be tapping the similar cognitive processes, raising the

Table 2

Intercorrelations for GPA, ACT, and Basic Perceptual Skills Collapsed AcrossAnalytical Tasks

	GPA	ACT	Type of basic perceptual skill					
			Slope	PCS	PNS	Area	Angle	Length
GPA	--	.57**	-.02	-.16	-.13	-.11	.01	.08
ACT	--	--	-.15	-.32**	-.13	-.26*	-.17	.01
Slope	--	--	--	.12	.10	.23*	.38**	.26**
PCS	--	--	--	--	.38**	.17	.21*	.38**
PNS	--	--	--	--	--	.18*	.23*	.17
Area	--	--	--	--	--	--	.31**	.12
Angle	--	--	--	--	--	--	--	.35**
Length	--	--	--	--	--	--	--	--

Note. PCS=positions on a common aligned scale; PNS=positions on identical nonaligned scales. Data for GPA and ACT represent overall, undergraduate grade point average and the Composite score on the American College Test.

* $p < .05$. ** $p < .01$.

issue of whether such basic perceptual skills as determining position along a common scale and determining length can be discussed as separate skills.

Given the number of statistically significant correlations, it was determined that the issue of overlapping skills should be further assessed. Correlations between each perceptual skill and all other perceptual skills were averaged for each perceptual skill. That is, the correlations between angle and each of the other perceptual skills were averaged, the correlations between length and each of the other perceptual skills were averaged, and so forth. This procedure allowed the researcher to determine whether the performance variance of any one perceptual skill overlapped significantly with all other skills, in general. An average correlation of greater than .3 between any one perceptual skill and all others was

taken as an indication of that perceptual skill's redundancy within the model. The averaging of correlation coefficients followed the methods prescribed by Glass and Hopkins (1996). This process involved the transformation of correlation coefficients into Z -scores via a logarithmic transformation based on Fisher's Z -transformation. Subsequently, the values were weighted and then averaged to offset the effects of skewed distributions and different sample sizes underlying the different data points. The formula for the calculations was as follows:

$$\bar{Z}_w = \frac{\sum w_j Z_j}{w}$$

where $w_j = n_j - 3$, $w = \sum w_j$, and $Z_j = 1.1513 \text{ Log} \left(\frac{1+|r|}{1-|r|} \right)$.

After applying the appropriate transformations and weightings, the average correlations between each perceptual skill and all other perceptual skills were as follows: slope (.22), position along a common scale (.25), positions on nonaligned identical scales (.21), area (.20), angle (.31), and length (.27). Only performance of angle judgments appears to have been moderately related to performance on all other perceptual judgments. This is an indication that the perceptual skills used to make angle judgments overlap to some extent with all other perceptual skills and that the information provided by judgments of angle is somewhat redundant, after the information from other perceptual judgments is taken into account. This will be addressed in more detail later in the paper.

Analytical Tasks Collapsed Across Basic Perceptual Skills

Of the six correlations between analytical tasks, depicted in Table 3, three of them were statistically significant at the .05 level. Of these correlations, the strongest was between synthesis tasks and global comparison tasks, $r(122) = .28$, $p < .01$. This should be considered as a weak correlation, as it is an indication that these two types of tasks share only about 8% performance variance. This supports the notion that each analytical task seems to be testing a different cognitive process.

Table 3

Intercorrelations for GPA, ACT, and Analytical Tasks Collapsed Across BasicPerceptual Skills

	ACT	GPA	Synthesis	Type of analytical task		
				Global comparison	Local comparison	Point reading
ACT	--	.57**	.09	-.26*	-.32**	-.18
GPA	--	--	.21	-.18	-.15	-.16
Synthesis	--	--	--	.28**	.22*	-.03
Global comparison	--	--	--	--	.26**	.03
Local comparison	--	--	--	--	--	.05
Point reading	--	--	--	--	--	--

Note. Data for GPA and ACT represent overall, undergraduate grade point average and the Composite score on the American College Test.

* $p < .05$. ** $p < .01$.

Rank Ordering of the Different Types of Tasks

Analysis Rationale

One of the driving concerns of this study was the following: Given the arguments and conceptions of graph reading advanced in this paper, are the basic perceptual skills and analytical tasks ranked (for performance difficulty) in the order proposed by Cleveland and McGill (1984,1986) and Carswell (1992)? Table E.1 shows means and standard deviations for the absolute-error performance associated with each test item. These data provide an initial indication as to how the above question will be answered.

Regarding analytical skills, the data in Table E.1 suggest that point reading tasks resulted in the most accurate performance. Point reading accounts for five of the first seven items in Table E.1, which

has the test items listed in terms of increasing difficulty. In other words, the error associated with subjects' responses was smallest for point reading tasks. However, performance on the basic perceptual skills is related to the type of underlying analytical task associated with each item, and beyond the first four items, it is difficult to extrapolate any cohesive performance trends relating to analysis and perception. A more effective way to approach this task is to look at the ranking of perceptual skills within each type of analytical task, as well as the ranking of analytical tasks within each type of basic perceptual skill.

Perceptual Skills

Perceptual skills collapsed across analytical tasks. The second research question posed at the outset of this study was: When performances for the basic perceptual skills are collapsed across analytical tasks, are the basic perceptual skills ranked in the order predicted by the basic perceptual skills model? Recall that the authors proposed that the basic perceptual skills in graph reading could be ranked for difficulty in the following way: (a) position along a common scale, (b) positions along identical nonaligned scales, (c) length, (d) slope, (e) angle, and (f) area.

To be precise, Cleveland and McGill maintained that judgments involving position along a common scale are only slightly easier to make judgments about than tasks involving points on identical nonaligned scales. Similarly, it was found that when proportions to be estimated are near 0 or 100%, angle estimates are slightly more accurate than slope estimates, and when proportions to be estimated are closer to 50%, estimates of slope are slightly more accurate than estimates of angle.

As was stated, a fundamental issue in the testing of the basic perceptual skills model was the determination of whether the model is predictive of performance when the data for each perceptual skill are collapsed across all levels of the analysis variable. That is, given that more than one type of analysis can be made in concert with any one type of perceptual skill, it is important to determine the difficulty level for each perceptual skill by finding the average performance across all types of analyses.

As Figure 3 shows, subjects were most effective in their judgments about graphs involving position on a common scale. That is, judgments related to such perceptual skills are easiest to make when

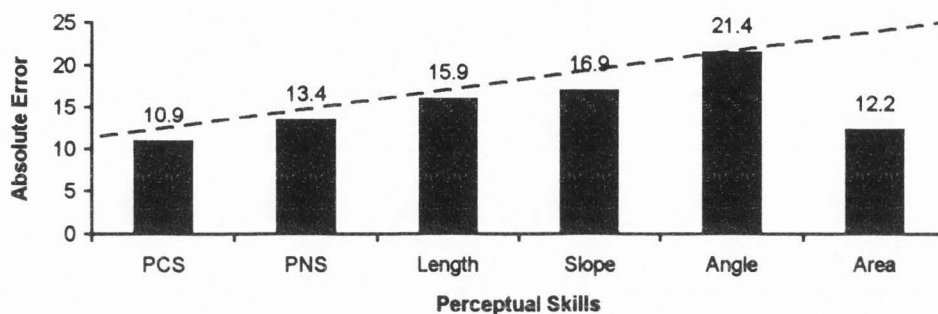


Figure 3. Mean absolute error for perceptual skills collapsed across analytical tasks.

Note. PCS=position on a common scale; PNS=positions along identical nonaligned scales. The value above each bar represents the mean, absolute judgment error associated with each type of perceptual skill when performance is collapsed across all levels of analysis. Standard deviations associated with each mean are as follows: PCS (5.6), PNS (7.9), Length (9.6), Slope (9.2), Angle (10.2), Area (8.7). The dashed line represents the approximate absolute-error trend that was expected across perceptual skills.

all forms of analysis are considered conjointly. Generally speaking, the predictions of the perceptual skills model are born out in this analysis. The exception to the predicted trend is manifest in performance data associated with items involving area judgment. Figure 3 indicates graphically how subjects' judgment of area diverged dramatically from that which would have been predicted by the model.

Perceptual skills ranked within each type of analytical task. The third experimental question posed prior to this investigation was: Within each type of analytical task, are the basic perceptual skills ranked (for performance difficulty) in the order proposed by Cleveland and McGill (1984, 1986)? Note that Cleveland and McGill (1984, 1986) developed the basic perceptual skills model through research deriving from a single type of analysis—a local comparison task where two specifiers were being compared. However, local comparison is only one of four types of analytical tasks. To determine whether this model is robust, it was necessary to assess its predictive power for all types of analytical tasks.

With regard to point reading tasks, the data in Table 4 indicate that performance on this test was inconsistent with the predictions of the basic perceptual skills model. Items involving the judgment of

Table 4

Mean Absolute Error Scores for Various Tasks

Type of basic perceptual skill	Type of analytical task			
	Point reading	Local comparison	Global comparison	Synthesis
Position on common scale	5.1 (3)	20.7 (4) ^a	6.9 (2)	11.0 (2)
Positions on nonaligned identical scales	4.3 (1)	31.6 (5) ^a	6.8 (1)	10.9 (1)
Length	4.4 (2)	18.2 (3) ^a	15.1 (6)	26.0 (4) ^a
Slope	7.7 (5)	11.5 (1)	14.8 (5)	33.2 (6) ^a
Angle	6.5 (4)	39.9 (6) ^a	11.9 (4)	27.3 (5) ^a
Area	--	14.1 (2)	10.1 (3)	11.5 (3)

Note. The values in parentheses in each column represent the actual difficulty ranking (1=easiest and 6=hardest) associated with the perceptual skills as they are ranked within each type of analytical task.

^aValues indicate tasks for which the proportion being judged was over 100%.

positions on nonaligned identical scales were performed with more accuracy than were length judgments, which were performed with more accuracy than were judgments focusing on position on a common scale.

For local comparison tasks, the findings of this study do not support the predictions of the basic perceptual skills model. Within the local comparison task, the perceptual skills are ranked for accuracy in the following way (from most to least accurate): slope, area, length, position along a common scale, position on nonaligned identical scales, and angle.

For global comparison analyses, absolute error scores for judgments involving positions on identical nonaligned scales and position on a common scale were the smallest, indicating that these were the easiest perceptual skills to perform. Given the marginal performance difference for these two items and the fact that Cleveland and McGill (1986) conceded only a small performance advantage in position along common scale tasks, it is reasonable to conclude that this finding is consistent with the predictions

of the basic perceptual skills model. However, the remaining perceptual skills were ranked for difficulty in such a way that was not predicted by the perceptual skills model. Recall that the perceptual skills model predicted that the remaining perceptual skills would be ranked as per the following: length, slope, angle and area. However, this study found that for global comparison tasks, area graphs were judged with more accuracy than were angle graphs, which were judged with more accuracy than slope and length. This ranking is inconsistent with Cleveland and McGill's difficulty ranking of the basic perceptual skills.

Regarding synthesis tasks, the data in Table 4 indicate inconsistency between the basic perceptual skills model and the findings of this study. The most glaring discrepancy is manifest within the judgment accuracy associated with area, length, angle, and slope skills, from which area emerges as the skill associated with the most judgment accuracy.

In summary, it appears that when the ranking of perceptual skills is considered within each level of analysis, the basic perceptual skills model is not predictive of performance. There are many potentially extenuating causes for the incongruence between findings in this study and the predictions of Cleveland and McGill (1984, 1986). These causes will be addressed in the "Discussion" section of this dissertation.

Analytical Tasks

Explication of the difficulty associated with different types of analysis. There were four difficulty levels of analysis on this test. According to Carswell (1992), these levels of difficulty varied along two dimensions: (a) the number of specifiers that had to be attended to in order to answer a question, and (b) whether an item required a comparison of graph features actually presented or a comparison involving some cognitive standard. Regarding the first dimension, point reading required attention to one specifier; local comparison required attention to two specifiers; global comparison required attention to three specifiers; synthesis required attention to four specifiers. With respect to the second dimension, point reading and local comparison tasks required a comparison of presented specifiers, and global comparison and synthesis tasks required a comparison of a cognitively imaged specifier with a presented specifier. Taken together, these dimensions suggest that the analytical tasks can be ranked for difficulty (from the

most accurately perceived to the least accurately perceived) in the following way: (a) point reading, (b) local comparison, (c) global comparison, and (d) synthesis.

Analytical tasks collapsed across perceptual skills. The fourth question posed at the outset of this study was: When performances for the analytical tasks are collapsed across the basic perceptual skills, are the analytical tasks ranked in the order predicted by the analytical tasks model? If Carswell's (1992) model of analytical tasks is sufficiently robust, then it should be predictive of not only analysis for individual perceptual skills, but also analysis across perceptual skills. Just as collapsing perceptual skills across analytical tasks increases the representativeness of a given perceptual skill, so to does the collapsing of analytical tasks increase our ability to more effectively discuss the general implications of such tasks.

Figure 4 depicts mean absolute error scores for analytical tasks collapsed across perceptual skills. The figure indicates that the findings from this study are inconsistent with the predictions of the analytical

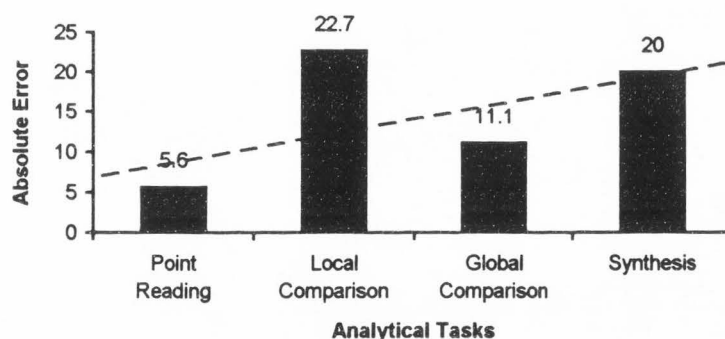


Figure 4. Mean absolute error scores for analytical tasks collapsed across perceptual skills.

Note. The standard deviation associated with each type of analytical task was as follows: Point Reading (2.4), Local Comparison (9.1), Global Comparison (9.3), Synthesis (9.8). The dashed line represents approximately the absolute error trend that was expected across the different analytical tasks.

tasks model. In fact, the trend in Figure 4 starkly contrasts with the predictions of the analytical tasks model. The contradiction involves the ranking of local comparison tasks as the most difficult type of analytical task. In other words, subjects were least accurate in their judgments of graphs requiring a local comparison. A combination of factors could have led to these contradictory findings. These will be discussed in the "Discussion" section of this dissertation.

Analytical tasks ranked within each perceptual skill. The fifth question posed at the outset of this study was: For each type of basic perceptual skill, are the analytical tasks ranked (for performance difficulty) in the order predicted by the analytical tasks model? Table 5 indicates discord between the predictions of the analytical tasks model and the findings from this study. The only instance in which the analytical tasks are ranked in accordance with the predictions of the analytical tasks model is for the perceptual skill of slope. In all other cases, the analytical tasks are ranked inconsistently.

Table 5

Analytical Tasks Ranked for Difficulty Within Each Perceptual Skill

Type of analytical task	Type of basic perceptual skill					
	Length	Position on nonaligned identical	Position on common scale	Angle	Slope	Area
Point reading	4.4 (1)	4.3 (1)	5.1 (1)	6.5 (1)	7.7 (1)	--
Local comparison	18.2 (3) ^a	31.6 (4) ^a	20.7 (4) ^a	39.9 (4) ^a	11.5 (2)	14.1 (3)
Global comparison	15.1 (2)	6.8 (2)	6.9 (2)	11.9 (2)	14.8 (3)	10.1 (1)
Synthesis	26.0 (4) ^a	11.0 (3)	11.0 (3)	27.3 (3) ^a	33.2 (4) ^a	11.5 (2)

Note. The values in parentheses in each column represent the actual difficulty ranking (1=easiest and 4=hardest) associated with the perceptual skills as they are ranked within each type of analytical task.

^a Values indicate tasks for which the proportion being judged was over 100%.

Over- and Under-100% Estimation

Although it was not a fundamental issue focused on at the outset of this paper, data analysis revealed distinct anchoring trends associated with proportion judgments above and below 100%. Table E.2 (see Appendix E) depicts the average error scores for each test item. Note that these values reflect "straight" error. That is, the means represent error that was not converted into absolute values. The data indicate that for every task where subjects needed to make a proportion judgment in excess of 100%, subjects (as a group) underestimated the proportion. Conversely, for every task where subjects needed to make a proportion judgment of less than 100%, subjects (as a group) overestimated the correct proportion. This issue will be addressed further in the "Discussion" section.

Gender-Based Performance Differences

Basic Perceptual Skills Collapsed Across Analytical Tasks

Beyond the topics pertinent to the research questions, other important issues emerged and should be addressed. Given the visual-analytical nature of this test and the empirically supported notion that males enjoy a performance advantage in analytical and visual-spatial tasks (see Bouchard & McGee, 1977; Sanders, Soares, & D'Aquila, 1982; Tapley & Bryden, 1977), it was important that part of the data analysis procedure be devoted to determining whether males scored significantly higher than females on the test.

To assess gender-based performance differences on the perceptual skills on this test, an independent-samples t test was conducted. Pertinent descriptive statistics underlying this procedure are contained in Table 6.

To begin, the Levene statistic was calculated to test for the assumption of homogeneity of variance between male and female performances for each perceptual skill. A conservative value of .1 was set as the critical alpha level for the Levene test. (For more discussion on the use of a conservative alpha value in the Levene procedure, the reader is referred to Glass and Hopkins, 1996.) For all but items

involving angle and area judgments, the Levene value was not statistically significant, indicating that a pooled-variance t test was appropriate. For angle and area judgments, the Levene values were statistically significant at the .1 level, indicating that a separate-variance t test was appropriate for assessing gender-based performance differences on these two variables.

For all t tests, an alpha level of .05 was used as the criterion for statistical significance. It was found that gender performance differences were not statistically significant for any comparison of means. This provides a strong indication that males did not have a performance advantage on the visual-spatial component of this test. Further, the graphical representation of data in Figure 5 visually supports the conclusion that males and females performed equally well on the visual-spatial component of this test.

Table 6

Descriptive Statistics for Mean Absolute Error for Basic Perceptual

Skills Broken Down by Gender

Type of perceptual skill	Gender	Statistics		
		<u>N</u>	<u>M</u>	<u>SD</u>
Slope	male	40	16.4	9.8
	female	82	17.2	8.9
PCS	male	40	10.5	4.8
	female	84	11.1	6.0
PNS	male	39	11.7	6.7
	female	84	14.1	8.3
Area	male	40	14.0	11.6
	female	84	11.3	6.8
Angle	male	39	21.1	12.6
	female	84	21.6	8.9
Length	male	40	16.7	8.6
	female	84	15.5	10.1

Note. PCS=points on a common scale; PNS=points on identical nonaligned scales.

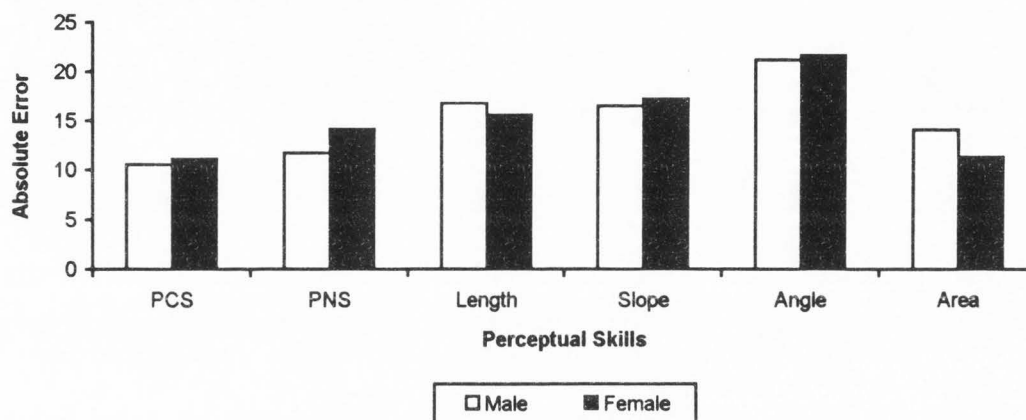


Figure 5. Mean absolute error for basic perceptual skills broken down by gender.

Note. PCS=position on a common scale; PNS=positions on nonaligned identical scales.

Additionally, standardized mean difference effect sizes were calculated for male-female performance comparisons for each basic perceptual skill. The following formula was used to calculate the effect sizes:

$$ES = t \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where n_1 =the number of male subjects and n_2 =the number of female subjects who participated in the experiment.

Cohen's (1988) generic criteria, for determining small, medium, and large effects sizes, were applied. The standardized mean difference effect sizes for items involving positions on a common scale, positions on nonaligned identical scales, length, slope, angle, and area were: .11, .31, .11, .08, .04, and .26, respectively. Even the largest of these effect sizes would be considered small according to Cohen's (1988) standards. Once again, the data indicate that males and females performed similarly in the analytical components of the graph reading tasks.

Analytical Tasks Collapsed Across
Basic Perceptual Skills

To assess gender-based performance differences on the analytical tasks on this test, independent-samples t tests were conducted, using the same protocol and conceptualizations outlined above. The descriptive statistics underlying this procedure are contained in Table 7.

The Levene statistic was calculated to test for the assumption of homogeneity of variance between male and female performances for each type of analytical task. Again, a conservative value of .1 was set as the critical alpha level for the Levene test. It was found that for all but tasks involving global comparisons, the Levene value was not statistically significant, indicating that a pooled-variance t test was appropriate. For global comparison tasks, the Levene values were statistically significant at the .1 level, indicating that a separate-variance t test was appropriate for assessing gender-based performance differences on global comparison tasks.

Table 7

Descriptive Statistics for Mean Absolute Error for Analytical Tasks

Broken Down by Gender

Type of analytical task	Gender	Statistics		
		N	M	SD
Synthesis	male	38	20.0	10.4
	female	83	19.9	9.5
Global comparison	male	40	12.5	11.4
	female	84	10.4	8.1
Local comparison	male	40	20.7	8.7
	female	84	23.6	9.2
Point reading	male	40	5.5	2.6
	female	83	5.7	2.4

For all t tests, an alpha level of .05 was used as the criterion for statistical significance. It was found that gender performance differences were not statistically significant for any comparison of means. This provides a strong indication that males did not have a performance advantage on the analytical component of this test.

Figure 6 depicts mean absolute error for analytical tasks broken down by gender. The figure graphically supports the contention that male and female subjects performed with relatively equal accuracy on the analytical components of this test.

In addition to t tests, standardized mean difference effect sizes were also calculated using the above-mentioned protocol for calculating and assessing effect size differences. The effect sizes for gender-based performance differences in synthesis, global comparison, local comparison, and point reading tasks were: .01, .23, .23, and .08, respectively. Once again, by Cohen's (1988) standards, such performance differences would be considered practically as small, indicating that males and females performed similarly in the analytical components of the graph reading tasks.

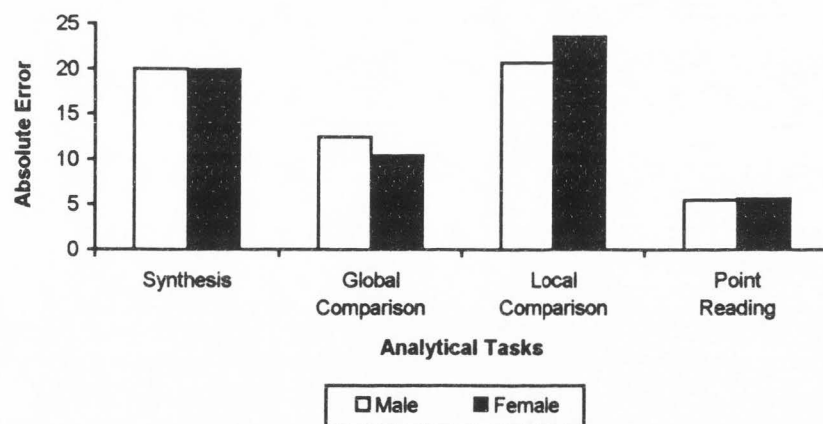


Figure 6. Mean absolute error for analytical tasks broken down by gender.

ACT Composite, GPA, and Performance

The ability to read graphs is important to the success of today's students. Most text books in junior and senior high schools include graphs and figures as a means of portraying quantitative information visually. Therefore, the more accurately one is able to make judgments about such graphs, the more successful one will be with relevant subject matter. Furthermore, the ability to accurately assess graphs has some bearing on performance on important standardized tests, such as the ACT. Unfortunately, graphicacy researchers have done little to investigate common educational measures that would be predictive of graph reading ability.

Given that the ACT Composite score represents verbal, analytical, and quantitative abilities, it was expected that subjects' ACT scores would correlate at least moderately and negatively with absolute error scores related to the analytical tasks. By that same token, it was expected that ACT scores would not correlate with performance on the perceptual skills being tapped in this graph reading task. Also, it was assumed that inasmuch as the vast majority of classes taken by undergraduates emphasize processes similar to analysis (e.g., induction, deduction), moderate and negative correlations would emerge between GPA and analytical tasks.

Correlations were run on ACT Composite scores, GPA, and test performance scores (see Tables 2 and 3) for perceptual skills collapsed across analytical tasks, and analytical tasks collapsed across perceptual skills. Correlations between GPA and perceptual skills/analytical tasks and between ACT and perceptual skills/analytical tasks were fairly consistently negative, as was expected. However, Table 2 indicates that only two of the six correlations between ACT scores and perceptual skills were statistically significant at the .05 level. Table 3 indicates that two of the four correlations between ACT scores and analytical tasks were statistically significant at the .05 level. No statistically significant correlations were found between GPA and analytical tasks. No statistically significant correlations were found between GPA and perceptual skills.

Of the statistically significant correlations, the strongest was $-.32$ between performance on local comparison tasks and ACT scores and between performance on tasks of positions along a common scale and ACT scores. The correlation of $-.32$ means that for both of these sets of variables, the statistically significantly correlated variables share about 10% variance, leaving about 90% variance to be explained by extraneous sources of error and different constructs. This correlation indicates that these sets of tasks are tapping the same cognitive processes, to a moderate degree.

However, aside from these two moderately strong correlations, there were 18 that were weak. It was not expected that so few moderate or even strong correlations between GPA/ACT scores and test performance would emerge. An explanation for this finding will be offered in the "Discussion" section of this dissertation.

Test-Taking Strategy for Synthesis Tasks

An issue that was identified in the midst of data collection was the fact that some subjects were not actually making synthesis judgments when they were asked to do so. Instead, of comparing specifiers/graphs B and D (as they were instructed to do), some subjects made comparisons between A and C. In reality, this latter strategy resulted in local comparisons instead of synthesis judgments. To examine the effects of subjects' choice of strategy, an additional item was added to all subsequent tests. This item was administered at the end of the test and consisted of the following question: For questions in which you were asked to compare graph B with graph D, did you actually do this? When the data were entered into the computer, an answer of "yes" was coded as 1 and an answer of "no" was coded as 2. Thirty-eight subjects indicated that they followed the instructions and extrapolated specifier/graph D. Twenty-three subjects indicated that they used A and C to answer synthesis questions.

To assess whether performance on synthesis tasks was related to subjects' test-taking strategies, a Pearson product-moment correlation was run. Given the above coding protocol and the assumption that the use of specifiers/graphs A and C (instead of B and D) to answer synthesis questions would increase

performance (i.e., lower absolute error), we would expect a negative correlation between item 25 and the synthesis items. Table 8 illustrates the absence of a practically significant correlation (i.e., stronger than absolute .3) between the type of strategy employed to answer synthesis items and mean absolute error performance on synthesis items. In short, test-taking strategy does not appear to have had a significant effect on judgment in synthesis tasks.

Table 8

Correlations Between Original Synthesis Tasks and Test-Taking Strategy Employed

for Such Items

	Synthesis- PCS	Synthesis- slope	Synthesis- angle	Synthesis- area	Synthesis- length	Synthesis- PNS
Synthesis strategy	.02	-.02	-.08	.12	.08	-.08

Note. PCS=positions on a common aligned scale; PNS=positions on identical nonaligned scales.

DISCUSSION

Intercorrelations

Uniqueness of Individual Items and Underlying Cognitive Processes

Cleveland and McGill (1984) maintained that they never intended for their taxonomy of basic perceptual skills to comprise cognitive processes that were structurally independent. Nonetheless, there is an important need to determine the degree of variance shared by these individual factors in light of the conceptualizations proposed in this dissertation. Without making such a determination, it is impossible to know whether and to what extent the individual basic perceptual skills should be considered to be independent of one another.

Surprisingly, the independence of the different test items was highly robust. Although it was not stated as such at the outset of this paper, it was suspected that there would be many statistically and even practically significant correlations between test items. This was not the case. As was indicated previously, of the 217 correlation coefficients in Table E.3, there were only 13 that were at least .34 in magnitude, and 11 of these were between items representing the same analytical task but different perceptual skills. One of the 13 correlations was between items representing the same perceptual skill but different analytical tasks. Four of the 13 correlations were between items representing different analytical skills and different perceptual skills. Given that only 6% of the inter-item correlations were marginally moderate to moderate in strength, this provides strong support for Cleveland and McGill's (1984, 1986) perceptual skills model as well as Carswell's (1992) analytical tasks taxonomy, because it indicates autonomy of the different cognitive processes utilized in the different types of graph reading tasks.

Perceptual Skills Collapsed Across Analytical Tasks: Overlapping Perceptual Processes

With regard to data relating to perceptual skills collapsed across analytical tasks, the strength of the fit between the data and the basic perceptual skills model seems to have received little support. Ten of

the 15 correlations between perceptual skills were statistically significant. The strongest of these correlations was between slope and angle items, items focusing on the judgment of positions on nonaligned identical scales and position along a common scale, and between position along a common scale and length items. All of these correlations were .38 and would be considered as small to moderate in strength. Some of these correlations were expected. Although Cleveland and McGill (1984, 1986) did not analyze shared variance between their perceptual skills, they did suggest that the perceptual demands of the different skills were necessarily somewhat similar. Correlations between different position items, angle and slope items, and position and length items could have been predicted from the outset due to the visual-spatial similarities of these types of test items.

However, some of the correlations were not expected. For example, moderate correlations were found between angle and length items (.35) and between angle and area items (.31). In all likelihood, the number of statistically and practically significant correlations between perceptual skills was increased by the manner in which these graphs were constructed. As was suspected at the outset of this study, equating critical elements such as frames, axes, and even specifier values across all graphs resulted in graphs that were more perceptually similar, which probably resulted in the large number of meaningful correlations. This is an important finding, because it highlights the importance of isolating the effects of the perceptual skills by controlling for as many factors as possible. It appears as though the perceptions of different types of graphs are highly dependent on consistent elements, as well as the types of specifiers themselves.

An additional explanation for these unexpected correlations (as well as all correlations deriving from this analysis) may be manifest in the analytical tasks. When perceptual skills were collapsed across analytical tasks, correlations between different perceptual skills were inherently linked to correlations of shared analytical tasks. It may not be that perceptual skills were correlated. It may have been that the correlations were attributable to shared analytical tasks. Certainly this remains a tenable alternative explanation for these findings.

There are a number of other critical design implementations in this study that may have increased the shared variance between different skills. Admittedly, the correlation between angle and

slope is most likely due to the spatial arrangement of specifiers and graph elements employed in this test. In this test, slope was created by adjusting the angle of a line, whose origin was the intersection of the y and x axes. This created an angle above the slope line and an angle below the slope line. Ideally, subjects should have been attending to the line for slope items, whereas they should have been attending to the blank space between two lines for angle items. However, some subjects probably used the angle created by the slope line and the x axis and the angle created by the slope line and the y axis to make some of their judgments of slope. In short, the slope test item may have been a disguised angle item.

The perceptual differences between slope and angle test items and the authenticity of slope items could have been increased perhaps by adjoining the slope line higher up on the y axis or simply placing the slope line in the middle of the graph space, touching neither axis, as was the case in the instrument used by Cleveland and McGill (1986). The bars in Figure 7 represent mean absolute error for each type of basic perceptual skill. This looks like a typical bar graph—a bar graph that you would find just about anywhere. Suppose that you were trying to judge the rising slope of the bars in Figure 7. Where would the slope line begin? Note the dashed slope line overlaying the bars in Figure 7. The line begins in the middle of the y axis at about a value of 11. It does not begin where x and y meet. In this study, beginning the slope line higher up on the y axis would have diminished the similarities between slope and angle test items, which could have reduced the amount of performance variance shared by the two skills. In retrospect, such a design implementation would have better served the arguments advanced in this paper by providing a more authentic representation of this type of perceptual skill.

In closing, the following must be stated: In light of the fact that 5 of the 15 correlations, among basic perceptual skills, were moderate in strength, it appears that at least some of the basic perceptual skills may not be tapping independent cognitive processes. Much of this overlapping variance is probably attributable to the fact that many graphical elements were held constant across test items. Further, it may be that this overlap is due to the fact that underlying data points in this analysis shared the same analytical tasks. At the very least, these findings are an indication that the perceptual skills model should receive further study to better determine the autonomy of the different perceptual skills.

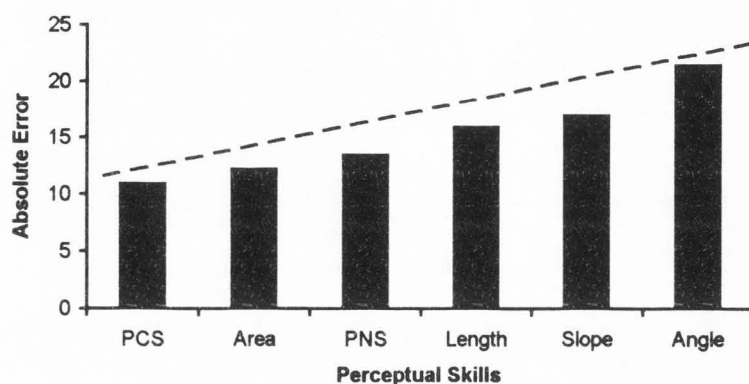


Figure 7. Depiction of more accurate slope line.

Note. PCS=Position on a common scale; PNS=Positions on nonaligned identical scales.

Analytical Tasks Collapsed Across Basic Perceptual Skills: Evidence of Unique Analyses

With regard to data relating to analytical tasks collapsed across perceptual skills, the strength of the fit between the data and the analytical tasks model is strong. Although three of the six correlations between analytical tasks (see Table 3) were statistically significant at the .05 level, these correlations were weak—the strongest of them being .28. This tends to support the idea that the different analytical tasks tap different cognitive processes.

Alternatively, the same concerns that were raised in the previous analysis must be raised again. Given that the correlations between analytical tasks were inherently tied correlations of shared perceptual skills, the lack of strong correlation coefficients may also support the autonomy of the different perceptual skills, as well as the autonomy of the analytical tasks. Once again, because of the design incorporated in this study, it is difficult to precisely determine which factor (perceptual skill or analytical task) plays the more central role in correlational associations or the lack thereof.

Nonetheless, this finding is somewhat inconsistent with informal feedback offered by subjects. After testing sessions, the experimenter debriefed subjects by explaining the analytical tasks and

perceptual skills models. It was commonplace for subjects to remark that they found global comparison and synthesis tasks to be very similar in terms of difficulty. These two tasks did have a correlation of .28, which is statistically significant at the .05 level. However, this indicates shared variance of only about 8%, which indicates that over 90% of the variance between these two tasks can be attributed to extraneous sources of error and the measuring of different cognitive processes.

Although the correlation between synthesis tasks and local comparison tasks was weak, it was statistically significant and is deserving of some discussion. Recall that midway through the experiment it was discovered that some subjects were not answering synthesis questions in the prescribed manner. About 38% of the subjects polled indicated that they were not extrapolating a fourth graph/specifier in synthesis tasks, but were instead simply using graphs/specifiers A and C to answer the question. In essence, using A and C to answer such questions amounted to a local comparison task, and this may have been the reason for a statistically significant correlation between local comparison and synthesis tasks.

There is one strong argument that could be made against such a conclusion. The correlation between the strategy employed on such tasks and task performance was extremely weak for all synthesis tasks (see Table 8). On the surface, this would appear to be an indication that the type of strategy employed was not associated with performance on synthesis tasks, which would seem to render the previous conclusion mute. However, there may have been more subjects who did not follow the prescribed instructions, but who did not indicate as much because of social desirability effects. Further, a number of subjects indicated that they went back and forth between the strategies that they used and that they were uncertain as to when and how often they used the different strategies. Subjects were told that if they could not decide which strategy they used more often, they should not respond to the question about the type of strategy employed. Only one subject left this item blank. However, judging from the number of individuals raising this concern, there should have been more subjects who left this item blank. Responses from these subjects may have been misleading. These factors may have reduced the correlations between test-taking strategy and performance on synthesis tasks.

There were some statistically significant correlations between analytical tasks, but these correlations were not practically significant. Even though it is impossible to fully determine the individual contributions of analytical tasks and perceptual skills performances to the dearth of strong correlations, the autonomy of the different analytical tasks appears to have been supported somewhat by this study.

Rank Ordering of Basic Perceptual Skills: Mixed Findings

Collapsing Data Across Analytical Tasks: Support for the Basic Perceptual Skills Model

Prior to delving into an explication of the rankings of perceptual skills and analytical tasks, there is one major issue that must be broached. The most consistent trend in the data is that of how subjects performed on items for which their response exceeded 100%. That is, for some items subjects were asked to estimate the proportion that a larger or greater specifier was of a smaller or lesser specifier. Obviously, the answer to these questions would be over 100%. What was not so obvious at the outset of this study was that there would be such distinct performance differences when these types of questions were compared with questions where subjects were asked to judge a proportion of less than 100%. In Table E.1 (see Appendix E), the last eight error scores listed derive from items where subjects were making proportional judgments that exceeded 100%. This is the main reason the results of this study should be interpreted with caution. Having addressed this redoubtable issue, the discussion can proceed.

When performances for perceptual skills were collapsed across analytical tasks, the perceptual skills model appears to be somewhat validated as a robust model for ranking the perceptual processes of graph reading. Cleveland and McGill (1986) did not include different types of analytical tasks in their study from which the perceptual skills model derives. Further, the perceptual skills model had not been tested with visual stimuli that closely approximated real-world graph reading. Yet, the ranking of the basic perceptual skills in Table 5 provides strong support for the perceptual skills model. The fact that the collapsed analysis (and not the ranking of perceptual skills within each type of analytical task) yielded results that were more consistent with the basic perceptual skills model may be an indication that the

individual items did not provide a representative sampling of each perceptual skill, as was the case when they were collapsed across analytical tasks. Before jumping to this conclusion, however, there are some issues that must be discussed.

There were two anomalies in the difficulty ranking. First, notice that in Table 4, angle is ranked as an easier judgment than slope for all but local comparison tasks. The average performance difference across the other three analytical tasks is 3.3. However, for local comparison tasks, slope is ranked as easier than angle, with a performance difference of 28.4. This tremendous disparity is most likely due to the fact that for local comparison, the slope item required a judgment of less than 100% and the angle item required a judgment of greater than 100%. Why does all of this matter to the discussion of performance collapsed across analytical tasks? The fact is that when performances for angle and slope test items are collapsed across analytical tasks, slope emerges as the easier skill, even though it was the more difficult skill in three of the four analytical tasks. There is a similar trend with regard to judgments of points on common and nonaligned scales (see Table 4).

When collapsing across analytical tasks, the findings from this study at first seem to closely approximate the predictions of the basic perceptual skills model. However, this apparent consistency between studies is an artifact of making judgments over and under 100%. Again, this underscores the influence of the 100% anchoring point.

The second anomaly was this: Overall, area judgment was ranked as the second easiest perceptual judgment to make. This is highly inconsistent with the predictions of the basic perceptual skills model, and there are several possible reasons for this finding. In all likelihood, a combination of these possibilities has had some effect on the data. The first explanation is that none of the items involving area judgment involved a proportion estimation greater than 100%, as was the case for other perceptual skills. As has been indicated, items requiring estimations of greater than 100% were more difficult for subjects. Because none of the area judgments involved an estimation that exceeded 100%, it makes sense that area would be ranked as easier than some of the other perceptual skills. Yet, it is not immediately sensible that judgment of area should be transformed from the most difficult perceptual skill to the second easiest. That

all area items involved a proportion judgment of less than 100% is probably not sufficient to have had the impact that the difficulty rankings in Table 5 might otherwise indicate.

The second possibility relates to the shape of the area being judged. The area test used in this study departed from Cleveland and McGill's in two important ways. First, Cleveland and McGill (1986) had subjects make proportion estimations about circles and even oddly shaped "blobs." As was noted by Cleveland and McGill (1984), circular and asymmetric shapes consist of areas that are more difficult to judge than the area circumscribed by a square, as was the case in this study. Second, when axes and borders are placed evenly around such a square, the cues enhancing perceptual judgment are increased. Therefore, area judgments in this study were easier to make than area judgments in Cleveland and McGill's study. The above design modifications notwithstanding, the findings are notable because they highlight the tenuous generalizability of the basic perceptual skills model in its prediction of area judgment. After all, if the model is only predictive of area judgment (relative to judgments about other perceptual skills) when the area judged is circular or asymmetric, then its usefulness will be limited.

In summary, the difficulty ranking of perceptual skills collapsed across analytical skills appears to more closely approximate the predictions of the perceptual skills model than do the rankings of perceptual skills within each analytical task. However, with respect to judgments involving area, slope, and angle, interpretation of the rankings should be undertaken with caution.

Ranking of Perceptual Skills Within Analytical Tasks: Failure to Support the Perceptual Skills Model

As was previously discussed, when perceptual skills were ranked within each type of analytical task, the perceptual skills model was not predictive of performance. This is most evident for the rankings within the point reading and global comparison tasks, in which the rankings of the different perceptual skills did not match the predictions of the model (see Table 4). Carswell (1992) maintained that the perceptual skills model poorly predicts performance difficulty for comparisons of positions (points aligned

and nonaligned), length, and angle. This appears to have been the case for point reading and global comparison tasks.

In addition to considering the ranking of all perceptual skills within a single level of analysis, it will be instructive to make independent assessments for those items that did and did not require a proportion judgment in excess of 100%. Refer back to Table 4 and note the values followed by a superscripted "a." These values denote those items for which proportions being judged were greater than 100%. The only types of analytical tasks requiring proportional judgments in excess of 100% were local comparison and synthesis tasks. Let us first examine the ranking of under-100% local comparison tasks. When we rank solely the under-100% items (on the basis of judgment accuracy associated with such items), slope emerges as an easier judgment to make than area. This is consistent with the predictions of the basic perceptual skills model. However, when we rank solely the over-100% items, length emerges as the easiest skill, followed by position on a common scale, positions on nonaligned identical scales, and angle. This ranking is inconsistent with the predictions of the basic perceptual skills model. Tables 9 and 10 should provide further clarification of this analysis where over- and under-100% items were separately considered.

Now consider the ranking of the perceptual skills within the "Synthesis" column in Table 9. There were three test items that required judgments of proportions in excess of 100. When we rank only these items, length emerges as the perceptual skill associated with the greatest judgment accuracy, but angle is ranked as easier than slope. This finding is incongruent with the predictions of the basic perceptual skills model. Likewise, when we rank solely the perceptual skills where judgments of under 100% were required, the judgment of positions along identical nonaligned scales emerges as the easiest perceptual skill, followed by position on a common scale, and area. Once again, this ranking contradicts the predictions of the perceptual skills model.

In summary, it appears as though the basic perceptual skills model is poorly predictive of performance within any one analytical task. This was true within every type of analytical task and in spite

Table 9

Mean Absolute Error Scores for Perceptual Skills Requiring Judgments in
Excess of 100%

Type of basic perceptual skill	Type of analytical task			
	Point reading	Local comparison	Global comparison	Synthesis
Position on common scale	--	20.7 (2)	--	--
Positions on nonaligned identical scales	--	31.6 (3)	--	--
Length	--	18.2 (1)	--	26.0 (1)
Slope	--	--	--	33.2 (3)
Angle	--	39.9 (4)	--	27.3 (2)
Area	--	--	--	--

Note. The values in parentheses in each column represent the actual difficulty ranking (1=easiest and 4=hardest) associated with the perceptual skills as they are ranked within each type of analytical task.

of whether proportional judgments being made were greater than or less than 100%. However, this conclusion must be interpreted with caution, inasmuch as the above analyses require a fragmented interpretation of the models under study, instead of a more holistic interpretation. Other factors must be considered.

Aside from the issues associated with the type of proportion being judged, there are other possible explanations for these findings. The first alternative explanation relates to timing. Cleveland (1985) maintained that the basic perceptual skills model was predictive of early rather than late processing. Given time, an individual attempting to decode a graph's specifiers may be able to utilize compensatory strategies to assess stimuli that might not be as easily decoded through preattentive processes. In short, it

Table 10

Mean Absolute Error Scores for Perceptual Skills Requiring Judgments of
Under 100%

Type of basic perceptual skill	Type of analytical task			
	Point reading	Local comparison	Global comparison	Synthesis
Position on common scale	5.1 (3)	--	6.9 (2)	11.0 (2)
Positions on nonaligned identical scales	4.3 (1)	--	6.8 (1)	10.9 (1)
Length	4.4 (2)	--	15.1 (6)	--
Slope	7.7 (5)	11.5 (1)	14.8 (5)	--
Angle	6.5 (4)	--	11.9 (4)	--
Area	--	14.1 (2)	10.1 (3)	11.5 (3)

Note. The values in parentheses in each column represent the actual difficulty ranking (1=easiest and 6=hardest) associated with the perceptual skills as they are ranked within each type of analytical task.

was maintained that lengthening exposure time to a graph reduces judgment differences for different perceptual skills.

In validation studies of their model, Cleveland and McGill (1984, 1986) instructed subjects to make rapid judgments about graphical stimuli presented. The authors do not indicate how quickly subjects made their responses, but a similar study by Spence and Lewandowsky (1991) sheds some light on this issue. In their study, Spence and Lewandowski allowed subjects only 1.5 seconds to make judgments about aligned bar graphs and angles (i.e., pie charts). In this study, subject performance that was most predictive of the basic perceptual skills model was that associated with the fastest response time. Conversely, in the present study, subjects were allowed as much time as they needed to make each judgment. If subjects took more time to make judgments about the more difficult perceptual skills (which is probably what

happened), then the performance differences between tasks could have been attenuated or even reversed, resulting in the precarious difficulty rankings that emerged from the data and undermining the fit between the data and the basic perceptual skills model.

Rank Ordering of Analytical Tasks: Mixed Support for Carswell's Taxonomy

The difficulty ranking of analytical skills for perceptions of length, positions on nonaligned identical scales, slope, angle, and position along a common scale must be broken down. Table 11 provides performance data solely for those items requiring a proportional judgment in excess of 100%. As can be seen, it is difficult to draw an interpretation from these data, as so few analytical tasks (within a single perceptual skill) required an over-100% judgment. Nonetheless, it should be noted that when angle judgments in excess of 100% were made, synthesis tasks emerged as easier than local comparison tasks—a trend that is inconsistent with the predictions of the analytical tasks model.

Table 11

Mean Absolute Error Scores for Analytical Tasks Requiring Judgments in Excess of 100%

Type of analytical task	Type of basic perceptual skill					
	Length	Position on nonaligned identical	Position on common scale	Angle	Slope	Area
Point reading	--	--	--	--	--	--
Local comparison	18.2 (1)	31.6	20.7	39.9 (2)	--	--
Global comparison	--	--	--	--	--	--
Synthesis	26.0 (2)	--	--	27.3 (1)	33.2	--

Note. The values in parentheses in each column represent the actual difficulty ranking (1=easiest and 2=hardest) associated with the perceptual skills as they are ranked within each analytical task.

When performance associated solely with under-100% test items is considered, there is a greater degree of model-data fit. Table 12 indicates that within almost every perceptual skill, the ranking of over-100% analytical tasks is in accord with the predictions associated with the model. On the other hand, it must be noted that for area judgments of under 100%, global comparisons emerged as easier for subjects to make than synthesis judgments, which were easier to make than were local comparisons. This finding is inconsistent with the tenets of the analytical tasks taxonomy.

The separate analysis of items requiring over- and under-100% judgments seems reasonable given the tremendous performance disparity between these two types of tasks. And such a fragmented analysis provides some indication that Carswell's (1992) taxonomy was predictive in this study. But having said this, the conclusions to be drawn from such an analysis are questionable. It was never the intent of this study to separately analyze such tasks independently of one another. Neither Carswell nor Cleveland and McGill (1984, 1986) discuss this issue. Therefore, the preceding discussion should be interpreted with caution.

Table 12

Mean Absolute Error Scores for Analytical Tasks Requiring Judgments of Under 100%

Type of analytical task	Type of basic perceptual skill					
	Length	Position on nonaligned identical	Position on common scale	Angle	Slope	Area
Point reading	4.4 (1)	4.3 (1)	5.1 (1)	6.5 (1)	7.7 (1)	--
Local comparison	--	--	--	--	11.5 (2)	14.1 (3)
Global comparison	15.1 (2)	6.8 (2)	6.9 (2)	11.9 (2)	14.8 (3)	10.1 (1)
Synthesis	--	11.0 (3)	11.0 (3)	--	--	11.5 (2)

Note. The values in parentheses in each column represent the actual difficulty ranking (1=easiest and 3=hardest) associated with the perceptual skills as they are ranked within each type of analytical task.

Given the above concerns, it is probably most useful to focus primarily on analytical tasks collapsed across perceptual skills. On the surface, it may appear that Figure 4 indicates inconsistency between the predictions of the analytical tasks model and performance in this study. However, there are additional factors that must be considered before such a rush to judgment is made. Refer back to Table 5 and note that for four of the six local comparison tasks, judgments of over 100% were required. As has been indicated, such proportion estimations are associated with a high degree of error. It is likely that local comparison tasks are ranked as most difficult because most local comparison tasks required a judgment in excess of 100%. This is unfortunate, because Carswell's methods and findings are most straightforward with respect to local comparison tasks.

In addition to the ranking taxonomy advanced by Carswell (1992), she also advanced the idea that performance differences between adjacently ranked analytical tasks are not great. She did not elaborate on this, but her data indicate that the performance difference between global comparisons and synthesis assessments is marginal. Contrary to Carswell's contentions, the graphical depiction of performance differences in Figure 4 would seem to indicate a tremendous performance disparity between synthesis and global comparison tasks. But even though synthesis tasks were more difficult than global comparison tasks (as would be predicted by the model), the noted performance disparity between the two tasks is probably misleading. Recall that there were no global comparison tasks requiring a proportion estimation above 100%, but three of the six synthesis tasks required the estimation of such a proportion. Given that the estimation of such a proportion is decidedly more difficult than the estimation of a proportion of less than 100%, it is easy to see how the size of the performance difference between global comparison and synthesis tasks could have been inflated due to the proportions being estimated and not the analytical tasks themselves.

Further, the data in Figure 4 do not include performance data for point reading tasks involving area judgment. It is likely that if a valid depiction of such a task could have been devised and included in the test, performance on such an item would have been characterized by high error. If such had been the

case, the performance disparity between point reading and local comparison tasks would have been attenuated.

In summary, when data were collapsed across perceptual skills, the findings of this study did not confirm the predictions of the analytical tasks taxonomy. However, it appears that when tasks requiring a judgment in excess of 100% and less than 100% are assessed separately, the ranking of analytical tasks within four of the six basic perceptual skills is consistent with the predictions of the analytical tasks model. Inconsistency is manifest in the rankings associated with over-100% angle judgments and judgments involving area. Nevertheless, it is difficult to offer conclusive insights about the extent of congruence or incongruence, given the interpretation problems introduced by the effects of the under-/over-100% proportions being judged.

Judgments of Proportions Above and Below 100%

The most pronounced trend in the data is the fact that test items requiring a proportion judgment of greater than 100% were associated methodically with more judgment error than those items requiring a judgment of less than 100%. In spite of the analytical task or the type of perceptual skill, items involving judgments in excess of 100% were more difficult than those requiring a judgment of less than 100% (see Table E.1). This is a significant finding, because in this experiment the models that have received considerable empirical attention were less influential on performance than was this *100% split*.

Another compelling question that must be addressed is one of whether the analytical tasks and perceptual skills models are more or less predictive depending on whether judgments of proportions above or below 100% are being made. Refer back to Figure 3 and separately assess the difficulty ranking for items requiring judgments above and below 100%. Notice that in either case, the rankings of perceptual skills are almost totally at odds with the predictions of the basic perceptual skills model. On the other hand, Table 5 indicates that most of the analytical tasks are ranked correctly (with the obvious exception of analytical tasks ranked for area judgments) in spite of whether proportions being judged were over or under 100%. On the surface, it seems as though the 100% dividing line does not serve to reduce or

increase the predictive power of either model. Unfortunately, with so few tasks incorporated into the ranking, it is difficult to draw a coherent conclusion about this issue. More research is needed to gain a clearer picture of this issue.

An important and related trend that Croxton and Stryker (1927) and Simkin and Hastie (1987) commented on was that of anchoring points. It has been suggested that the most influential anchoring point is manifest in a 1-to-1 ratio or, in terms of this experiment, a proportional judgment that most closely approximates 100%. Recall that the results of this study showed that for every task where subjects needed to make a proportion judgment in excess of 100%, subjects (as a group) underestimated the proportion. Conversely, for every item where subjects needed to make a proportion judgment of less than 100%, subjects (as a group) overestimated the correct proportion. This is another indication that the 100% split is a highly influential, perceptual factor in the reading of graphs.

Explication of Male and Female Graph Reading Equivalencies

Assessment via independent-groups *t* tests did not yield a single statistically significant nor a single practically significant difference between males and females, with regard to their test performance. It is difficult to ascertain whether this finding is consistent with the findings of graphicacy studies in general, given that Eells' (1926) study was the only one that publicized data relating to gender differences in graph reading performance. Nonetheless, Eells did not find statistically significant performance differences for males and females in graph reading tasks focusing on bars and pie charts. The fact that relatively few graphicacy studies have publicized data relating to gender differences in performance is compelling. The main cognitive processes implicated in graph reading are those where males have traditionally been viewed as having a natural advantage over females. These processes are analysis, spatial perception, and quantitative calculation. Because the graphicacy literature has not expanded on this issue in any detail, it is appropriate at this time to compartmentalize the different processes used in this study and attempt to determine whether gender-performance trends found in this study are consistent with past studies in similar domains.

The types of graph reading tasks required different levels of analysis. As Carswell (1992) has suggested, these levels of analysis represent varying levels of difficulty. These levels of difficulty vary along two dimensions: (a) the number of specifiers that must be attended to in order to answer a question, and (b) whether an item requires a comparison of graph features actually presented or a comparison involving some cognitively extrapolated or hypothetical standard.

Second, the items in this study required the processing of visual-spatial characteristics. It has typically been thought that males have a performance advantage when it comes to such processing tasks. In fact, a number of studies have reported gender-based performance differences in mental rotation tasks. Bouchard and McGee (1977), Maccoby and Jacklin (1974), Sanders et al. (1982), and Tapley and Bryden (1977) are just a few of the researchers who have utilized mental rotation tasks and come to the conclusion that males have a performance advantage. The problem with discussing the findings of such studies in terms of the current research is that mental rotation tasks require subjects to make a comparison between a rotated figure and its standard form, and to decide whether the rotated figure is the same as its standard or the mirror image.

This seems like a different task than the graphicacy tasks implemented in this study, and there is an important distinction to be made here. McGee (1979) maintained that visual-spatial tasks may consist of a *visualization factor*, an *orientation factor*, or both factors. The visualization factor includes the imaging of the rotation or unfolding of objects, as well as the imaging of depth. The orientation factor includes the ability to judge relationships and patterns. The test items in this study clearly were more closely associated with McGee's orientation factor.

The most recent studies and reviews of visual-spatial processing have commented on the gender issue, with deference to McGee's (1979) ideas. Authors such as Halpern (1986), Linn and Hyde (1989), and Newcombe and Baenninger (1990) have more recently concluded that insofar as visual-spatial abilities are concerned, males have a consistent performance advantage in tasks involving mental rotation of three-dimensional objects, the navigation through cognitive maps, and the mental figuring of trajectories. All of these types of tasks consist primarily of McGee's visualization factor. Conversely, these

performance differences do not exist when tasks are based more on orientation factors, as was the test in the current study.

Additionally, the manner in which the instructions were administered prior to this test may have reduced gender differences in performance. There are some indications that the nature of an instructional protocol may increase or diminish gender-related expectancy effects. That is, females who have come to believe that they have less spatial-processing ability may expect to perform poorly on a test, when the instructions preceding the test accentuate its spatial characteristics. Sharps, Welton, and Price (1993) employed a mental rotation task in conjunction with varying the content of administered instructions. For one group of subjects, instructions were imbued with spatial terms, such as "distance" and "orientation." Another group of subjects received "nonspatial" instructions. For the "spatial" instruction group, males outperformed females, but for the nonspatial instruction group, no sex differences were observed. Sharps, Price, and Williams (1994) replicated the instructional component of the previous study and again found no performance differences between males and females when the spatial nature of the task was not emphasized in the experimenter's instructions.

In the current experiment, an equal amount of instructional time was devoted to explicating both the spatial and analytical components of the test items. For example, after attempting practice item D, subjects were told, "The answer to the question is 159.5%. Does everyone understand how I arrived at that answer? Again, to answer this type of question, you just mentally estimate the distances between the horizontal axis and each point. Then you mentally estimate how much larger point C is than point A. In this example, point C is 159.5% of point A." Notice that both the spatial and analytical characteristics of the task are emphasized. It is possible that because the spatial characteristics of these tasks were not the primary concern, the instructions were less likely to encourage gender-stereotyped, expectancy effects in female subjects, thereby diminishing performance differences between males and females.

Third, the tasks comprising this test required an understanding of quantitative relationships. That is, subjects were not simply asked to identify the greatest or least specifiers presented. This would have amounted to a simple spatial task. Instead, subjects were asked to convert their judgment of the spatial

relationship between certain specifiers into a percentage. For point reading tasks, the extent of calculation required was negligible. However, in order to respond to local comparison, global comparison, and synthesis tasks, it would have been necessary for a subject to use some combination of division, addition, subtraction, and (especially where synthesis tasks were at issue) even multiplication. Therefore, mathematical abilities figured prominently into performance on this test.

The literature focusing on gender differences in math performance has become increasingly clear over the last few years. Halpern (1986), Maccoby and Jacklin (1974), and Flake, Loyd, and Hoover (1981) are often cited for their early findings supporting the notion that males have an advantage in mathematics. However, Hyde, Fennema, and Lamon (1990) have indicated that even though the gender gap in math may have been larger at one time, the gap is now virtually nonexistent. In their excellent meta-analysis, the authors included 100 studies, representing nearly 4 million students from various socioeconomic, grade, and intelligence levels, as well as different cultural backgrounds. For studies published in or prior to 1974, the researchers obtained a standardized mean difference effect size of .14, which represents the performance advantage that high-school males have over females. This performance difference is extremely small. Further, there is recent evidence that when male and female high school students take the same number of and type of math classes, the achievement differences of males and females diminish or disappear altogether (Oakes, 1990). Finally, several authors, including Eccles (1989) and Hyde et al. (1990), have found that when calculation (one of the fundamental abilities inherent to mathematics) is analyzed by itself, the gender gap is reduced to less than a .05 effect size.

If the quantitative characteristics of test items resulted in a performance advantage for male subjects, then such an advantage should have emerged for the most quantitatively complex items—synthesis or global comparison tasks wherein the proportions being estimated exceeded 100%. That is, if there was such an advantage, male subjects should have performed more efficiently on items that required the most calculating. Statistical analyses indicated that this was not the case.

In summary, it appears as though the findings of the current investigation are consistent with other related studies and reviews that have provided commentary on gender effects. Where analytical,

visual-spatial, and quantitative task characteristics are concerned, the relevant literature indicates that males do not have a consistently significant performance advantage over females. With respect to each of these characteristics, as they were manifest in the various items on this test, congruent findings have been obtained in this study.

General ACT Scores, GPA, and Graph Reading

Researchers (e.g., Dunn, Griggs, Olson, & Beasley, 1995; Kampwirth & Bates, 1989; Kavale & Forness, 1987) who have discussed the concept of learning styles have maintained that different individuals have different cognitive processing strengths. For example, some individuals are more efficient at processing visual information while others are more efficient at processing verbal information. Dunn, Dunn, and Perrin (1994) have concluded that the format in which information is presented will have much to do with how that information is understood and retained.

The relationship between ACT Composite scores and GPA was moderately strong. The two shared about 25% variance. It is likely that much of the unexplained variance is due to the fact that the sample used in this experiment consisted primarily of freshman university students. Allow me to explain. Generally speaking, ACT scores and GPA can be taken as good measures of intellectual ability. However, there is one concern that must be addressed. Because GPA is (to a certain extent) dependent upon the types of classes one takes and how quickly one acclimates to university-level expectations, it may not be the best measure of freshman intellectual ability—at least insofar as intellectual ability has traditionally been conceptualized. This is the most likely explanation for the finding of only a moderate correlation between ACT score and GPA. It is also the best argument for focusing discussion on the correlations between ACT score and test performance instead of GPA and test performance.

With regard to basic perceptual skills collapsed across analytical tasks, the relationship between ACT scores and test performance is compelling. What is of interest is the fact that there were only two statistically significant correlations, and these were small. This finding probably has much to do with what is measured by the ACT Composite score and what was measured by the test used in this study.

The ACT Composite score comprises English (i.e., grammar skills), science reasoning (i.e., analytical skills), math (i.e., quantitative skills), and reading (i.e., comprehension skills) subtests. Although these subtests provide for a good sampling of intellectual skills (as they have been traditionally conceptualized), they do not directly test visual-spatial abilities.

As far as the ACT is concerned, it may be true that test items are presented visually to examinees. However, the problems do not focus on visual-spatial issues. In this study, not only was information presented visually, but the problems confronting subjects were primarily visual-spatial in nature. Graphicacy researchers may need to look to other tests of ability to make more meaningful predictions related to graph reading.

CONCLUSION

Summary

The findings of this study are mixed. On the one hand, the analytical tasks and perceptual skills models appear to have predicted performance difficulty with only a moderate degree of precision. Much of this reduction in predictive precision can be traced to performance differences associated with proportion judgments above and below 100%. With regard to the independence of individual analytical tasks and perceptual skills and the combinations thereof, the findings (as a whole) are difficult to interpret. There were only a handful of practically significant correlations between individual test items. This is impressive and tends to support the idea that the different test items were tapping different cognitive processes. The problem with staying with this conclusion is the fact that judgments of over and under 100% may have yielded some very spurious performance data. Given this fact, it is probably more appropriate to draw conclusions from analyses of analytical tasks (collapsed across perceptual skills) and perceptual skills (collapsed across analytical tasks).

With regard to perceptual skills, there are a number of statistically significant and practically significant correlations, suggesting that the perceptual skills may not be highly differentiated. However, the correlations between collapsed perceptual skills may be due in part to the fact that different collapsed skills shared the same analytical tasks. With regard to analytical tasks, the independence of the tasks was demonstrated to some extent, inasmuch as the practical significance of even the strongest correlations was small. Three of the six correlations were statistically significant but were also less than .3 in strength. The data underlying the difficulty rankings of perceptual skills and analytical tasks were supportive of Carswell's (1992) taxonomy but not of Cleveland and McGill's (1986). In either event, ranking inferences should be made cautiously given the effects of the 100% split.

The data relating to gender and performance are consistent with the latest research relating to gender-based performance differences in mathematics and analytical skills. That is, there were no statistically or practically significant, gender-based performance differences for virtually all types of tasks.

Finally, it was found that intelligence measures (ACT Composite and GPA) are not predictive of performance on this test. This is most likely due to the fact that these traditional measures of intellectual ability do not test visual perception in the same manner as this test.

Taken as a whole, the findings from this study are only moderately supportive of the analytical tasks and basic perceptual skills models. This is probably due in large part to the fact that so many graphical elements were held constant and subjects were allowed as much time as they needed to answer test items. This likely resulted in higher correlations and difficulty rankings that were inconsistent with model predictions. Ultimately, the findings serve to vindicate the importance of this study. As was asserted at the outset of this study, the reading of graphs is a complex process. In this study alone, there were many explanations offered as tenable alternatives for the findings contained herein. The number of defensible explanations for these findings highlights the need to more carefully study graph analysis and perception.

Limitations of the Study

There are some critical limitations in this study, some of which have and have not been noted previously. First, although one of the claims made at the outset of this paper was that the instrument used in this test was more practical than those used by Cleveland and McGill (1986) and those reviewed by Carswell (1992), there were certain impractical components inherent to this test. For example, test questions and the graphs themselves did not involve any concrete concepts. That is, the questions asked about relationships between bars or points, and not (for example) about relationships between the gross national products of two different countries. It is possible that increasing the meaningfulness of the quantities represented in the graph would increase an examinee's interest, thereby improving performance.

Another issue related to practicality is the fact that subjects in this study were able to take as much time as they needed to complete each test item. As was noted previously, this may have been a factor that led to a difficulty ranking of perceptual skills that was slightly different than the one proposed

by the basic perceptual skills model. More importantly, there are many situations where the reader of a graph must do so within a certain time limit that might force her to work more quickly than she otherwise would. The findings of this study may not be as readily applicable to such instances.

A highly influential factor in this study was the magnitude of the proportion being judged. When proportion judgments of greater than 100% were being made, subjects performed more poorly. The findings of this study are insufficient to determine whether this trend was due to anchoring effects, numeracy effects, or some other variable. Further, the 100% phenomenon led to analyses and conclusions that were incomplete or that should be interpreted cautiously.

Issues related to the manner in which task independence was established must also be addressed. Cohen's (1988) generic guidelines for identifying practically small, moderate, and strong correlations were used. But Cohen's yardstick was not designed to be a catch-all standard. In fact, Cohen (1988), Glass and Hopkins (1996), Hinkle, Wiersma, and Jurs (1979), and others have maintained that the use of Cohen's guidelines should be superseded if the literature avails itself of different standards deriving from a line of inquiry that has systematically built upon itself. Unfortunately, no similar studies could be located that could have provided a foundation on which this analysis of construct overlap could more solidly rest. The generic usage of Cohen's guidelines necessitated the use of .3 as a criterion for dividing weak and moderately strong correlations. Therefore, even if correlations of .28 and .32 were side by side, the former was considered practically insignificant while the latter was taken as an indication of a practically significant trend. This is an important issue because most of the moderately strong correlations in this study were not much stronger than .3. Because so many correlations were just above or below .3, the cases made for or against key theories (on the basis of practical significance or a lack thereof) are weakened.

Finally, only one item was used to represent each type of analytical task/perceptual skill. Each graph comprised value specifiers that were between 30 and 60. This is important because it eliminated the effects of some anchoring positions. However, because of this implementation, it would be difficult to claim that the findings herein should be generalized to all graph reading situations. Currently, it is not known whether the two taxonomies become more or less predictive of graph reading as proportions and

values to be judged are moved closer to or farther from critical anchoring points/positions. It may be that the basic perceptual skills model is less predictive of tasks in which specifiers closely approximate anchoring points. This study was not designed to determine this. It is possible that the findings herein apply to a limited range of values coded within specifiers.

Implications of This Study

Regarding future study of the analytical tasks and basic perceptual skills models, there were three conspicuous findings that emanated from this study. The most prominent of these is related to authenticity. From the outset of this study, the author was critical of Cleveland and McGill's (1984, 1986) validation of their model through the use of graphs that did not closely approximate real-world graphs. The author was also critical of Carswell's (1992) study of the basic perceptual skills model and analytical tasks taxonomy, in which she used data from studies of multicue information displays and visual-perceptual, decision-making tasks to draw certain inferences about graph reading. These criticisms diminish confidence in the conclusions drawn from such studies.

It has already been suggested that the more systematic use of practical graph construction would erode the predictiveness of these models, and regarding the basic perceptual skills model, this may have been the case. Given the amount of performance overlap between perceptual skills (collapsed across analytical tasks) and the ranking anomalies associated with them, one should be cautious about using Cleveland and McGill's (1984, 1986) model to drive discussions of graph reading in all real-world situations. Their model may be helpful as a loose heuristic, but it cannot be used as an absolute explanation for graph reading in general.

On the other hand, Carswell's (1992) model received support on two key fronts: task independence and difficulty rankings. Regarding both of these issues, the findings from the current study are moderately congruent with the predictions of the analytical tasks taxonomy.

In light of the findings from this study and the discussions advanced herein, the place of Cleveland and McGill's (1984, 1986) model in the graphicacy literature becomes difficult to surmise. The

model derives from graphical stimuli that are a hybrid of pure visual-perceptual tasks and real-world graphicacy tasks. Which domain is the model best suited for? Perception? Graphicacy? Unfortunately, the answer may be neither. Granted, graph reading requires visual perception, but researchers have been inclined to discuss these overlapping processes in different ways. Given that this is the case, the logical fit of Cleveland and McGill's conceptions within either domain is questionable. The basic perceptual skills model is caught between these two areas of interest, and as such, its explanatory power is weakened with regard to both of them. This assertion further underscores the importance of this study as a signal to graphicacy researchers to focus on authenticity, so as to avoid being caught in the purgatory of trying to serve two theoretical ends. What is needed are conceptions or definitions of authenticity that more graphicacy researchers utilize. Only after such definitions have been developed, should the broader issues of perception and analysis be addressed. It is the author's position that this fundamental issue should have been broached decades ago and that it is one of the causes of the fragmented state of the graphicacy literature. With more graphicacy research converging on efforts to define authenticity (in its various forms) and use such authenticity to direct instrument design, the field will be empowered. The comparability between studies will be increased and graphicacy researchers will be able to focus on the all-important issue of practical application, which should be the driving concern for this field.

All of the foregoing notwithstanding, the basic perceptual skills model should not be thrown out. Instead, Cleveland and McGill's (1984, 1986) model should be used as a springboard for future investigations of perceptual skills that are implicit in graph reading. The current study provides for a useful departure from the basic perceptual skills model, the limitations of which were bound to surface once it was more widely tested.

The second issue that is critical to the future of the graphicacy research was that of unique cognitive processes. More specifically, the amount of performance overlap between certain basic perceptual skills leaves one with the feeling that the model is perhaps too ambitious. As has been mentioned, Cleveland and McGill (1986) maintained that for their purposes, the defining of autonomous

perceptual skills was not an issue. Their model is driven more by the ranking of perceptual skills based on judgment accuracy.

The problem with this approach is that it is conceptually backward. When discussing the difficulty with which different perceptual processes are made, one first must establish whether the perceptual processes are truly different. Look back at Table 2 and notice the correlations between angle and other types of perceptual skills. All five correlations between angle and other perceptual skills are statistically significant. Three of these correlations are stronger than .3. The median of this group of correlations was .31, as was the mean derived using Glass and Hopkins' (1996) recommended method for weighting, transforming, and averaging correlation coefficients. A correlation of .31 is somewhere between small and moderate in strength. This is an indication that much of the information provided by angle judgments may be redundant, in light of the information being provided by other judgments related to the basic perceptual skills. Nonetheless, angle judgments are an important part of many graphs. So, it would be irresponsible to omit them entirely from the basic perceptual skills model. A more parsimonious conception of the basic perceptual skills model could feature angle and slope judgments (that correlated at .38) as a single skill within the model.

To this end, it is proposed that angle and slope judgments could be subsumed into a single category referred to as: *inclined plane*. The term captures the essence of angle and slope judgments that essentially require the graph reader to assess the graded degree to which two planes or lines diverge from one another in two-dimensional space. More importantly, discussing angle and slope judgments as manifestations of the same perceptual skill returns the basic perceptual skills model to the course where it should have been heading. Aside from the alternative explanations for the performance overlap between collapsed perceptual skills, these skills ultimately must be discussed in terms of the constructs or processes underlying them and not the information being processed. Simply because the visual-spatial arrangement of angles, slopes, or points along a common scale appears to be different on paper, this does not mean that such stimuli are processed any differently by human beings. Cleveland and McGill (1984) have discussed the similarity between perceptions of angle and slope but failed to unite them within the same skill. The

findings of this study suggest the need for the consolidation of terms within Cleveland and McGill's model. Future efforts to more closely investigate the basic perceptual skills model would do well to further determine whether other perceptual skills should be similarly consolidated.

The final issue that will bear on future graphicacy studies is the 100% split, which seems to clearly divide tasks associated with less and more accurate judgments. To date, there has been no mention of this split in the literature concerned with graphicacy or visual perception. This is highly intriguing, because this split was more predictive of high and low performance than were the two models primarily focused on. Recall that four of the six local comparison tasks involved judgments of proportions that exceeded 100% (see Table 4). This appears to have been the main reason that local comparison judgments (as a whole) were ranked as the most difficult analytical task in this study. It could not be determined whether the 100% split occurs because of numeracy/calculation or perceptual problems, but for graphicacy researchers, this is not the most important issue. What is of paramount consequence is determining how the 100% split affects the predictiveness of the two models focused upon in this paper. In all likelihood, continued study of the 100% split will reveal further limitations of the analytical tasks and basic perceptual skills models. Furthermore, such a focus is apt to reveal that separate graphicacy models will need to be constructed for judgments of proportions above and below 100%. In light of the effect of the 100% split, it is difficult to believe that any one model will be able to account for graph reading in all situations. The amount of variability between the values of different graphs is simply too great.

Directions for Future Research

The first step beyond this study should be a replication. This study was the first of its kind—a primary research study where analytical and perceptual cognitive processes were jointly considered under the rubric of established taxonomies. To this end, the most important endeavor would be to develop a more systematic usage of over-/under-100% judgments in test items. The most probable outcome of such a study would be to determine how accurately the analytical and perceptual models predict proportion

judgments above and below 100%. The findings of this study indicate that both types of judgments may lead to similar difficulty rankings, but this cannot be concluded with any absolute certainty at this point.

As has been indicated, it is difficult to determine whether numeracy (i.e., mathematical ability) figures prominently into the difficulty associated with making judgments about proportions that are in excess of 100%. The current literature and the results of this study do not provide any indication as to whether this difficulty is due to numeracy or perceptions being pulled toward anchoring points, which is an equally viable alternative explanation. Future studies should look more closely at the effects of mathematical ability as it bears on different types of graph reading skills. In particular, it would be important to know whether the performance disparity associated with making graphical judgments about proportions above and below 100% is due more to perceptual or numeracy factors. This will be a difficult undertaking, considering the close relationship between mathematical and spatial abilities. Nonetheless, a useful starting point for such research might simply be to look at the correlation between scores on the math component of the ACT and performance in the over- and under-100% test items.

Previously, it was noted that the ACT is measuring abilities that are different than those used to perform in this study. In light of the different constructs being tested by the ACT and this graphicacy test, it would be beneficial for future researchers to concentrate their efforts on the relationship between graph reading and visual learning style. There are numerous learning styles instruments that measure visual ability. Examples of such tests include: the *Learning Style Profile* (NASSP, 1986), *Swassing-Barbe Modality Index* (Barbe & Swassing, 1988), *Learning Style Inventory* (Dunn, Dunn, & Price, 1982), and *Trio* (Van Dusen, Spach, Brown, & Hansen, in press).

Finally, a taxonomy of authenticity must also be developed for research in the field of graph reading. This study has demonstrated that the use of more authentic graphical components and holding them constant across graphs may have much to do with the structure of graph reading models. To date, there have been no attempts to discuss graphical elements that are most representative of graph reading in applied settings. Such a taxonomy would help to provide a cohesive framework for future graphicacy research.

A CAUTIONARY FABLE FOR USERS OF GRAPHS

Early in the study of graphicacy there was a tendency for the typical researcher to advocate only one or two types of graphs for the effective display of data. However, there is an undeniable danger in becoming too comfortable with a particular type of graph or explanatory theory, for undoubtedly there will come a time when that graph or theory will be insufficient to convey or explain important information. In a subtle but very real way, the ideas set forth in this paper were driven by the philosophy that graphs serve us best when we use the right one for the right job. The following fable is intended to animate this notion.

It was a blistering summer evening on the Serengeti. At the end of each working day, animals of every sort and variety would run or scurry or stampede to their favorite watering holes. On this particular day, in a patch of field surrounded by ancient, desiccated trees, an allogiter (which is a white, fluffy toothless, clawless animal, roughly the same size and shape of an alligator) and a giroffe walked into a bar. In this particular pub, patrons were required to be of a certain height if they were to be served drinks.

"This is a very peculiar rule, indeed," remarked the giroffe to the allogiter.

"Strangely peculiar," echoed the allogiter.

Well, neither the allogiter nor the giroffe wanted to cause any distress, so they amicably agreed to subject themselves to this rather strange ritual. But as soon as they resolved to allow themselves to be measured, a calamity ensued, for in order to be served drinks, one needed to be at least two and a half feet in height, and the allogiter was no taller than two feet. Everyone knew this because the top of the allogiter's head did not reach a line on the wall that was painted precisely two and a half feet above the floor.

"You may stay," offered a gorilla, who looked to be the owner of this fine establishment, "but you may not order anything to drink."

The allogiter and giroffe had never hear of such a thing. The allogiter shot back, "Surely, you can see that if you measured me from the tip of my tail to the tip of my snout, I would be two and a half feet and more. Why don't you measure me thusly?"

At this point, a prairie rabbit, who had been minding her own business in a dark corner of the room, jumped into the conversation. "How could we do that?" she asked incredulously. "Do you not see where the mark is? It simply could not be done! This is how we know what you are and what you are not. If I were to move the mark up three feet, and you were as tall as that, why then, I would know you were a monkey. And if I were to move the mark down just a little more than two feet, and you were as tall as that, then everyone would know that you were a king salamander. So, you see, if we were to move the mark down to the floor, then you would be a flea. And that just wouldn't do!" And with that, the prairie rabbit punctuated the end of her monologue by slamming her mug down on the table in front of her.

The giroffe didn't fully understand what had just been said, but it certainly sounded as though they had reached an impasse. Never before had the allogiter and the giroffe been so offended. Without saying another word, they quickly turned around and left the bar in search of another watering hole.

The evening hours came and went, and finally the prairie rabbit decided it was time to repair to her burrow. On her way home, the prairie rabbit passed by a muddy river, as was her custom. On this particular night, the moon's brightness lit the way almost as well as the sun at noon. The path wandered closer and closer to the river until the mud from the river's edge mingled with the mud from the trail. Suddenly and seemingly with the speed and grace of a gazelle, an alligator leapt from the mud. The prairie rabbit froze. Even with the moon's luminance, she could not see the alligator very well as he was camouflaged with the mud from the river and the mud from the trail.

The alligator offered a toothy grin and a rumbling "Good evening," which sounded more like an earthquake than an attempt at conversation.

The rabbit, still struggling to see this interloper, returned the salutation. And, with her bravado mounting, she slowly approached the stranger who somehow seemed to be hiding in the penumbra of the brightest moon that she had ever seen. "How are you on this fine evening?" the prairie rabbit asked in an

attempt to find out more about this mysterious new friend. There was no response. The prairie rabbit continued to advance. At last, she looked into the alligator's face—a collection of brown wrinkles and scales and teeth. Again she queried the alligator, "How are you on this fine evening?"

"I am hungry," replied the alligator with all the purpose and conviction of an angry monsoon. The alligator continued, "You do not run away. Do you not know who I am?"

"Why, yes. I believe I know who you are," the rabbit returned confidently. "But just so I can be certain, would you please move over by that tree?" This was a peculiar request, thought the alligator; but now, his curiosity had gotten the better of him. So, he complied with the prairie rabbit's wishes. The rabbit immediately proceeded to scratch a mark on a desiccated tree, just above the alligator's head. Finally, the rabbit proclaimed, "Why, you're no more than two feet tall! Of course! You're that creature from the bar!" And then, with a surge of superiority and condescension, she added, "Had to go down to the river to do your drinking did you? Where's that tall friend of yours?"

And of course, you guessed it. That was the last question that the prairie rabbit ever uttered, for you know what happens when height is your only measure of others.

In this fable, the reader is brought to an understanding that certain types of measurement are appropriate within a certain context and given certain conditions, but when the context and conditions change, the static measurement device may be ineffective and ultimately lead to wrongheaded conclusions. Similarly, if the tenets of the basic perceptual skills model, in its current form, continue to be embraced, it may do more harm than good—taking researchers away from authentic graph reading, which the current model seems poorly equipped to handle.

REFERENCES

- American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. American Psychologist, 47, 1597-1611.
- Baird, J. C. (1970). Psychophysical analysis of visual space. New York: Pergamon.
- Baird, J. C., & Noma, E. (1978). Fundamentals of scaling and psychophysics. New York: Wiley.
- Barbe, W. B., & Swassing, R. H. (1988). Teaching through modality strengths: Concepts and practices. Columbus, OH: Zaner-Bloser.
- Barnett, B. J., & Wickens, C. D. (1988). Display proximity in multicue information integration: The benefits of boxes. Human Factors, 30, 15-24.
- Bertin, J. (1981). Graphics and graphic information processing (W. J. Berg and P. Scott, Trans.). New York: Walter de Gruyter.
- Bertin, J. (1983). Semiology of graphics (W. J. Berg, Trans.). Madison, WI: University of Wisconsin Press. (Original work published in 1973)
- Bouchard, T. J., Jr., & McGee, M. G. (1977). Sex differences in human spatial ability: Not an x-linked recessive gene effect. Social Biology, 24, 332-335.
- Carswell, M. C. (1992). Choosing specifiers: An evaluation of the basic tasks model of graphical perception. Human Factors, 34, 535-554.
- Carswell, M. C., & Wickens, C. D. (1987). Information integration and the object display: An interaction of task demands and display superiority. Ergonomics, 30, 511-527.
- Cleveland, W. S. (1985). The elements of graphing data. Monterey, CA: Wadsworth.
- Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American Statistical Association, 79, 531-554.
- Cleveland, W. S., & McGill, R. (1986). An experiment in graphical perception. International Journal of Man-Machine Studies, 25, 491-500.

- Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum.
- Coury, B. G., Boulette, M. D., & Smith, R. A. (1989). Effect of uncertainty and diagnosticity on classification of multidimensional data with integral and separable displays of systems status. Human Factors, 31, 551-569.
- Croxtan, F. E. (1927). Further studies in the graphic use of circles and bars II: Some additional data. Journal of the American Statistical Association, 22, 36-39.
- Croxtan, F. E., & Stein, H. (1932). Graphic comparison by bars, squares, circles and cubes. Journal of the American Statistical Association, 27, 54-60.
- Croxtan, F. E., & Stryker, R. E. (1927). Bar charts versus circle diagrams. Journal of the American Statistical Association, 22, 473-482.
- Culbertson, H. M., & Powers, R. D. (1959). A study of graph comprehension difficulties. Audio-Visual Communication Review, 7, 97-100.
- Dunn, R., Dunn, K., & Perrin, J. (1994). Teaching young children through their individual learning styles: Practical approaches for grades K-2. Boston, MA: Allyn & Bacon.
- Dunn, R., Dunn, K., & Price, G. E. (1982). Productivity environmental preference survey. Lawrence, KS: Price Systems.
- Dunn, R., Griggs, S. A., Olson, J., & Beasley, M. (1995). A meta-analytic validation of the Dunn and Dunn model of learning preferences. Journal of Educational Research, 88(6), 353-362.
- Eccles, J. S. (1989). Bringing young women into math and science. In M. Crawford & M. Gentry (Eds.), Gender and thought: Psychological perspectives (pp. 38-57). New York: Springer-Verlag.
- Eells, W. C. (1926). The relative merits of circles and bars for representing component parts. Journal of the American Statistical Association, 21, 119-132.
- Enrick, N. L. (1972). Effective graphic communication. Princeton, NJ: Auerback.
- Feliciano, G. D., Powers, R. D., & Kearl, B. E. (1963). The presentation of statistical information. Audio-Visual Communication Review, 11, 32-39.

- Fisher, R. A., & Yates, E. H. (1974). Statistical tables for biological, agricultural, and medical research (6th ed.). London, UK: Longman.
- Glass, G. V., & Hopkins, K. D. (1996). Statistical methods in education and psychology (3rd ed.). Boston, MA: Allyn & Bacon.
- Goettl, B. P., Kramer, A. F., & Wickens, C. D. (1991). Integrated displays and the perception of graphical data. Ergonomics, *34*, 1047-1063.
- Goldsmith, T. E., & Schvaneveldt, R. W. (1984). Facilitating multicue judgments with integral information displays. In J. Thomas & M. Schneider (Eds.), Human factors in computer systems (pp. 123-141). Norwood, NJ: Ablex.
- Gregory, R. L. (1966). Eye and brain: The psychology of seeing. New York: McGraw-Hill.
- Halpern, D. F. (1986). Sex differences in cognitive abilities. Hillsdale, NJ: Erlbaum.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1979). Applied statistics for the behavioral sciences. Boston, MA: Houghton Mifflin.
- Hollands, J. G., & Spence, I. (1992). Judgments of change and proportion in graphical perception. Human Factors, *34*, 313-334.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. Psychological Bulletin, *107*, 139-155.
- Kampwirth, T. J., & Bates, M. (1989). Modality preference and teaching method: A review of the research. Academic Therapy, *15*(5), 597-605.
- Kavale, K. A., & Forness, S. R. (1987). Substance over style: Assessing the effectiveness of modality testing and teaching. Exceptional Children, *54*, 228-239.
- Kosslyn, S. M. (1989). Understanding charts and graphs. Applied Cognitive Psychology, *3*, 185-226.
- Kosslyn, S. M. (1994). Elements of graph design. New York: W. H. Freeman.
- Kruskal, W. H. (1982). Criteria for judging statistical graphics. Utilitas Mathematica, *21*, 283-310.
- Legge, G. E., Gu, Y., & Luebker, A. (1989). Efficiency of graphical perception. Perception & Psychophysics, *46*, 365-374.

- Lewandowsky, S., & Spence, I. (1990). The perception of statistical graphs. Sociological Methods and Research, 18, 200-242.
- Linn, M. C., & Hyde, J. S. (1989). Gender, mathematics, and science. Educational Researcher, 18, 17-27.
- Lohse, G., Walker, N., Biolsi, K., & Rueter, H. (1991). Classifying graphical information. Behaviour & Information Technology, 10, 419-436.
- Maccoby, E. E., & Jacklin, C. N. (1974). The psychology of sex differences. Stanford, CA: Stanford University Press.
- MacDonald-Ross, M. (1977). How numbers are shown: A review of research on the presentation of quantitative data in texts. Audio-Visual Communication Review, 25, 359-409.
- McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. Psychological Bulletin, 86, 889-918.
- National Association of Secondary School Principals. (1986). Learning style profile. Reston, VA: Author.
- Newcombe, N., & Baenninger, M. (1990). The role of expectations in spatial test performance: A meta-analysis. Sex Roles, 16, 25-37.
- Oakes, J. (1990). Opportunities, achievement, and choice: Women and minority students in science and math. Review of Research in Education, 16, 153-162.
- Orwell, G. (1949). 1984. New York: Harcourt, Brace, Jovanovich.
- Paivio, A. (1975). Coding distinctions and repetition effects in memory. In G. H. Bower (Ed.), Psychology of learning and motivation (Vol. 9, pp. 72-89). Orlando, FL: Academic Press.
- Paivio, A. (1986). Mental representations: A dual coding approach. New York: Oxford University Press.
- Peterson, L. V., & Schramm, W. (1955). How accurately are different kinds of graphs read? Audio-Visual Communication Review, 2, 178-189.
- Plake, B. S., Loyd, B. H., & Hoover, H. D. (1981). Sex differences in mathematics components of the Iowa Test of Basic Skills. Psychology of Women Quarterly, 5, 780-784.
- Sanders, B., Soares, M. P., & D'Aquila, J. M. (1982). The sex difference on one test of spatial visualization: A nontrivial difference. Child Development, 53, 1106-1110.

- Sanderson, P. M., Flach, J. M., Buttigieg, M. A., & Casey, E. J. (1989). Object displays do not always support better integrated task performance. Human Factors, *31*, 183-198.
- Schmid, C. F. (1983). Statistical graphics: Design principles and practice. New York: Wiley.
- Seidler, K. S., & Wickens, C. D. (1992). Distance and organization in multifunction displays. Special issue: Visual displays. Human Factors, *34*, 555-569.
- Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. Journal of the American Statistical Association, *82*, 454-465.
- Sharps, M. J., Welton, A. L., & Price, J. L. (1993). Gender and task in the determination of spatial cognitive performance. Psychology of Women Quarterly, *17*(1), 71-83.
- Sharps, M. J., Price, J. L., & Williams, J. K. (1994). Spatial cognition and gender: Instructional and stimulus influences on mental image rotation performance. Psychology of Women Quarterly, *18*(3), 413-425.
- Sparrow, J. A. (1989). Graphical displays in information systems: Some data properties influencing the effectiveness of alternative forms. Behaviour and Information Technology, *8*, 43-56.
- Spence, I. (1990). Visual psychophysics of simple graphical elements. Journal of Experimental Psychology: Human Perception and Performance, *16*, 683-692.
- Spence, J., & Lewandowsky, S. (1991). Displaying proportions and percentages. Applied Cognitive Psychology, *5*, 61-77.
- Stevens, S. S. (1975). Psychophysics. New York: Wiley.
- Tapley, S. M., & Bryden, M. P. (1977). An investigation of sex differences in spatial ability: Mental rotation of three-dimensional objects. Canadian Journal of Psychology, *31* (3), 122-130.
- Teghtsoonian, M. (1965). The judgment of size. American Journal of Psychology, *78*, 392-402.
- Tufte, E. R. (1983). The visual display of quantitative information. Cheshire, CT: Graphics Press.
- Tukey, J. W. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley.
- Tukey, J. W. (1990). Data-based graphics: Visual display in the decades to come. Statistical Science, *5*, 327-339.

- Van Dusen, L., Spach, D., Brown, J., & Hansen, M. (in press). A validation study of the Trio measure of visual processing ability. Journal of Educational and Psychological Measurement.
- von Huhn, R. (1927). Further studies in the graphic use of circle and bars. Journal of the American Statistical Association, 22, 31-36.
- Washburne, J. N. (1927). An experimental study of various graphic, tabular and textual methods of presenting quantitative material. Journal of Educational Psychology, 18, 465-476.
- Wickens, C. D. (1989). Attention and skilled performance. In D. Holding (Ed.), Human skills (pp. 196-211). Chichester, UK: Wiley.
- Wickens, C. D., & Andre, A. D. (1990). Proximity compatibility and information display: Effects of color, space, and objectness on information integration. Human Factors, 32, 61-77.

APPENDICES

Appendix A. Statement of Informed Consent

I am over 18 years of age and I give my informed consent to voluntarily participate in this study of graph judgment. I consent to publication of study results so long as the information remains anonymous. I further understand that although a record will be kept of my having participated in the experiment, all experimental data will be identified by number only. I understand that test data will be kept in a locked office in the Psychology Department and such data will be kept indefinitely.

My participation in this experiment will involve making judgments about graphs that will be presented to me on pages in a test booklet. My participation in the experiment will last about 60 minutes.

I understand that if I will be receiving some form of course credit for my participation in this experiment, certification of my participation will be given to my instructor. The exact nature of this course credit is at the discretion of my instructor and not the experimenter.

I understand that the experimenter will need my grade point average at Utah State University and my General Score from the American College Test. I give my consent to the experimenter to acquire this information through the Admissions and Records Department of Utah State University.

The general purpose of this experiment is to study the efficiency with which people are able to make judgments about the quantitative properties comprising graphs. I understand that although the exact nature of the experiment will not be revealed until a later time, there are no disguised procedures.

There are no known discomforts or risks associated with my participation in this experiment. This judgment is based upon a relatively large body of research with people engaged in similar activities.

I am free to withdraw from the experiment at any time without penalty of any kind.

Following my participation, I agree not to discuss the true nature or any aspect of the experiment with anyone, until such time that all subjects have completed their participation and the study is concluded. I understand that there will be at least 120 participants in this study.

The experimenter will gladly answer any questions regarding the procedures of this study when the experimental session is completed. I further understand that if my questions are not adequately addressed by the experimenter, then the project director, Dr. Lani Van Dusen, is willing to answer any questions I may have concerning the experimental procedures. I also understand that if I am uncomfortable with any aspect of this project, I can express my concerns to Dr. Lani Van Dusen, whose phone number is (435) 797-1402.

If you understand and agree to all of the above, please sign and date this document.

(Experimenter)

(Experimental Participant)

(Principal Investigator)

(Date)

Appendix B. Experiment Script

When subjects enter the laboratory or classroom, the experimenter should greet them and invite them to take a seat. Before the experiment begins, the experimenter should allow the participants to talk among themselves, if they feel so inclined.

The experimenter will have a list of names for those individuals who have signed up to participate in the session. As the participants arrive, the experimenter should get their names to determine if everyone who has signed up for the session is actually present. If all of the participants have not arrived by the designated time, the experimenter should wait for 5 minutes before beginning the experiment. The experimenter should inform the participants that the experiment will be delayed for 5 minutes to accommodate those who may be running a little late.

During the experiment, the experimenter should continue to project items when the answers to such items are being explained. The experimenter should supplement explanations with the graph.

Usage of Certain Terms

During the administration of practice items, the experimenter must use the term "size" in addition to "area," where appropriate. The experimenter must also use the term "percentage" in addition to "proportion." The synonymous usage of these terms will clarify certain problems.

Pilot Subjects

We are developing a new test to measure certain aspects of graph reading. We need your help in improving the clarity of the instructions and tasks that comprise the test. Your job today will be to provide critical feedback about the instructions and tasks that you will be exposed to. So, we will be working together to make this a better test. Now is not the time to be shy. If you have any concerns or suggestions about anything, it's likely that you are not alone and you should feel free to speak up at any time. Any feedback that you can provide will be quite valuable. Now, if there are no questions, I will take you through the practice items and then the first six items of the test.

General Introduction

When it is time to begin, the experimenter must follow this script:

Before we begin, I want to thank you for being here. We cannot conduct research unless we have individuals, like you, who are willing to participate in the

research process. Today, we are going to be testing your graph reading abilities. In a typical graph reading situation, you are trying to assess and interpret the quantitative relationships between the different elements in the graph. In just a minute, I will project some graphs on to the projection screen. Each graph will be somewhat different, as will the type of interpretation you will be asked to make about each graph. Don't worry, we will give some examples and allow for discussion before we begin the actual testing.

Now, I will hand out the answer sheets. Please fill in your name, gender, and subject major, in the appropriate space provided. If you have not declared a major, please write "Undeclared" in this space. [The experimenter should emphasize this next point.] **Do not** fill in the spaces for "Grade Point Average" or "General ACT Score." We will collect this information from the Admissions and Records Department here at the university. If you do want us to collect this information from the Admissions and Records Department, please make a note indicating this at the top of your answer sheet. [Wait for everyone to finish this before continuing. The experimenter should hand out the test booklets now.]

Now, let's read through the instructions together. [The experimenter should pick up an answer sheet and slowly and clearly read the first set of instructions. The experimenter should be looking for facial expressions or other body language that would indicate confusion on the part of participants. After reading the instructions, the experimenter should continue.] *Are there any questions?* [At this point, any questions that do not require the exposition of the experimental hypothesis should be addressed. The experimenter should then continue.]

Item A Pre

Please look up at the projection screen and let's attack this item together. This is practice item A. Notice the question below the graph. This is one type of question that you will be asked. Go ahead. See if you can answer the question. When you think you have it, write your answer in the space provided on your answer sheet. [The experimenter should allow the subjects some time to do this. When it looks as though most of the participants have done this, the experimenter should continue.]

Item A Post

The answer to the question is 33. Does everyone understand how I got that answer? [The experimenter should pause to wait for a response.] Notice the scale on the vertical axis of the graph. It ranges from 0 to 100. For this first example, Bar A has a length of 33 on this scale. Do not worry if you did not get this answer exactly correct. The important thing is that you understand why the correct answer is 33. Does everyone understand? Please do not be bashful about speaking up. [The experimenter should pause to wait for a response. If there is confusion, the experimenter should restate the above explanation.]

Item B Pre

Good. I think we have it. Let's move on to practice item B. Practice item B has the same type of question. However in this case, you will be estimating angles. See if you can answer the question for practice item B. [The experimenter should allow the subjects some time to do this. When it looks as though most of the participants have done this, the experimenter should continue.]

Item B Post

The answer to the question is 52 degrees. For this example, Angle B is open to a 52 degree angle. Does everyone understand how I got that answer? [The experimenter should pause to wait for a response. If there is confusion, the experimenter should briefly review how values are assigned to angles and the differences between 45, 90, and 180 degree angles.] Do not worry if you did not get this answer exactly correct. The important thing is that you understand why the correct answer is 52.

Item C Pre

Let's move on to another type of question which involves slopes. Here is practice item C. For this item, you will be comparing slopes. The value of a slope represents the magnitude and direction that a line deviates from a horizontal plane. In this experiment, if a line is exactly horizontal, it has a slope of 0 degrees. If a line is exactly vertical, then it has a slope of 90 degrees. So, as a line moves from a horizontal position to a vertical position, the slope increases. Now, try to answer the question. When you think you have it, write your answer in the space provided on your answer sheet. [The experimenter should allow the subjects some time to do this. When it looks as though most of the participants have done this, the experimenter should continue.]

Item C Post

The answer to the question is 62.7%. For these types of questions, your response must be in terms of percentages. Let me repeat that. When the test asks for a proportion, your answer must be in terms of a percent. Does everyone understand how I got that answer? [The experimenter should pause to wait for a response.] To answer this question, you just mentally estimate the individual slopes and then mentally estimate the proportion (or percentage) that slope A is of slope C. Are there questions about this? [The experimenter should pause to wait for a response. If there is confusion, the experimenter should restate the above slope explanation.]

Item D Pre

O.K. Let's continue. Here is practice item D. Practice item D has the same type of question. However in this case, you will be judging distances between the horizontal axis and the points above it. Remember to use the scale on the vertical axis to help you with your estimation. Now, see if you can answer the question for practice item D. [The experimenter should allow the subjects some time to do this. When it looks as though most of the participants have done this, the experimenter should continue.]

Item D Post

*The answer to the question is 159.5%. Does everyone understand how I got that answer? [Experimenter should pause to wait for response.] Again, to answer this type of question, you just mentally estimate the distances between the horizontal axis and each point. Then you mentally estimate how much larger point C is than point A. In this example, point C is 159.5% of point A. **Notice that the correct answer is 159.5% and not 59.5%. On this test, if you are asked to estimate the proportion that a larger value is of a smaller value, then your answer must always be greater than 100%. If point C really was 59.5% of point A, then point C would be smaller than point A. And you can see that point C is definitely larger than point A. Are there questions about this?** [The experimenter should pause to wait for a response. If there is confusion, the experimenter should briefly restate or rephrase the above explanations.]*

Item E Pre

If we are all ready to move on, let's continue to another type of question. Here is practice item E. Practice item E involves graphs that are similar to the one presented in practice item D. The difference here is that you will mentally add the values of points B and C. Then, you will compare point A to the sum of points and B and C. Go ahead and answer the question for practice item E.

Item E Post

The precise answer to the question is 37.1%. Does everyone understand how I got that answer? [The experimenter should pause to wait for a response.] To answer this question, you might begin by mentally adding the distances between point B and the horizontal axis and point C and the horizontal axis. Once you have mentally summed these values, point A is proportionally compared to that sum, and point A is 37.1% of point B plus point C. Are there questions about this. [The experimenter should pause to wait for a response. If there is confusion, the experimenter should go through the above explanation again.]

Item F Pre

*Why don't we practice this type of question again. Here is practice item F. In this example, you will be estimating the sizes of different squares. Even though you are comparing the sizes of different squares, **you shouldn't be trying to do the math for area**. That would be overwhelming without a calculator. Just mentally combine the sizes of the squares when you need to. Now, answer the question. [The experimenter should allow the subjects some time to do this. When it looks as though most of the participants have done this, the experimenter should continue.]*

Item F Post

The precise answer to the question is 76.2%. Can you see how I arrived at that answer? [The experimenter should pause to wait for a response.] If you mentally add squares A and B together, you will find that square C is 76.2% of the A-B square that you are mentally imagining. Are there questions about this? [Experimenter should pause to wait for response. If there is confusion, the experimenter should briefly discuss how the area of a parallelogram is derived from length and width.]

Item G Pre

*There is one more type of question that we need to cover. Here is practice item G. This type of question requires you to estimate a trend and then make some decisions about a graphical element that is hypothetical. In this example you will see squares A, B and C. The size of these squares increases at a uniform, proportional rate. Your task in this type of question involves three parts. **First**, You will need to mentally estimate the uniform rate that the squares increase by. **Second**, based on this estimate, you will need to imagine what the size of the next square in the series should be. **Third**, you would then compare square B to the hypothetical square which we will call square D. Go ahead and answer the question. [The experimenter should allow the subjects some time to do this. When it looks as though most of the participants have done this, the experimenter should continue.]*

Item G Post

*The precise answer to the question is 56.4%. Let's go through it step by step. **First**, the size of these squares increases at a rate of 33%. That is, square B is 33% larger than square A, and square C is 33% larger than square B. **Second**, if you take that 33% and apply it to square C, then you will be able to imagine the size of hypothetical square D. **Third**, once you have done this, you can mentally compare square B with square D. You will find that square B is 56.4% of the mentally imaged square D. Are there questions about this? [The experimenter should pause to wait for a response. If there is confusion about the trend or the extrapolated square D, the experimenter should restate relevant instructions.]*

Item H Pre

Let's practice this type of problem again. Only, this time let's use angles. Approach the problem in the same way. Go through the same three steps to answer the question.

Item H Post

*The precise answer to the question is 177%. Let's go through it step by step. **First**, the angles increase at a rate of 31%. **Second**, if you take that 31% and apply it to angle C, then you will be able to imagine the size of the hypothetical angle. **Third**, once you have done this, you can mentally compare angle D with angle B. If you measured precisely, you would find that square D is 177% of square B. Are there questions about this? [The experimenter should pause to wait for a response. If there is confusion, the experimenter should answer subjects' questions.]*

Final Instructions

We are ready to move on. Look back at your answer sheet. Let's read the next set of instructions. [The experimenter should read through these instructions.] Does everyone understand? Good. Let's begin. When we are through with the test, I will collect your answer sheets and then you will be free to leave. If you would like to know more about the theories being tested in this experiment, just ask when you turn in your test. I would be happy to provide you with more information.

Post Test

At the end of the experiment, the experimenter should sign any certificates which will be used in exchange for course credit. Subjects should be told that they can find out about their performance after the data have been analyzed. Examinees should be given Derek Borman's office phone number, which is 797-3817.

Appendix C. Answer Sheet and Pilot Checklists

Name _____ Sex _____ Subject Major _____

Overall Grade Point Average _____ General ACT Score _____

Immediately below, are spaces in which you will write your responses to the practice items. We will do the practice items together. The spaces below are labeled to correspond with the labels for the graphs presented on the projection screen. Notice that you must proceed from the left side of the page to the right side of the page, to answer the items in order. Below each projected graph there will be a question for which you should give your best response, even if that means giving a rough estimate or a guess.

Practice Items

A _____ B _____ C _____ D _____
 E _____ F _____ G _____ H _____

Below, are spaces in which you will write your responses to test items. This is the beginning of actual testing. During the test, use the projector to show each item on the screen in front of you. We will move through the test as a group. That is, I will not move on to a new test item until everyone has had enough time to answer the previous one. If I move too quickly from one test item to the next, simply raise your hand and let me know that you need more time. If you struggle with an item, you should simply consider it to the best of your ability and then make your best guess. **Remember**, when writing your answers on this sheet, you must move from the left to the right side of the answer sheet. If you have any questions, you should ask them now. However, some help can be given once the test has started. In addition, there is one item (item #5) for which you will be identifying the area of a square. For this item, you will probably need to calculate the answer with a pencil and paper or with a calculator. Feel free to do so.

Test Items

1 _____ 2 _____ 3 _____ 4 _____
 5 _____ 6 _____ 7 _____ 8 _____
 9 _____ 10 _____ 11 _____ 12 _____
 13 _____ 14 _____ 15 _____ 16 _____
 17 _____ 18 _____ 19 _____ 20 _____
 21 _____ 22 _____ 23 _____ 24 _____

Problems

	Unable	Lag Time	Vocab	Other
Introduction				
Answer Sheet				
Item A (Pre)				
Item A (Post)				
Item B (Pre)				
Item B (Post)				
Item C (Pre)				
Item C (Post)				
Item D (Pre)				
Item D (Post)				
Item E (Pre)				
Item E (Post)				
Item F (Pre)				
Item F (Post)				
Item G (Pre)				
Item G (Post)				
Item H (Pre)				
Item H (Post)				

Resolution of Problems

Introduction	
Answer Sheet	
Item A (Pre)	
Item A (Post)	
Item B (Pre)	
Item B (Post)	
Item C (Pre)	
Item C (Post)	
Item D (Pre)	
Item D (Post)	
Item E (Pre)	
Item E (Post)	
Item F (Pre)	
Item F (Post)	
Item G (Pre)	
Item G (Post)	
Item H (Pre)	
Item H (Post)	

Appendix D. Answers to Test Items and Test Items

It is important to note that for graphs comprising distance, length and/or area information, the values below reflect units on a graph with a 2-inch ordinate that has a scale of 100.

Practice Items

- A. Point reading: Length
Graph information for unit length: A (33), B (48), C (52)
Answer: The value of bar A is 33.
- B. Point reading: Angle
Graph information in degrees: A (33), B (48), C (52)
Answer: The value of angle C is 52.
- C. Local comparison: Slope
Graph information in degrees: A (37), B (42), C (59)
Answer: Slope A is 62.7% of slope C.
- D. Local comparison: Position on common aligned scale
Graph information for unit distance: A (37), B (42), C (59)
Answer: Point C is 159.5% of point A.
- E. Global comparison: Positions on non-aligned identical scales
Graph information for unit distance: A (36), B (46), C (51)
Answer: Point A is 37.1% of point B + point C.
- F. Global comparison: Area
Graph information for unit distance: A (36; 1296), B (46; 2116), C (51; 2601)
Answer: Square C is 76.2% of square A + square B.
- G. Synthesis: Area
Trend: 33% increases
Graph information for unit length/width and area: A (32; 1024), B (36.9; 1361.61)
C (42.6; 1814.76), D (49.13; 2413.76)
Answer: Square B is 56.4% of square D.
- H. Synthesis: Angle
Trend: 31%
Graph information in degrees: A (32), B (42.6), C (56.7), D (75.4)
Answer: Angle D is 177% of angle B.

Test Items

1. Local comparison: Slope
Graph information in degrees: A (37), B (44), C (58)
Answer: Slope A is 63.8% of slope C.
2. Synthesis: Position on common aligned scale
Trend: 27% increases
Graph information: A (34), B (43.18), C (54.84), D (69.65)
Answer: Point B is 62% of point D.

3. Point reading: Position on common aligned scale
Graph information for unit distance: A (38), B (49), C (59)
Answer: The value of point C is 59.
4. Local comparison: Positions on non-aligned identical scales
Graph information for distance: A (31), B (46), C (56)
Answer: Point C is 180.6% of point A.
5. Point reading: Area
Graph information for unit length/width and area: A (37; 1369), B (45; 2025), C (53; 2809)
Answer: The area of square A is 1369 square units.
6. Global comparison: Angle
Graph information in degrees: A (35), B (42), C (54)
Answer: Angle A is 36.5% of angle B + angle C.
7. Synthesis: Slope
Trend: 32% increases
Graph information: A (37), B (48.8), C (64.4), D (85)
Answer: Slope D is 174.2% of slope B.
8. Global comparison: Length
Graph information for unit length: A (39), B (44), C (51)
Answer: Bar C is 61.4% of bar A + bar B.
9. Point reading: Slope
Graph information in degrees: A (33), B (42), C (57)
Answer: Slope C is 57 degrees.
10. Local comparison: area
Graph information for unit length/width and area: A (38; 1444), B (49; 2401), C (59; 3481)
Answer: Square A is 41.5% of square C.
11. Global comparison: Positions on non-aligned, identical scales.
Graph information for unit distance: A (32), B (40), C (54)
Answer: Point A is 34% of point B + point C.
12. Local comparison: Length
Graph information for unit length: A (37), B (48), C (55)
Answer: Bar C is 148.6% of bar A.
13. Synthesis: Angle
Trend: 29% increases
Graph information in degrees: A (33), B (42.6), C (55), D (71)
Answer: Angle D is 167.7% of angle B.
14. Local comparison: Position on common aligned scale
Graph information for unit distance: A (34), B (41), C (55)
Answer: Point C is 161.8% of point A.

15. Point reading: length
Graph information for unit length: A (30), B (49), C (57)
Answer: The value of bar A is 30.
16. Synthesis: Area
Trend: 28% increases
Graph information for unit length/width and area: A (35; 1225), B (39.6; 1568.16),
C (44.8; 2007.04); D (50.7; 2570.49)
Answer: Square B is 61% of square D.
17. Global comparison: Area
Graph information for unit length/width and area: A (33; 1089), B (48; 2304), C (51; 2601)
Answer: Square A is 22.2% of square B + square C.
18. Point reading: Positions on non-aligned identical scales
Graph information for unit distance: A (36), B (43), C (54)
Answer: The value of point C is 54.
19. Global comparison: Slope
Graph information in degrees: A (39), B (42), C (50)
Answer: Slope C is 61.7% of slope A + slope B.
20. Synthesis: Length
Trend: 30% increases
Graph information for unit length: A (34), B (44.2), C (57.46), D (74.7)
Answer: Bar D is 169% of bar B.
21. Synthesis: Positions on non-aligned identical scales
Trend: 34% increases
Graph information for unit distance: A (35), B (46.9), C (62.85), D (84.22)
Answer: Point B is 55.7% of point D.
22. Point reading: Angle
Graph information in degrees: A (37), B (45), C (58)
Answer: Angle A is 37 degrees.
23. Global comparison: Position on common aligned scale
Graph information for unit distance: A (33), B (41), C (56)
Answer: Point A is 34% of point B + point C.
24. Local comparison: Angle
Graph information in degrees: A (30), B (44), C (58)
Answer: Angle C is 193.3% of angle A.

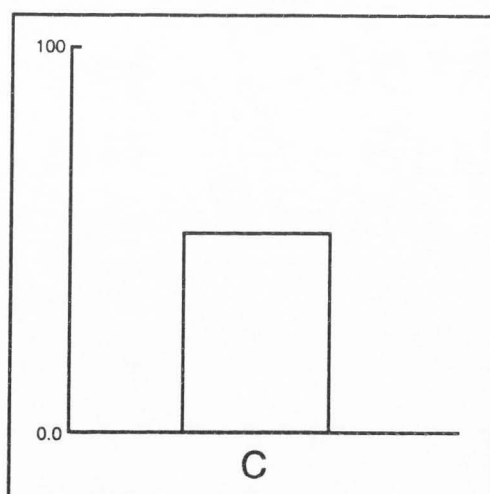
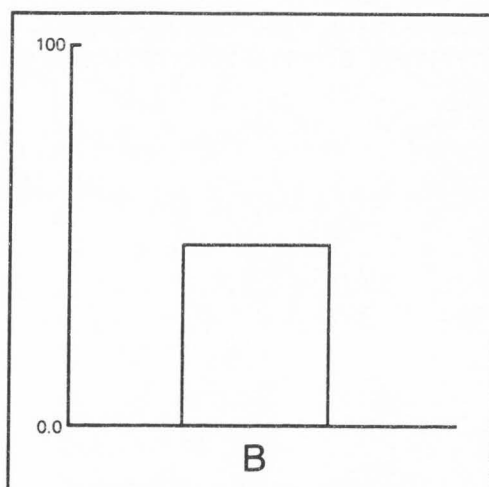
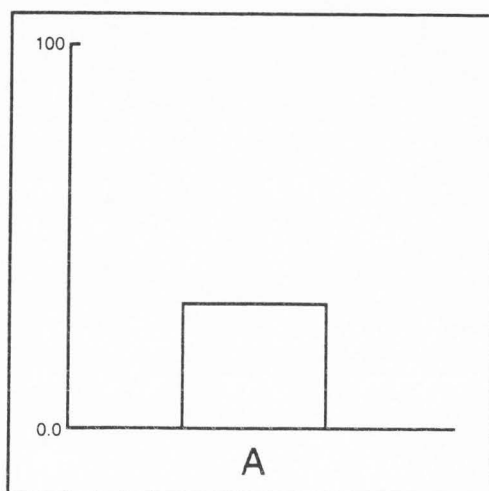
Additional Item

The following item was added to the test in the middle of data collection. It was noted by some subjects that for assessing synthesis tasks in this study, it wasn't necessary to mentally image specifier/graph D. Because the increase in the size of the specifiers was uniform, some subjects simply chose to compare specifiers/graphs A and C instead of B and D. Although it is true that the proportional relationship between A and C was the same as that between B and D, any answers based on comparisons of A and C could not be considered as valid responses to synthesis tasks. By definition, a comparison

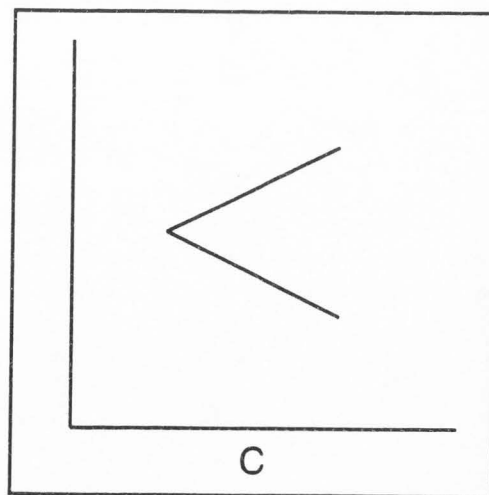
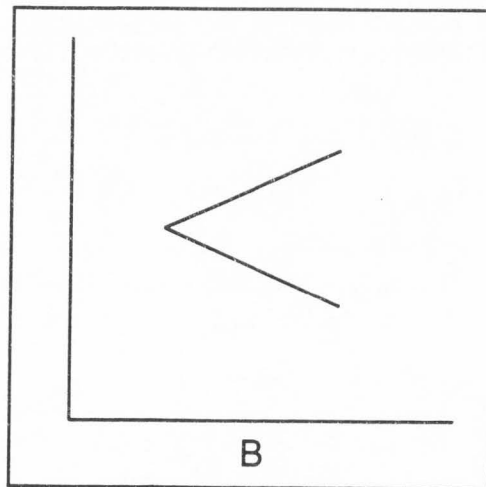
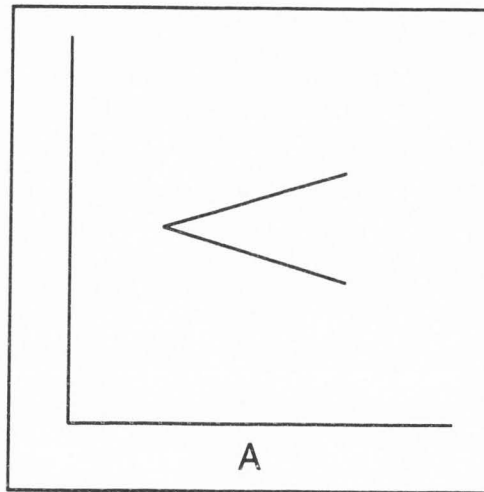
between two points that are physically present is simply a point comparison task. The fundamental component of a synthesis task involves: 1) the extrapolation of a trend relationship across specifiers and 2) the subsequent imaging of the next specifier in the series, based on the estimated trend. Subjects who compared A and C were most likely performing neither of these tasks.

In light of the above discovery, an additional item was constructed. This question was asked so that one more data point could be used to assess the possible effects of some subjects' failure to approach synthesis tasks in the intended manner. The question was not depicted on an overhead transparency, but was instead administered orally by the experimenter. This last question was rephrased and reiterated to the extent that it was clearly understood by all participants. The additional test item is listed below:

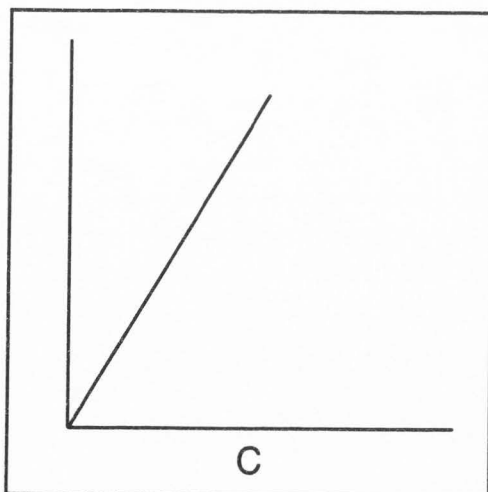
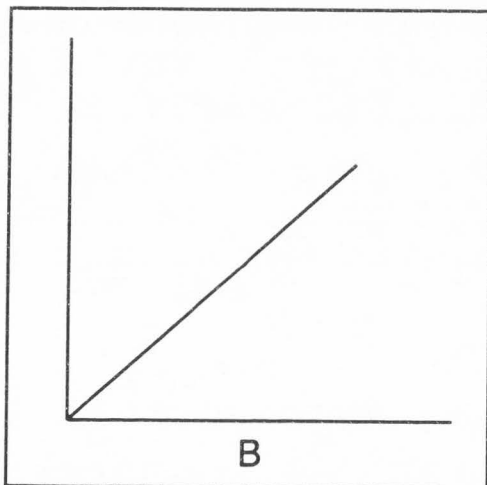
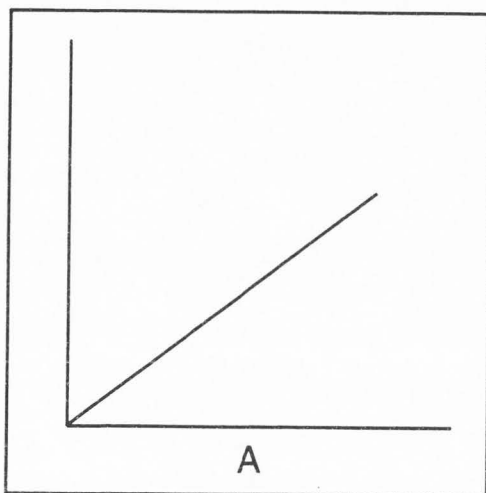
25. "For questions in which you were asked to compare graph B with graph D, did you actually do this?"



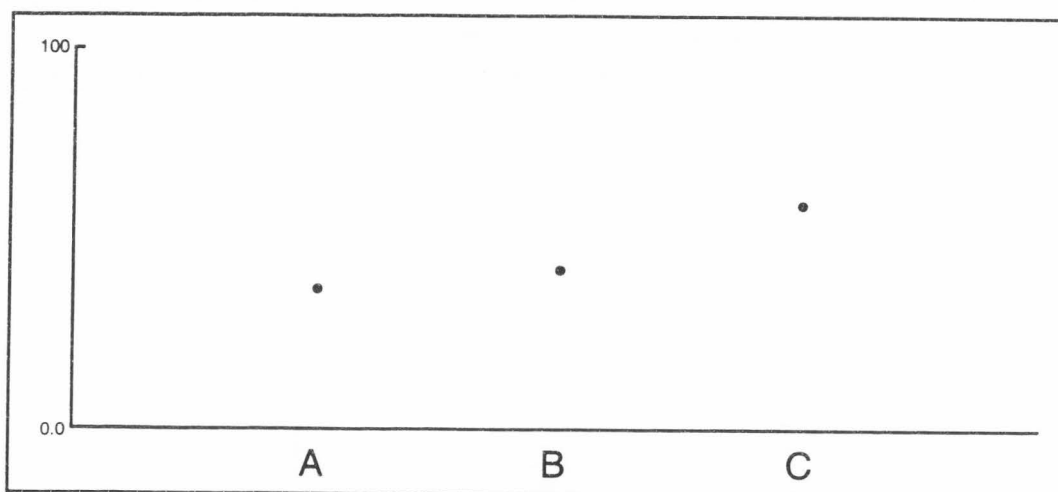
Practice Item A. What is the value of bar A?



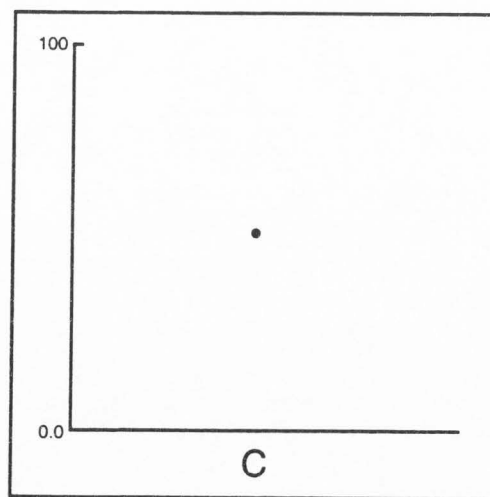
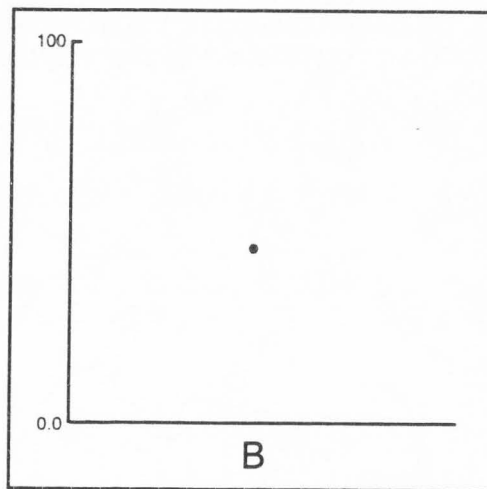
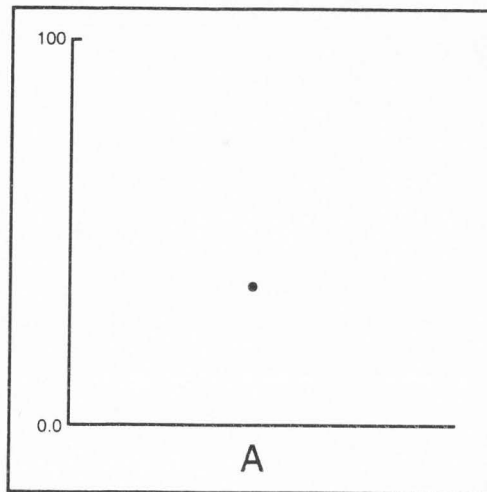
Practice Item B. What is the value of angle C?



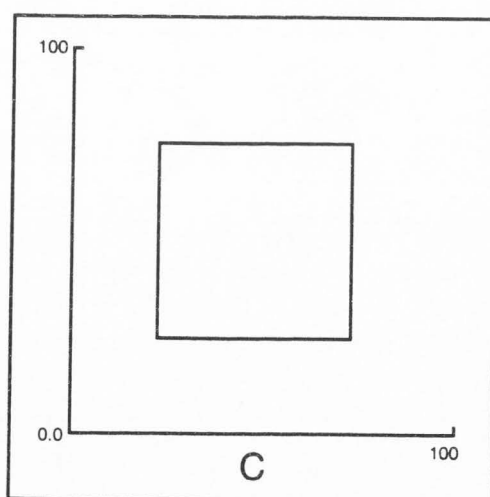
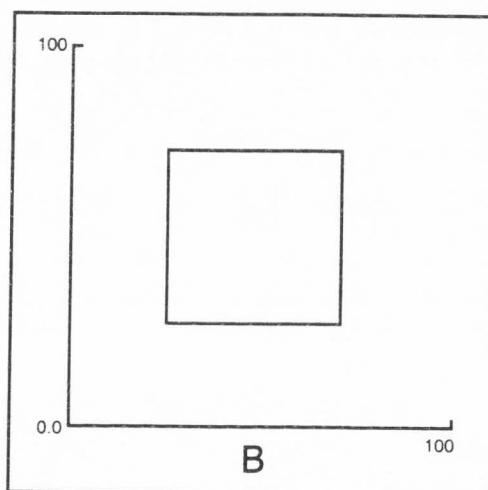
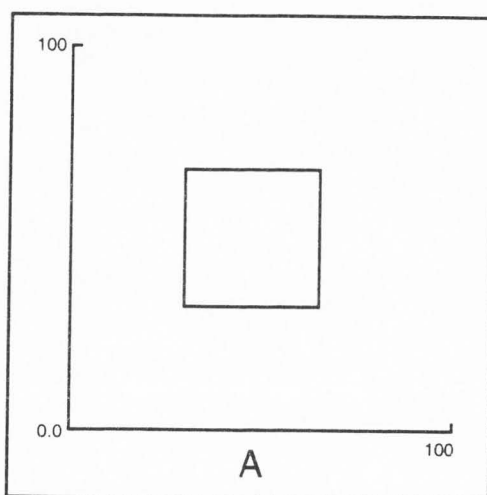
Practice Item C. Slope A is what proportion of slope C?



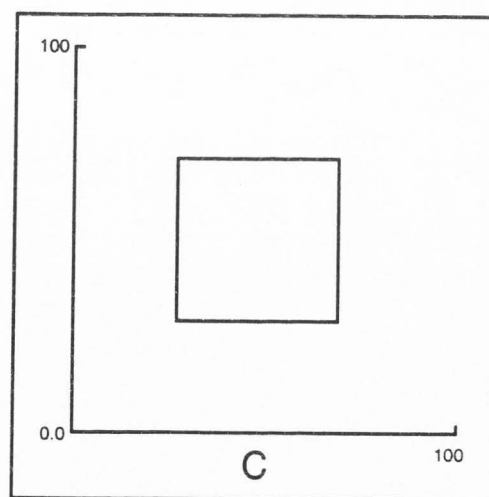
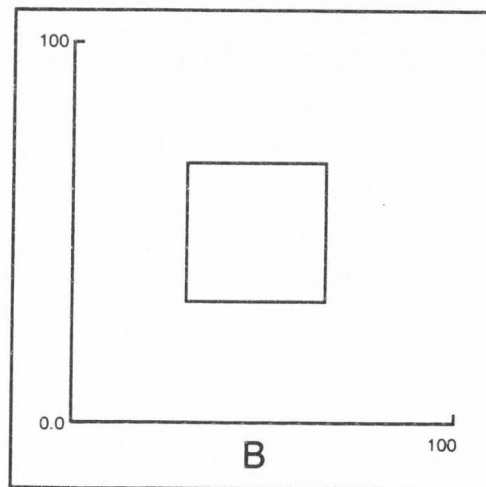
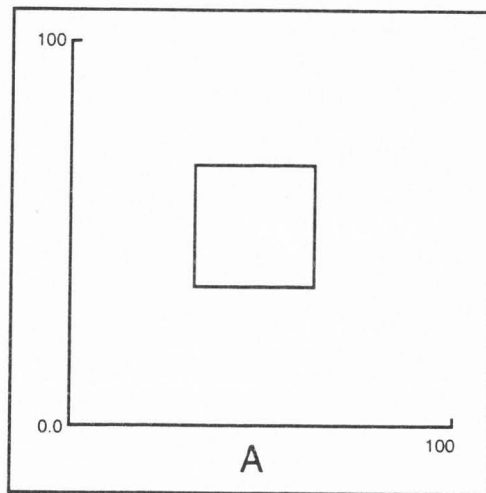
Practice Item D. Point C is what proportion of point A?



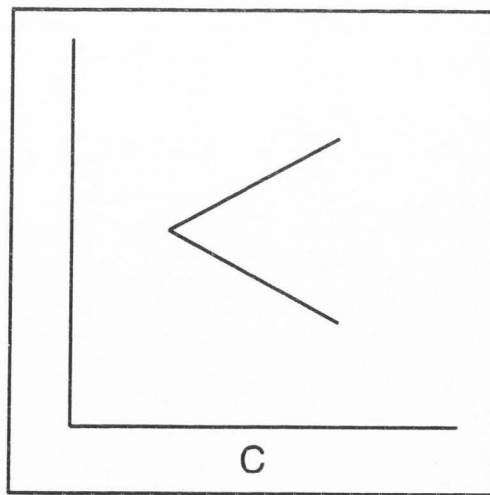
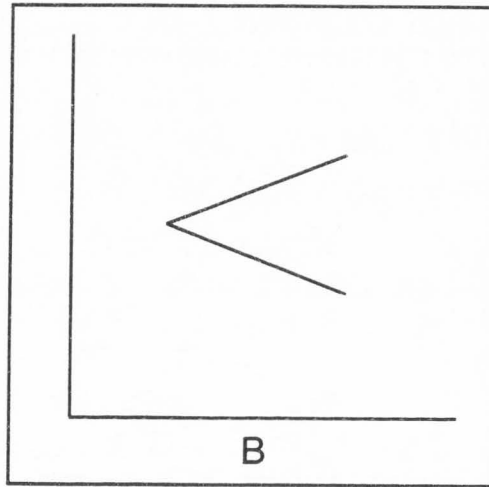
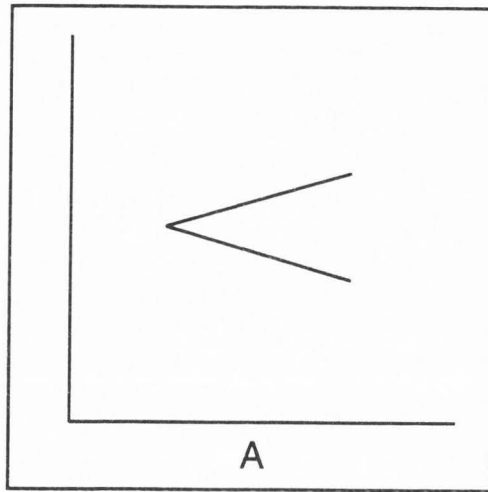
Practice Item E. Point A is what proportion of point B + point C?



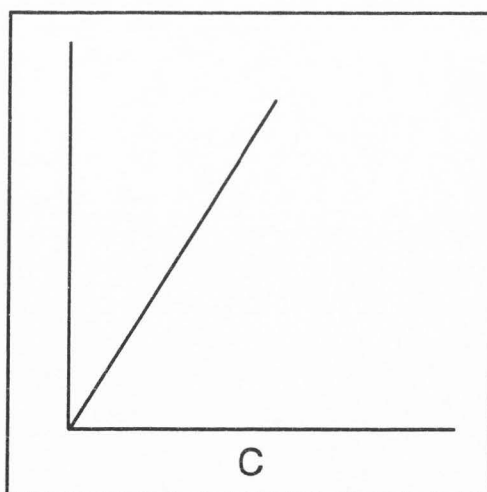
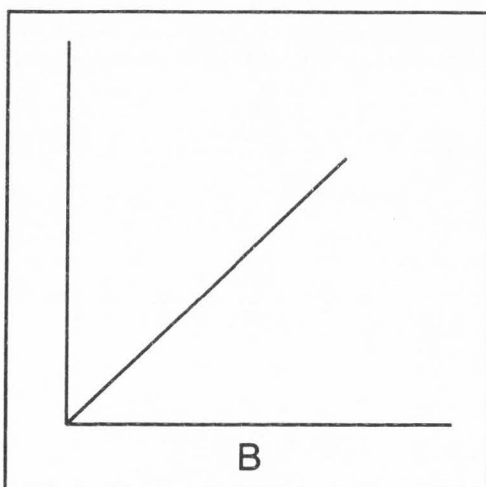
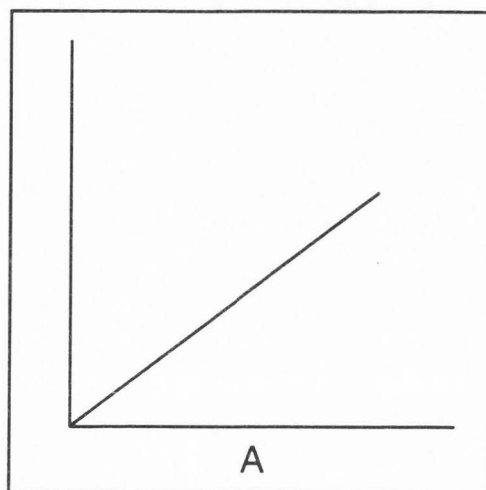
Practice Item F. Square C is what proportion of square A + square B?



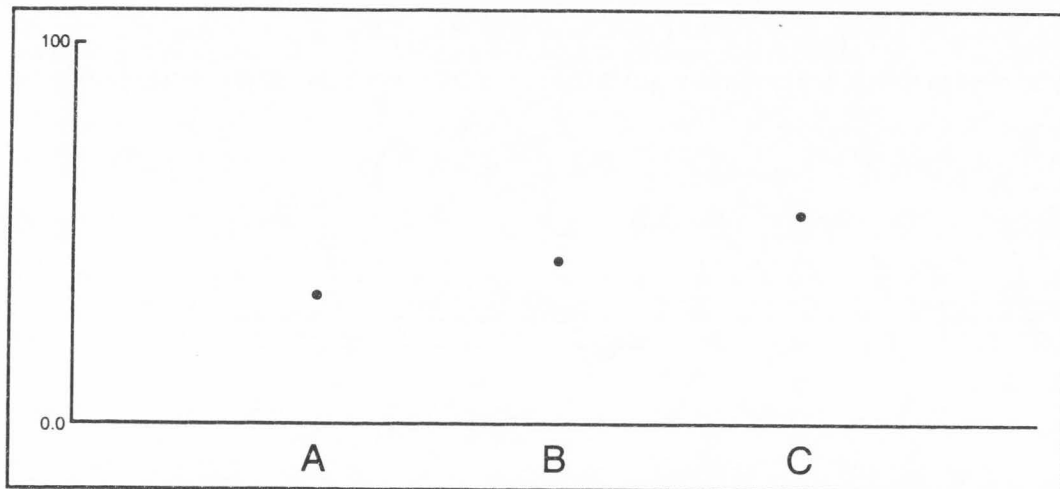
Practice Item G. Square B is what proportion of square D?



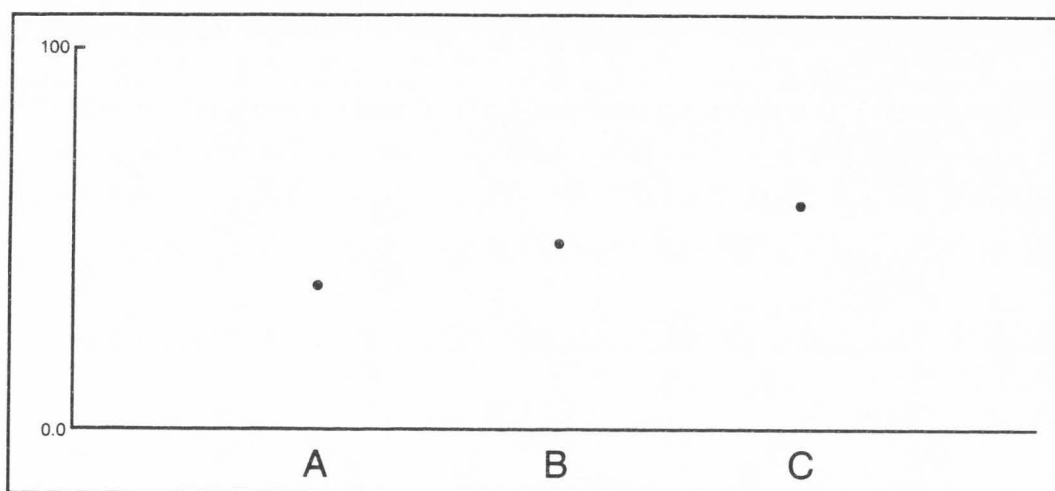
Practice Item H. Angle D is what proportion of angle B?



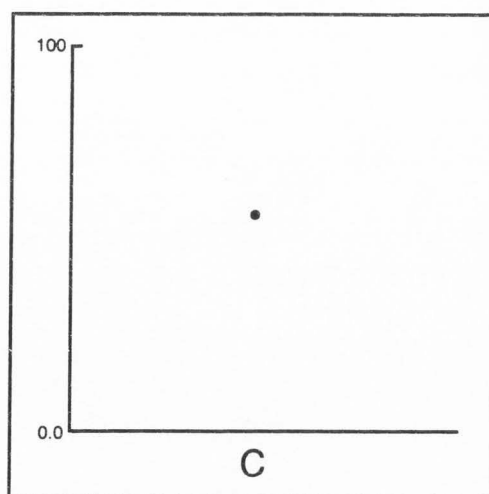
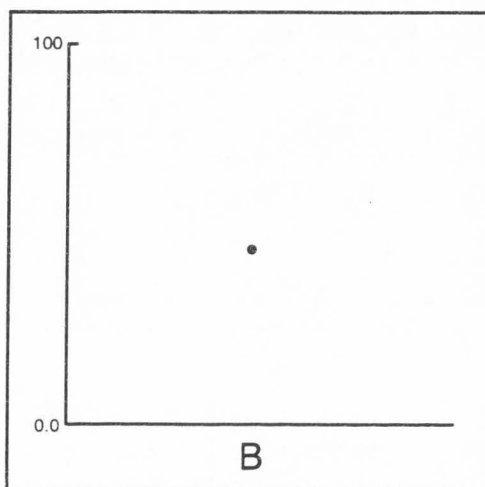
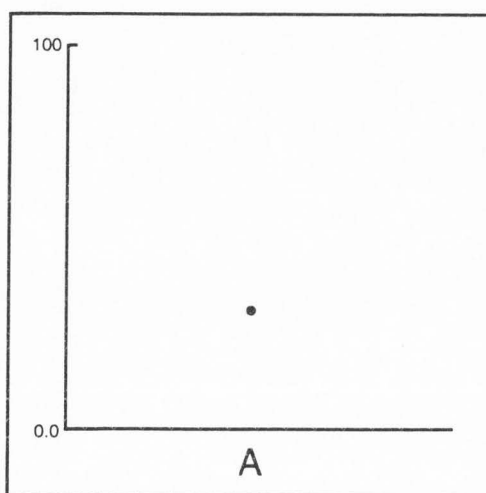
1. Slope A is what proportion of slope C?



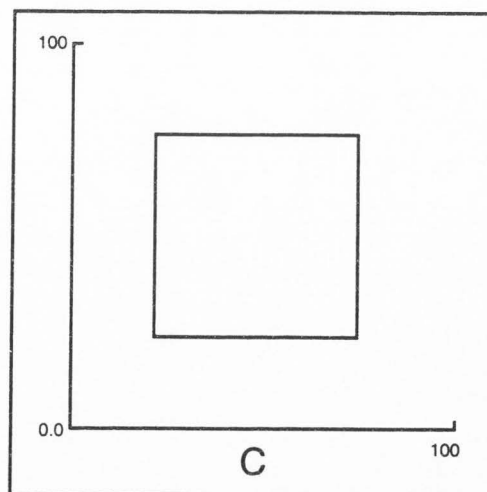
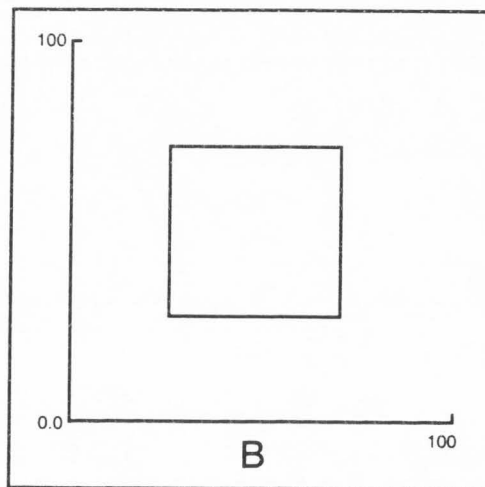
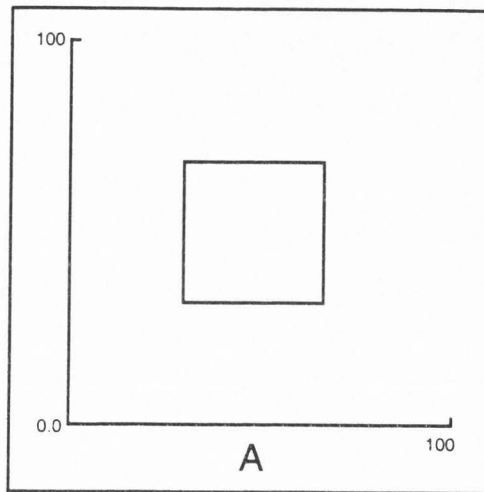
2. Point B is what proportion of point D?



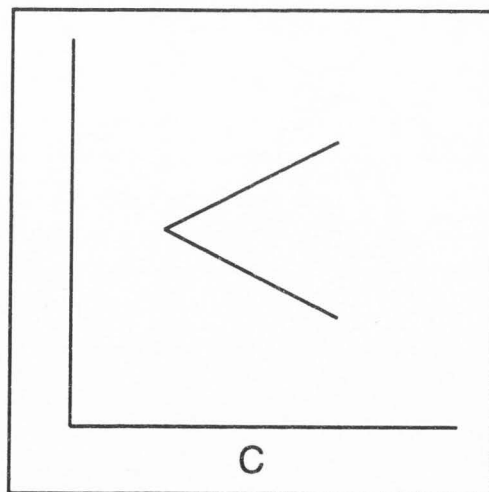
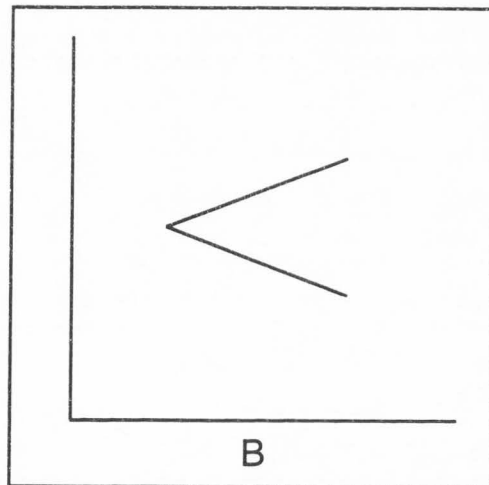
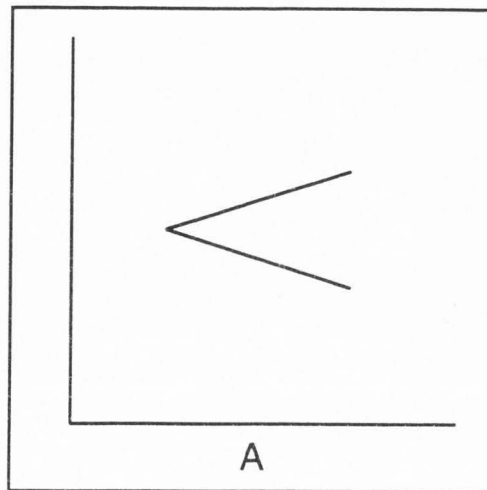
3. What is the value of point C?



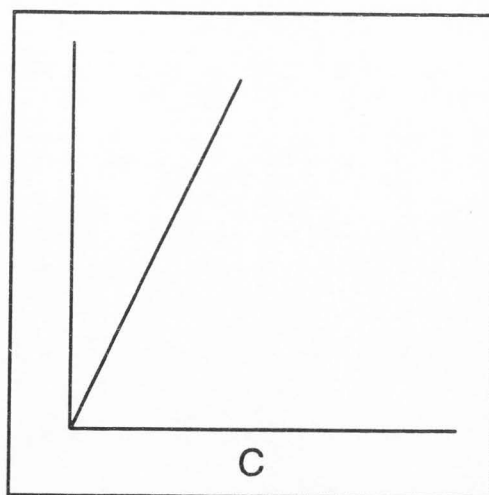
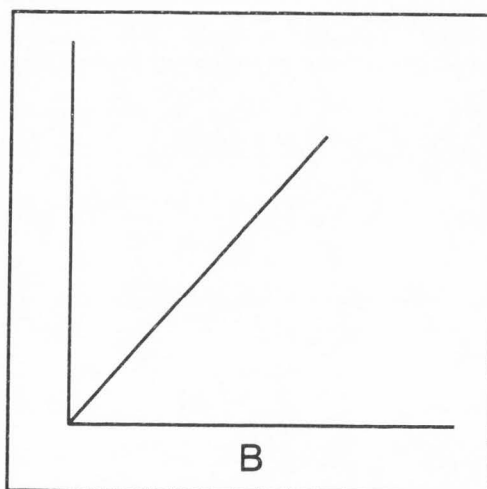
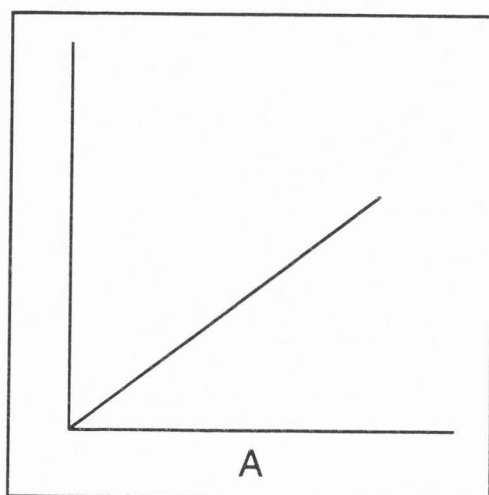
4. Point C is what proportion of point A?



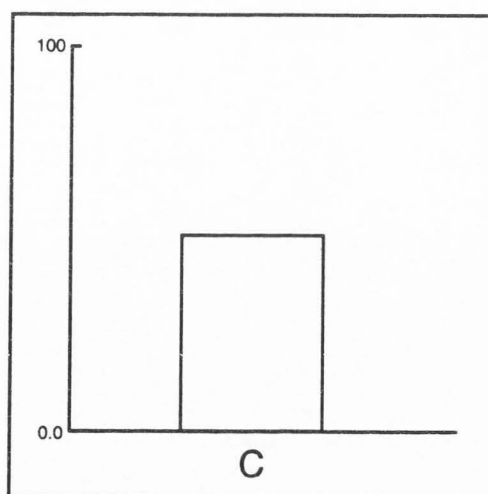
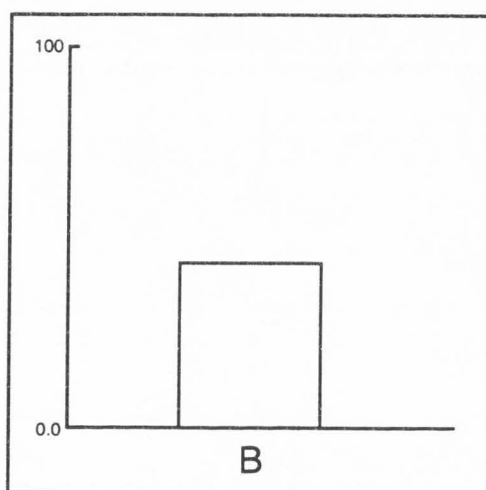
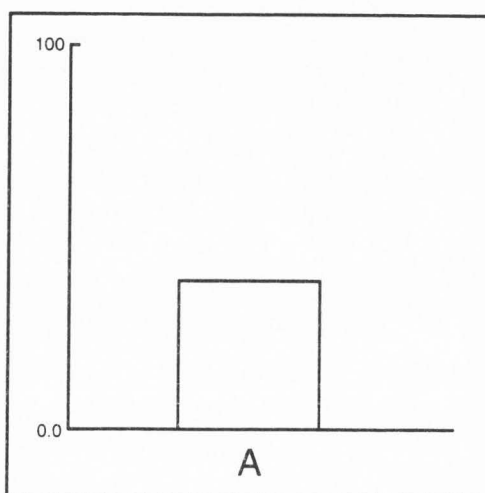
5. What is the area of square A?



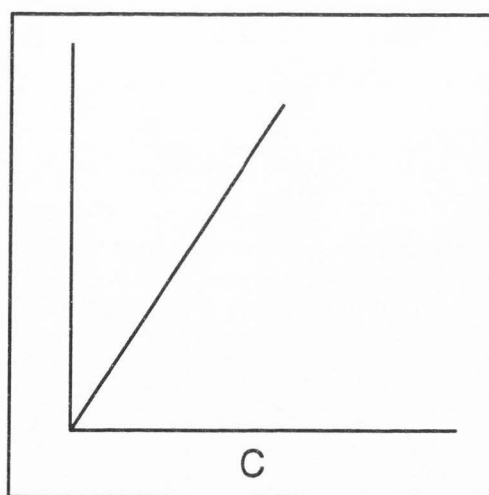
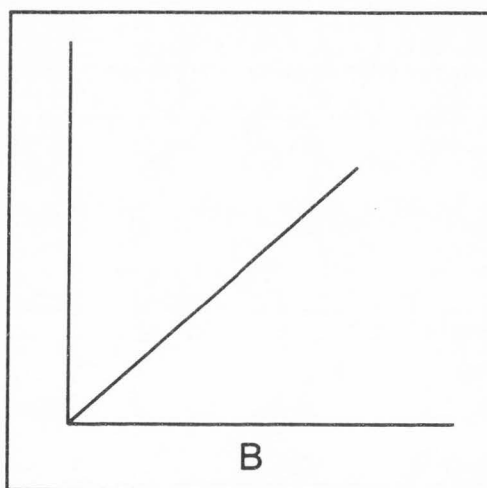
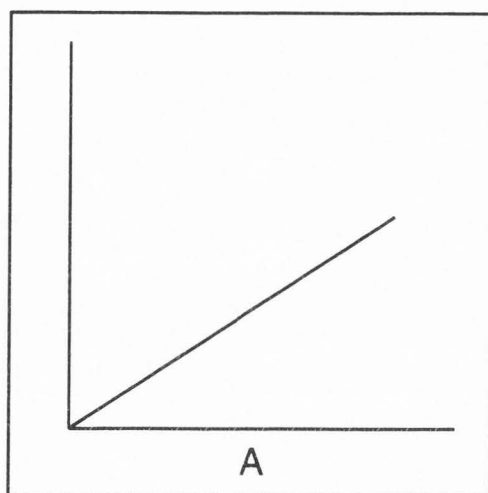
6. Angle A is what proportion of angle B + angle C?



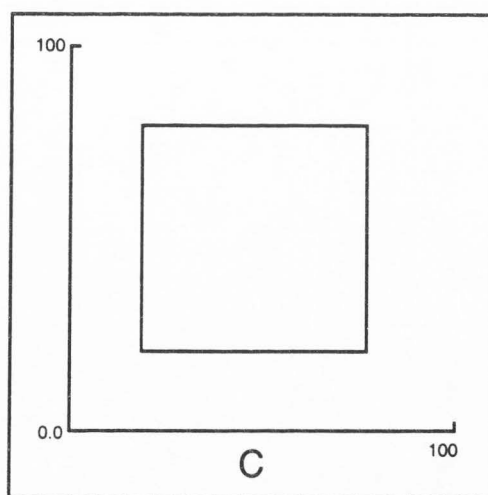
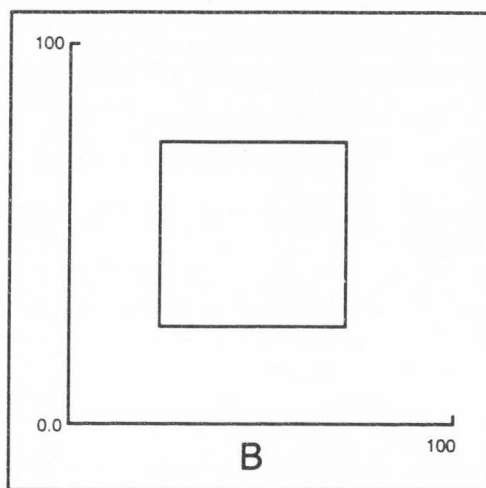
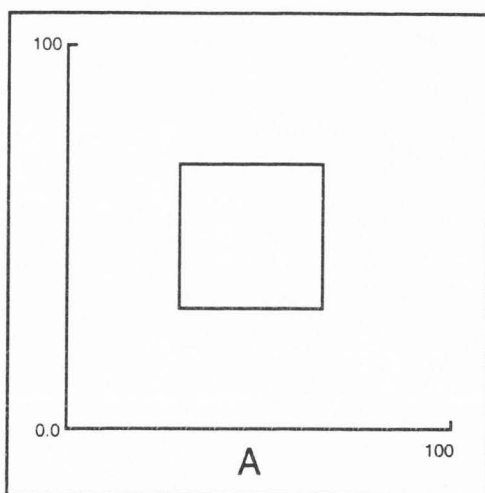
7. Slope D is what proportion of slope B?



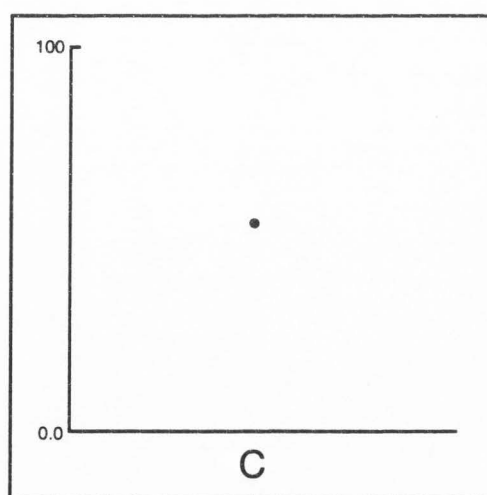
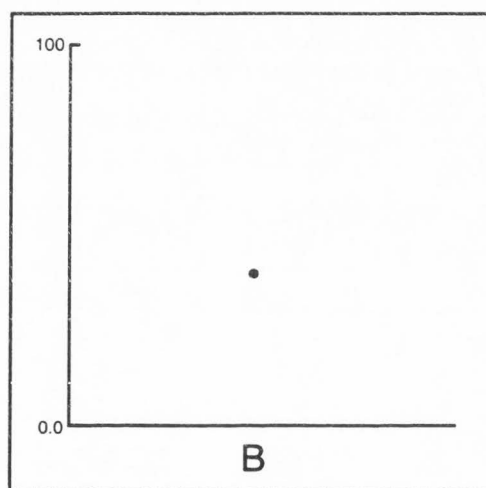
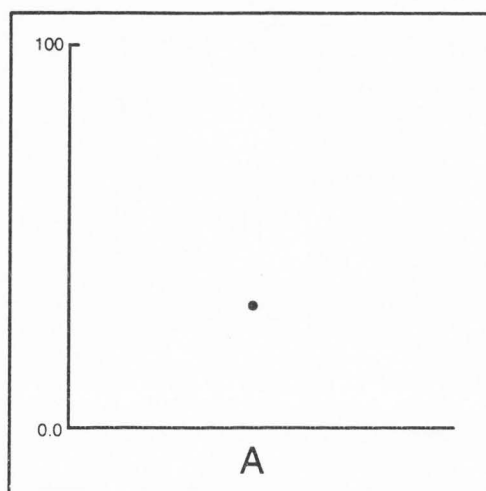
8. Bar C is what proportion of bar A + bar B?



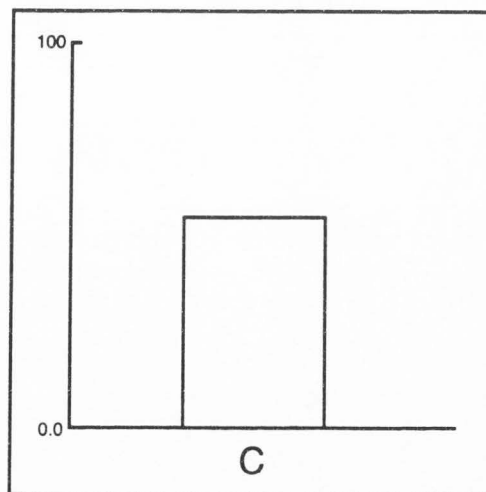
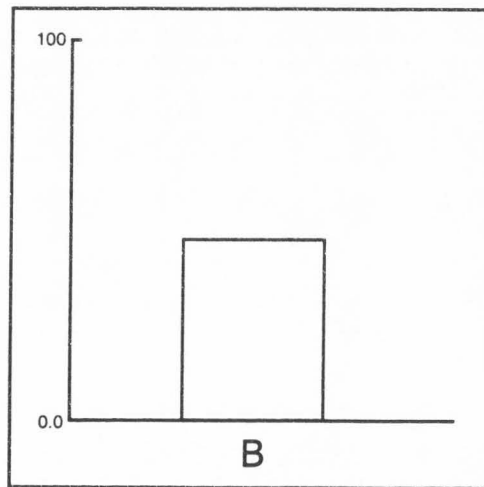
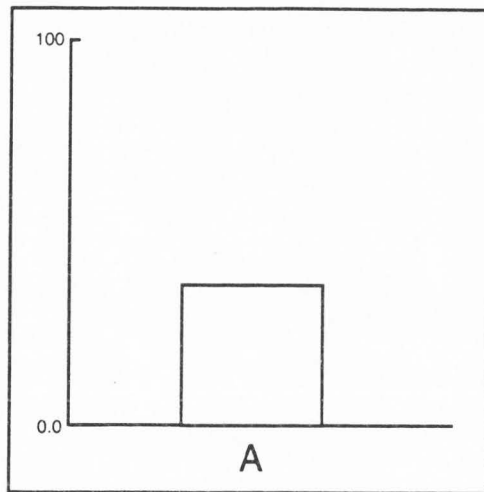
9. What is the value of slope C?



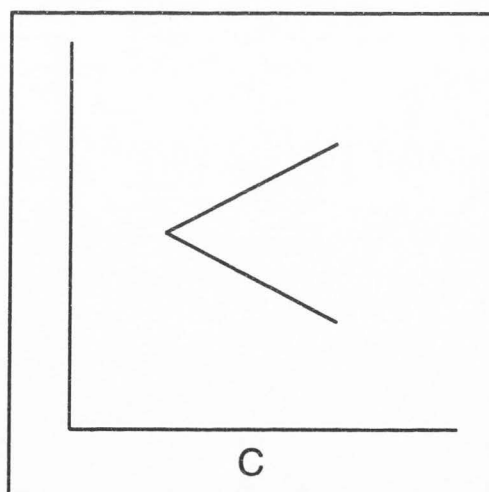
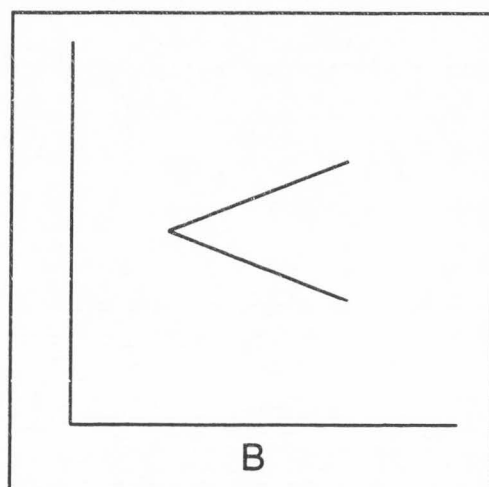
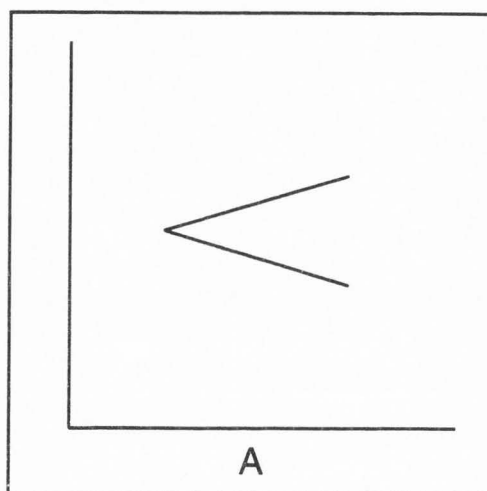
10. Square A is what proportion of square C?



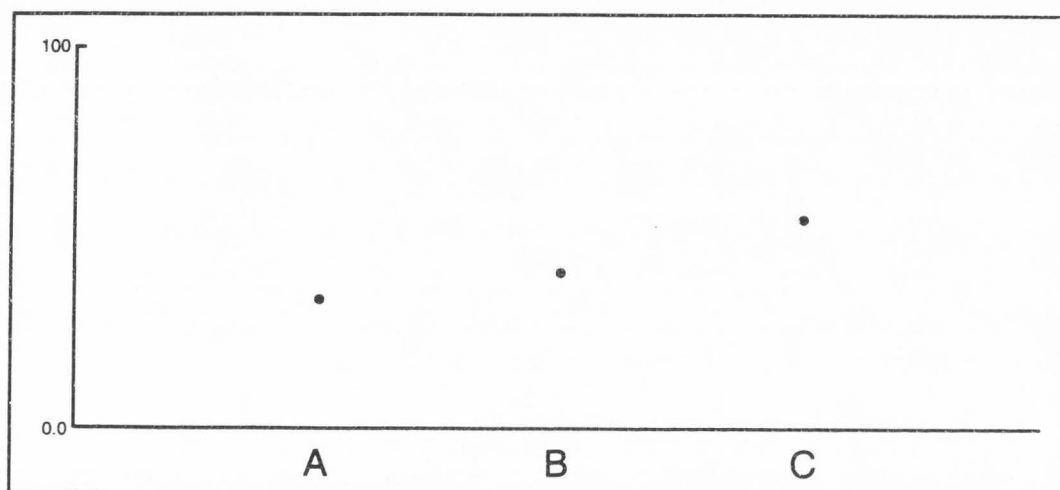
11. Point A is what proportion of point B + point C?



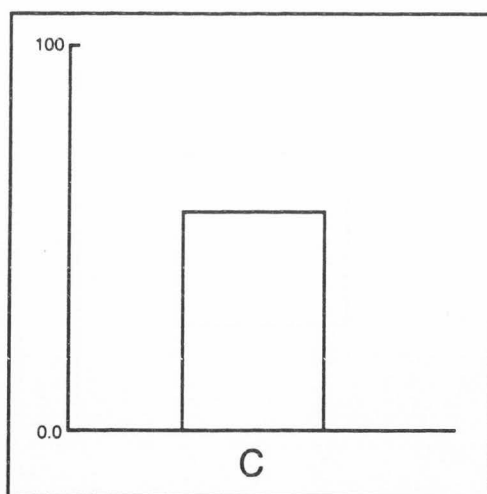
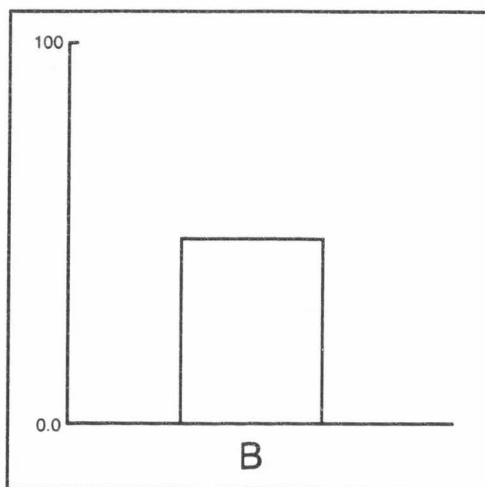
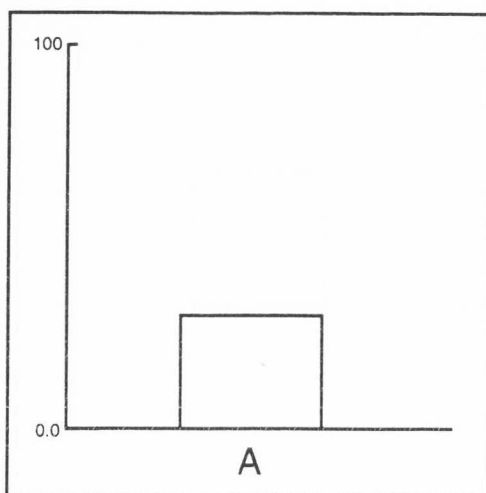
12. Bar C is what proportion of bar A?



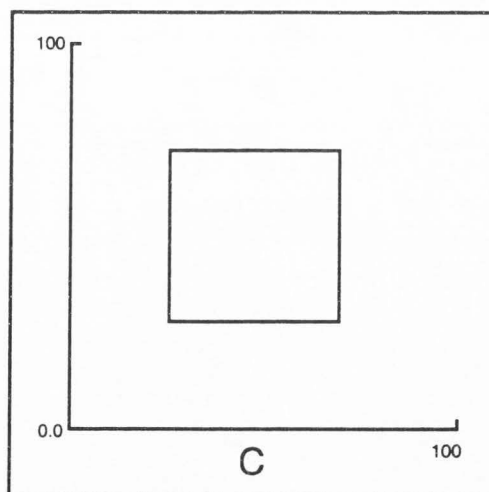
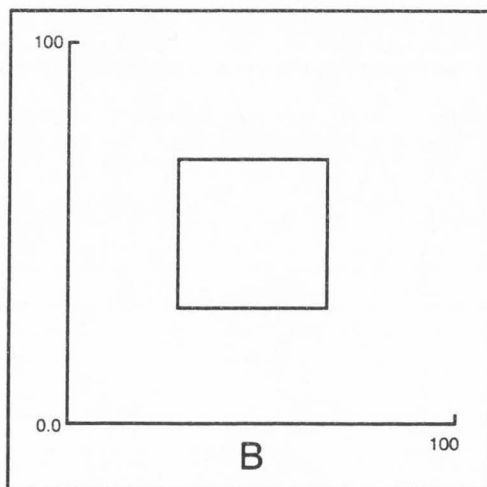
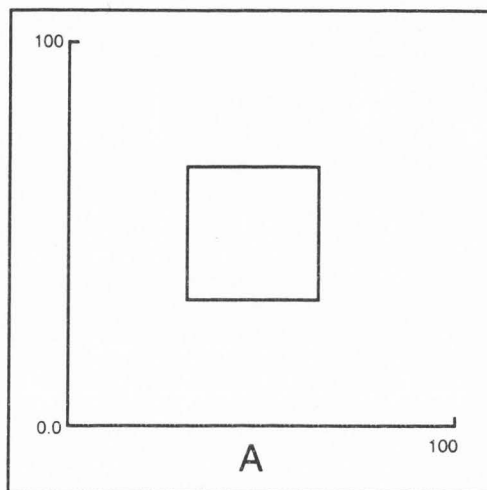
13. Angle D is what proportion of angle B?



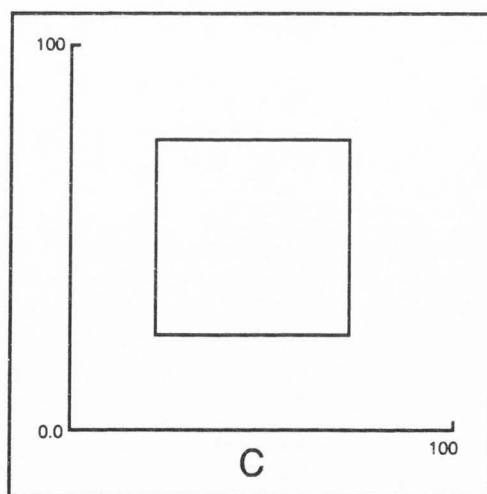
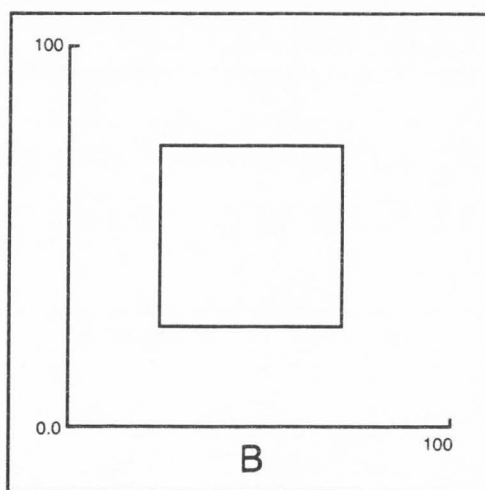
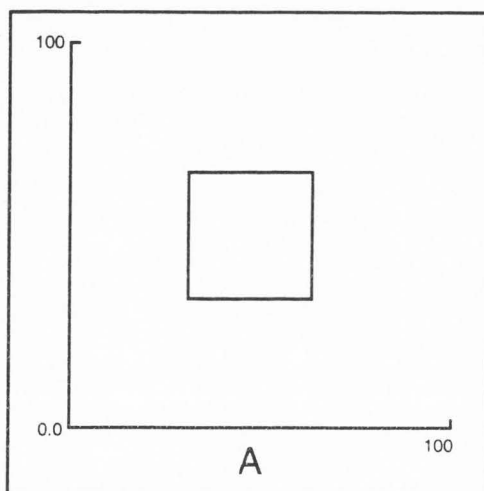
14. Point C is what proportion of point A?



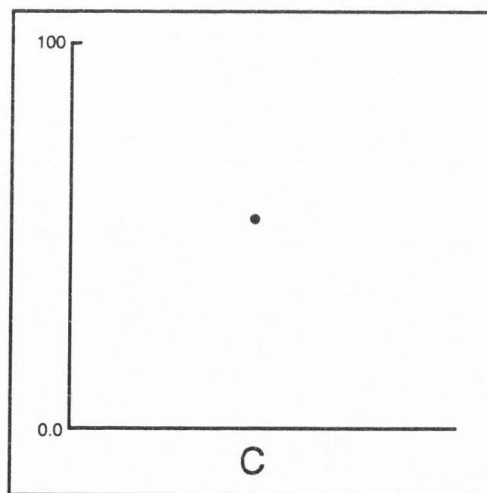
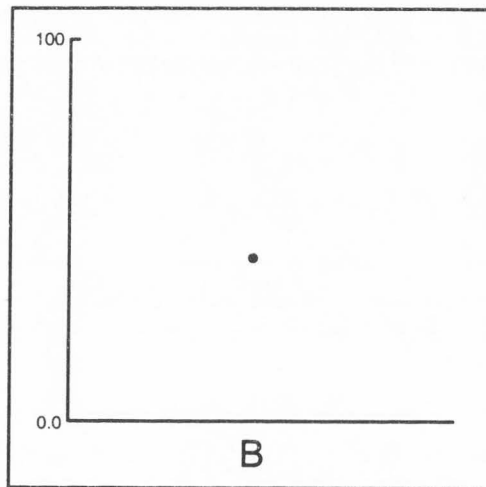
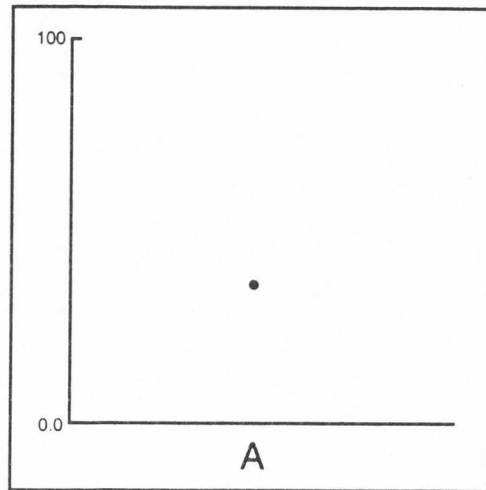
15. What is the value of bar A?



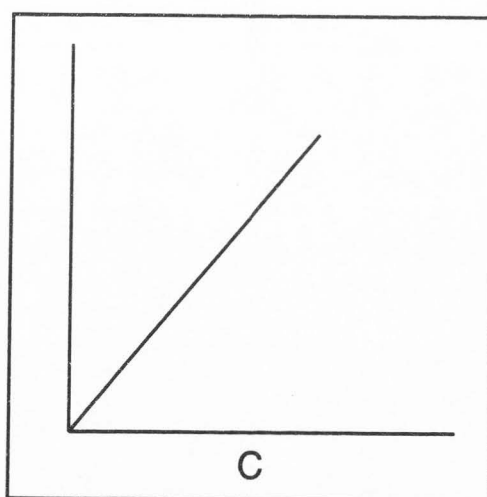
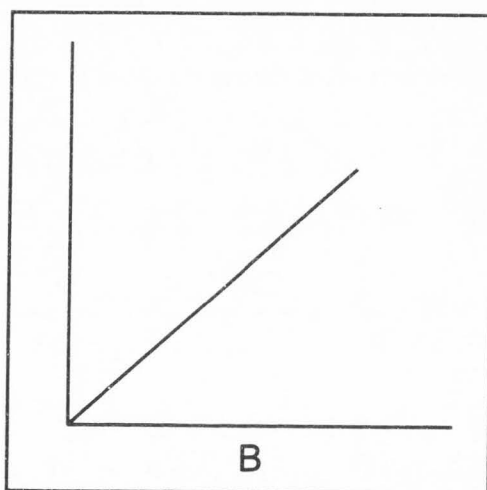
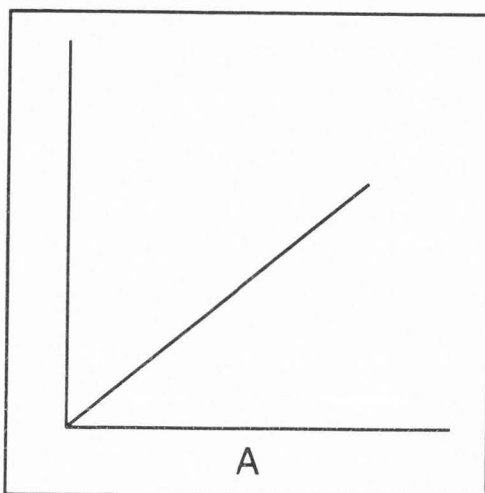
16. Square B is what proportion of square D?



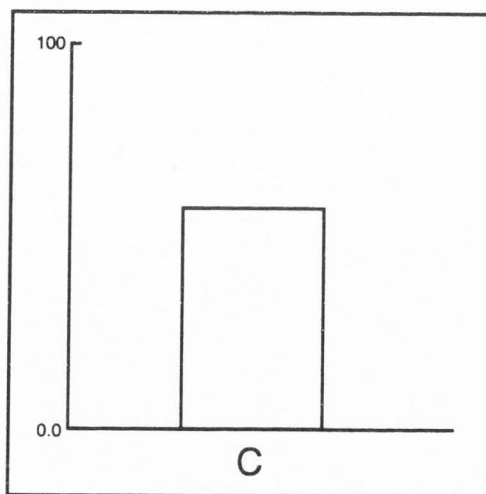
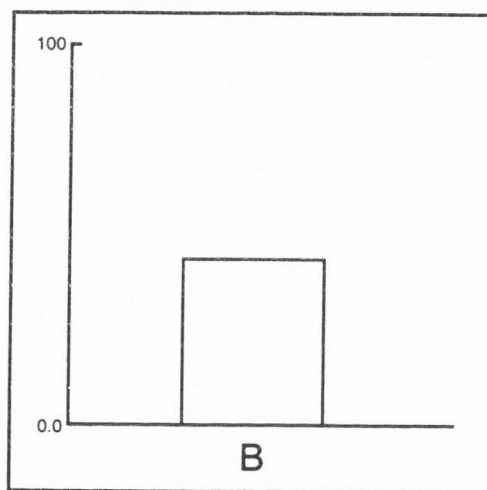
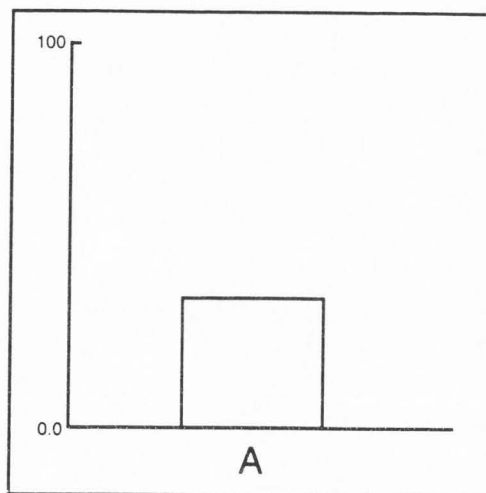
17. Square A is what proportion of square B + square C?



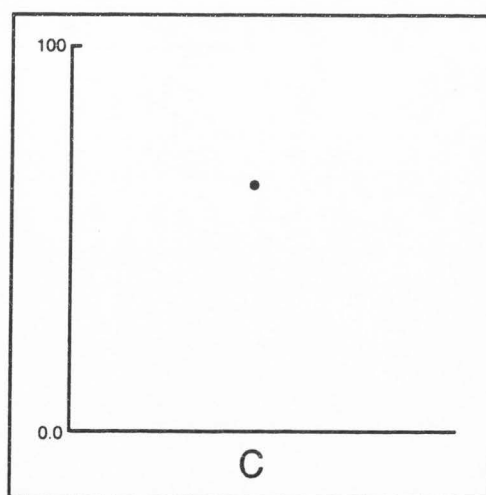
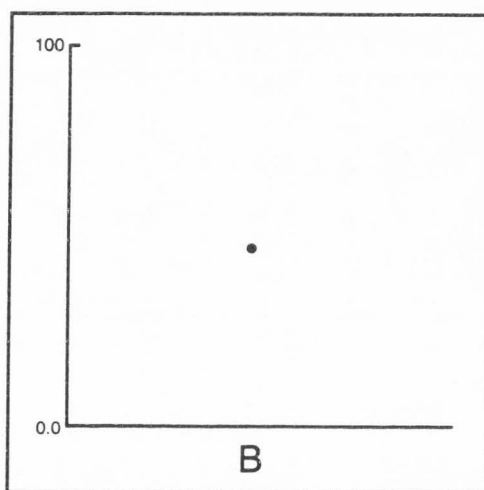
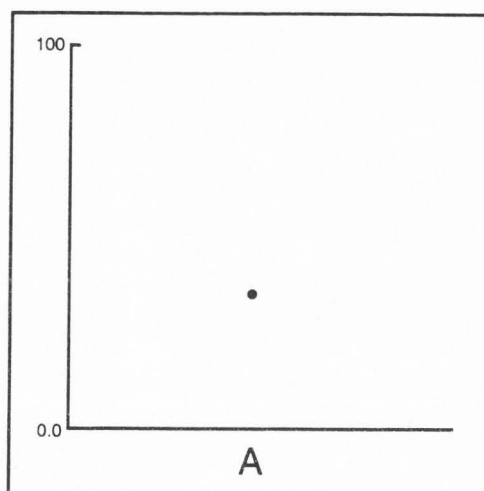
18. What is the value of point C?



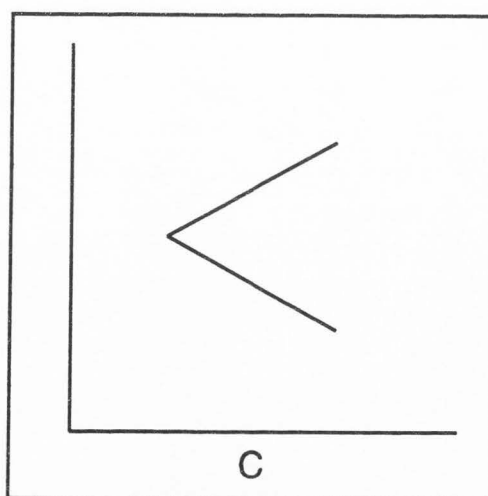
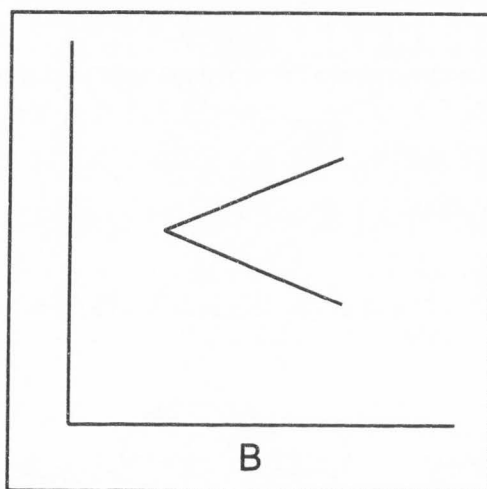
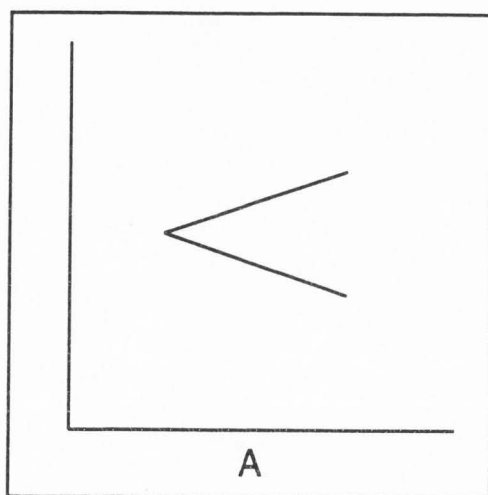
19. Slope C is what proportion of slope A + slope B?



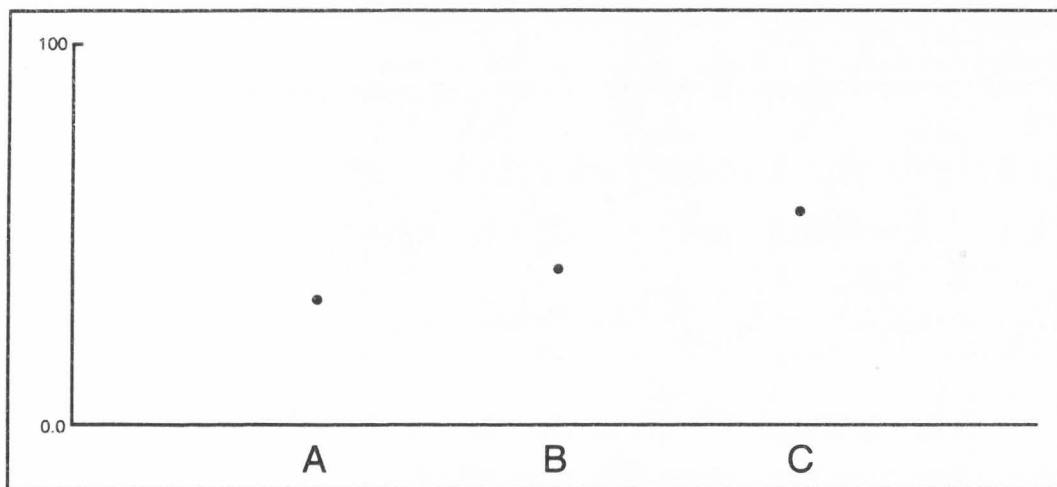
20. Bar D is what proportion of bar B?



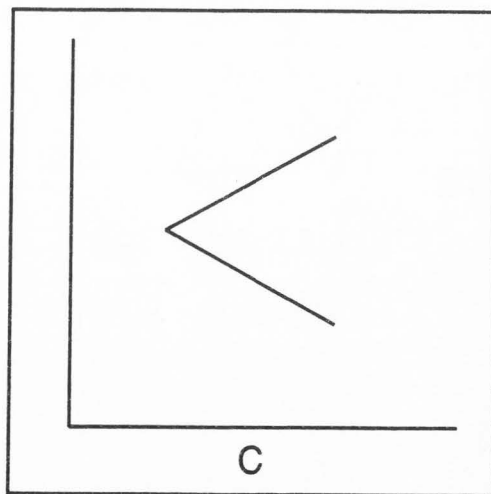
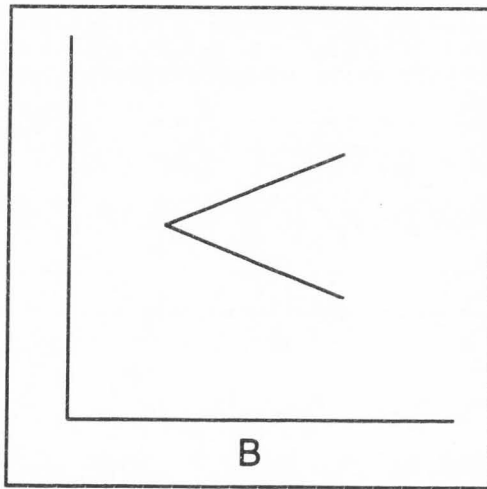
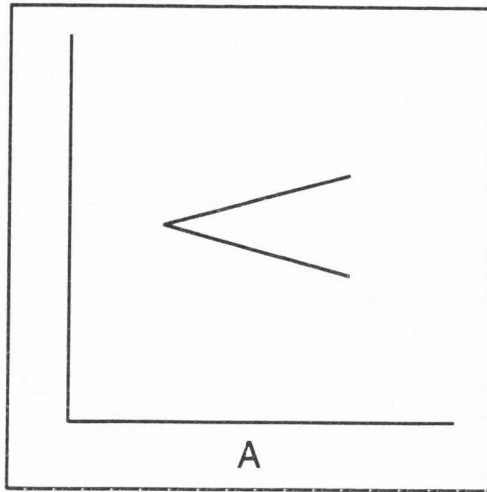
21. Point B is what proportion of point D?



22. What is the value of angle A?



23. Point A is what proportion of point B + point C?



24. Angle C is what proportion of angle A?

Appendix E. Statistical Analyses

Table E.1

Descriptive Statistics for Absolute Error Ranked in Accordance with the
Judgment Accuracy Associated with Each Item

Type of task	Statistics		
	N	M	SD
PR-PNS	124	4.3	3.7
PR-Length	124	4.4	4.7
PR-PCS	124	5.1	4.2
PR-Angle	124	6.5	5.6
GC-PNS	124	6.8	9.3
GC-PCS	124	6.9	8.2
PR-Slope	123	7.7	6.2
S-PNS	123	10.9	12.4
GC-Area	124	10.1	18.0
S-PCS	124	11.0	9.3
LC-Slope	124	11.5	9.2
S-Area	124	11.5	11.2
GC-Angle	124	11.9	23.1
LC-Area	124	14.1	12.1
GC-Slope	124	14.8	16.5
GC-Length	124	15.1	14.3
LC-Length ^a	124	18.2	12.6
LC-PCS ^a	124	20.7	16.6
S-Length ^a	124	26.0	27.8
S-Angle ^a	123	27.3	20.9
LC-PNS ^a	124	31.6	24.7

(Table Continues)

Type of task	Statistics		
	<u>N</u>	<u>M</u>	<u>SD</u>
S-Slope ^a	123	33.2	27.1
LC-Angle ^a	124	39.9	26.2

Note. PCS=positions along a common scale; PNS=positions along identical nonaligned scales; LC=local comparison; PR=point reading; GC=global comparison; S=synthesis.

^aValues indicate test items for which the proportions being judged were greater than 100%.

Table E.2

Straight-Error Descriptive Statistics for Test Items

Type of task	Statistics		
	<u>N</u>	<u>M</u>	<u>SD</u>
LC-Angle ^a	124	-39.2	27.2
S-Slope ^a	53	-34.1	19.6
LC-PNS ^a	124	-26.8	29.9
S-Angle ^a	123	-25.6	22.9
S-Length ^a	124	-22.1	31.0
LC-PCS ^a	124	-14.9	21.9
LC-Length ^a	124	-11.6	18.9
LC-Slope	124	1.1	14.7
S-PNS	123	1.3	16.5
PR-Length	124	1.5	6.2
GC-PNS	124	1.6	11.4
GC-PCS	124	1.8	10.6
S-Area	124	2.3	15.9
PR-Angle	124	2.4	8.3
PR-PNS	124	2.9	4.9
PR-PCS	124	3.7	5.4
PR-Slope	123	4.9	8.6
GC-Angle	124	5.0	25.5
GC-Length	124	6.0	19.9
GC-Slope	124	6.3	21.3
GC-Area	124	7.3	19.8

(Table Continues)

Type of task	Statistics		
	N	M	SD
LC-Area	124	10.6	15.3
S-PCS	124	11.0	14.2

Note. PCS=positions along a common scale; PNS=positions along identical nonaligned scales; LC=local comparison; PR=point reading; GC=global comparison; S=synthesis.

^aValues indicate test items for which the proportions being judged were greater than 100%.

Table E.3.

Intercorrelations Between Test Items

Type of task	PR-Angle	GC-PCS	LC-Angle ^a
LC-Slope	.05	.15	.18*
S-PCS	.00	.07	.07
PR-PCS	.01	-.03	.05
LC-PNS ^a	.07	.15	.09
PR-Area	-.06	.00	-.05
GC-Angle	.04	.12	-.06
GC-Length	.10	.34**	.22*
PR-Slope	-.11	.02	.01
LC-Area	.14	-.01	.20*
GC-PNS	-.07	.56**	.07
LC-Length ^a	-.08	.04	.13
S-Arngle ^a	-.00	-.01	.08
LC-PCS ^a	-.03	.10	.25*
PR-Length	-.04	.00	.12
S-Area	-.04	.02	.00
GC-Area	.14	.38**	.18*
PR-PNS	-.13	.12	.02
GC-Slope	.10	.34**	.11
S-Length ^a	.07	.08	.09
S-PNS	.16	.17	.03
PR-Angle	--	.09	.01
GC-PCS	--	--	.21*

(Table Continues)

Type of task	LC-Slope	S-PCS	PR-PCS	LC-PNS ^a	GC-Angle	S-Slope ^a
LC-Slope	--	.21*	.14	.06	.28**	.12
S-PCS	--	--	-.08	.24**	.04	-.01
PR-PCS	--	--	--	-.10	-.14	.00
LC-PNS ^a	--	--	--	--	.07	-.06
GC-Angle	--	--	--	--	--	.42**
S-Slope ^a	--	--	--	--	--	--

Type of task	GC-Length	PR-Slope	LC-Area	GC-PNS	LC-Length ^a	S-Angle ^a	LC-PCS ^a
LC-Slope	.32**	.10	.07	.09	-.08	.02	-.05
S-PCS	.34**	.34**	.03	.04	.10	.14	.13
PR-PCS	-.04	.22*	-.08	-.11	.04	-.10	-.06
LC-PNS	.28**	.17	.09	.09	-.00	.24**	.20**
GC-Angle	.17	-.05	-.00	.20*	-.06	-.07	-.09
S-Slope	.05	-.16	-.01	-.03	-.05	.06	-.08
GC-Length	--	.04	.09	.21*	.15	.11	.14
PR-Slope	--	--	-.15	.05	-.09	-.01	.11
LC-Area	--	--	--	-.05	-.10	.13	.09
GC-PNS	--	--	--	--	-.12	-.05	.12
LC-Length ^a	--	--	--	--	--	.19*	.40**
S-Angle ^a	--	--	--	--	--	--	.03
LC-PCS	--	--	--	--	--	--	--

Type of Task	PR-Length	S-Area	GC-Area	PR-PNS	GC-Slope	S-Length ^a	S-PNS
LC-Slope	-.02	-.05	.01	-.05	.28	.22**	-.01
S-PCS	-.08	.05	.03	-.02	.19*	.03	.25**

(Table Continues)

Type of task	PR-Length	S-Area	GC-Area	PR-PNS	GC-Slope	S-Length ⁵	S-PNS
PR-PCS	.19*	-.02	-.04	.37**	-.07	.12	.10
LC-PNS ^a	-.13	.06	.06	.09	.19*	.01	.11
GC-Angle	.04	.07	.15	-.16	.21*	.28**	-.03
S-Slope ^a	-.01	.35**	-.02	-.10	.13	.23**	-.03
GC-Length	-.05	.09	.11	-.01	.30*	.20*	.24**
PR-Slope	-.02	-.01	-.02	.17	-.06	-.03	-.09
LC-Area	.11	.11	.00	-.11	-.02	-.14	.13
GC-PNS	.10	.00	.36**	.26**	.47**	.00	-.03
LC-Length ^a	-.05	.04	.04	-.07	.01	.16	.18*
S-Angle ^a	-.06	.09	.04	.06	.01	.05	.03
LC-PCS ^a	.03	.02	.05	-.11	.11	.15	.02
PR-Length	--	.10	.10	.08	-.02	.02	.03
S-Area	--	--	.13	-.07	.20*	.20*	.06
GC-Area	--	--	--	.08	.34**	.05	.00
PR-PNS	--	--	--	--	.03	-.07	.01
GC-Slope	--	--	--	--	--	.10	-.02
S-Length ^a	--	--	--	--	--	--	.10
S-PNS	--	--	--	--	--	--	--

Note. PCS=positions along a common scale; PNS=positions along identical nonaligned scales; LC=local comparison; PR=point reading; GC=global comparison; S=synthesis. Bolded variables represent those questions for which the proportions being judged were greater than 100%. Correlation coefficients deriving from items have also been bolded. Correlation coefficients deriving from such items have also been italicized.

^aValues indicate test items for which the proportions being judged were greater than 100%.

* $p < .05$. ** $p < .01$.

Appendix F. Pilot Testing: Methodology and Findings

Issues

Prior to actual experimentation, it was necessary to ensure that the instructions given to participants were clear. As is the case for any study, subjects must understand the instructions given to them or they may produce responses or behaviors that confound the data set or yield, artifactual findings. The focus of the pilot study was on determining whether the instructions would enable subjects to participate in the intended manner.

Materials, Design and Procedure

The materials, design, procedure and subject selection used in pilot testing were identical to those that were outlined in the "Experimental Design and Methodology" section contained in the body of the paper, with the following exceptions:

1. For the pilot test, only 16 subjects participated. There were nine subjects in the first pilot test and seven in the second test.
2. Pilot tests were conducted on two occasions, so that data collected during the first testing session could be integrated into the experimental procedure and then evaluated during the second pilot test.
3. At the beginning of the experiment, subjects were told that one of the purposes of the experiment was to determine the adequacy of the instructional script. Therefore, subjects were asked to pay particularly close attention to the instructions given. Subjects were encouraged to voice concerns and thoughts about how to improve the instructions at any time during testing.
4. Subjects were taken through only the eight practice items instead of the entire test.

Analysis

In this study, the experimenter used the *Pilot Group Problem Checklist* to quickly identify certain types of problems raised by pilot subjects. The problem categories listed on this checklist were: *unable* (i.e., inability of the subject to accomplish the task in spite of solutions offered), *lag time* (i.e., the time span between the administration of the instructions and the task), *vocabulary* (i.e., wording used in the instructional script), and *other*. Similarly, the *Resolution of Problems Raised by Pilot Subjects* (also

located in Appendix C) was used to write a brief description of how such problems listed on the Group Problem Checklist would be solved.

When such concerns were raised, they were noted (see the Pilot Group Problem Checklist in Appendix C) and discussed with the entire group. To this end, the data collected were qualitative in nature and the data collection procedure resembled that for a focus group. To analyze subjects' responses about the clarity of the instructions, discussion was conducted regarding each concern raised and a solution to the problem was then advanced by either the experimenter or pilot subjects. The proposed solution was then discussed until a general consensus about a solution was reached.

Pilot Testing: Problems and Resolutions

Pilot test 1. The first issue of concern related to the wording on the answer sheet. The answer sheet originally had a space for "Subject Name." Comments were made to the effect that such wording was intimidating. The wording was changed to read, "Name."

The second issue related to the answer sheet as well. Pilot participants were willing to let the experimenter collect their GRE Composite score and GPA from the Admissions and Records Department. However, participants suggested that their willingness might be increased if they knew something about why such data were necessary. Participants were informed that the experimenter could not discuss the need to correlate data such as academic achievement with performance on the graphicacy test, because this would expose some of the theoretical underpinnings of the study. To this end the script was not changed.

The flow of the presented instructions was a functional point of consideration. The experimenter had a difficult time returning to his instructions after manipulating the transparencies on the overhead projector. To make the transition between transparencies and instructions more seamless, the instructions individual practice items were placed on separate pages.

Subjects agreed that the discussion following the presentation of practice item A was lengthy and redundant. A section addressing the y-axis values, in detail, was omitted with the understanding that the

participants would be subsequently probed to ensure that they understood how the scale on the y axis should be utilized.

The discussion following practice item D was related to the responses expected of subjects. When subjects were asked to make proportional comparisons between graphs/specifiers, they were expected to write down percentages that represented their judgments. At first, subjects indicated that they were confused about what they should actually write in response to the item question. This issue was elaborated upon and reiterated in this part of the script.

The prologue to practice item F and the subsequent global comparison task were confusing to subjects. Four of the participants reported that they were actually attempting to mentally perform area calculations for squares being combined and compared. Of course, the point of the experiment was not to assess proficiency of mental calculating. Participants suggested that the word "size" should be used in conjunction with the word "area" wherever it occurs in the script. This amendment was made. Additionally, it was agreed that subjects should be warned against performing area calculations mentally. The script was amended to address such concerns. This amendment was made.

The last area of confusion related to practice items focusing on synthesis tasks. Participants were understandably intimidated by these tasks and they felt that if they were provided certain steps that were integral to successfully solving this type of problem, they would better understand what was expected of them in this type of task and be able to perform with less anxiety. To improve the instructions for the two synthesis tasks, the tasks were broken down into three steps: 1) estimation of the trend, 2) imaging of the next specifier in the series, based on this trend, and 3) answering the question posed in the problem. These three steps were discussed before and after the presentation of practice items G and H. Participants agreed that this approach would improve the comprehensibility of the instructions for synthesis tasks.

Pilot test 2. The only new problem that was raised during the second pilot test related to the term "proportion" which appears in most of the test questions. After some discussion, it was agreed that for the practice items, the experimenter should also use the term "percentage." For some of the pilot subjects, this

term helped with the comprehensibility of the tasks. No pilot subjects indicated that the joint use of these terms during the administration of practice items increased confusion. The amendment was made.

CURRICULUM VITAE

Derek Borman
Department of Psychology
Utah State University

ADDRESS, E-MAIL and PHONE:

182 W 1100 N
 Logan, UT 84341
 E-Mail: SLCPS@CC.USU.EDU
 (435) 750-6047

EDUCATION:

- | | | |
|--------------|---------------|---|
| 1999, May | Ph.D.: | Utah State University - Logan, UT.
<ul style="list-style-type: none"> • Program: Psychology (Research and Evaluation Methodology) • Dissertation: <i>An Investigation of Perceptual and Analytical Tasks Which Affect Graphicacy</i> |
| 1995, May | M.A.: | Humboldt State University - Arcata, CA.
<ul style="list-style-type: none"> • Program: Psychology (Academic Research) • Thesis: <i>An Investigation of Schema and Exemplar Explanations of Motor-Skills Memory</i> |
| 1990, August | B.S.: | Brigham Young University - Provo, UT.
<ul style="list-style-type: none"> • Minor: English |

TEACHING INTERESTS:

- Research/Evaluation Design and Methodology
- Cognition
- Social Psychology
- General Psychology
- Educational Psychology
- Learning and Motivation

TEACHING EXPERIENCE:

- 1999, Spring **Instructor** (4420), Utah State University - Logan, UT.
Undergraduate course in cognitive theories.
- 1998 - Present **Honors Advisor**, Utah State University - Logan, UT. Direct Honors
students in the development of independent psychology projects, including
research design, data collection and interpretation of data.
- 1998, Fall **Instructor** (Psych 6650), Utah State University - Logan, UT.
Graduate course covering behavioral principles of learning.
- 1998, Summer **Instructor** (Psych 366). Utah State University - Logan, UT.
Undergraduate, distance education class in educational psychology.
- 1996 - 1998 **Teaching Assistant** (Psych. 366). Utah State University - Logan, UT.
Undergraduate laboratory instruction in principles of educational
psychology.
- 1996, Winter **Guest Speaker** (Psych. 712). Utah State University - Logan, UT.
Lectured on the development of working memory.
- 1994, Fall **Guest Speaker** (Psych. 625). Humboldt State University - Arcata, CA.
Lectured on the skeletomotor system and collection of electromyogram
signals.
- 1994, Fall **Instructor** (Psych. 104). Humboldt State University - Arcata, CA.
Team taught an introductory psychology class.
- 1994, Spring **Teaching Assistant** (Psych. 104). Humboldt State University - Arcata,
CA. Undergraduate laboratory instruction for introductory psychology.
- 1994, Spring **Guest Speaker** (Psych. 623). Humboldt State University - Arcata, CA.
Lectured on cognitive mechanisms mediating the optical extraction of
- 1993, Fall **Guest Speaker** (Psych. 681). Humboldt State University - Arcata, CA.
Lectured on memory concepts, research, and theory.
- 1990 - 1991 **Instructor** (Remedial English). Loral Ruff High School - Sacramento,
CA. Taught grammar to at-risk youths.

RESEARCH INTERESTS:

- Memory
- Social appraisal
- Learning styles
- Social influences of mass media
- Metacognition

RESEARCH EXPERIENCE:

- 1998 - Present **Content Specialist.** Research and Correlation Division (L.D.S. Church) - Salt Lake City, UT. Performed literature review of issues related to the influences of mass media on adolescents.
- 1998 - 1999 **Content Specialist.** Research and Correlation Division (L.D.S. Church) - Salt Lake City, UT. Performed literature review and meta-analyses of factors affecting survey response rates to mailed surveys.
- 1997 - Present **Dissertation Research.** Utah State University - Logan, UT. Performed literature review of graphicacy research, developed graphicacy test, collected and statistically analyzed data.
- 1994 - 1995 **Thesis Research.** Humboldt State University - Arcata, CA. Performed literature review of theories related to the acquisition, retention and reproduction of motor skills, constructed data collection apparatus, wrote computer program to collect data, statistically analyzed data.
- 1994 - 1995 **Psychology Department Assistant.** Humboldt State University - Arcata, CA. Advised undergraduate students regarding the design and statistical analysis for psychology research projects.
- 1993 - 1995 **Graduate Research Assistant.** Humboldt State University - Arcata, CA. Created job-satisfaction questionnaire for Simpson, Inc. Developed inservice curricula for Eureka City Police Department.
- 1993 - 1993 **Data Analyst.** Heidi Stromberg - Arcata, CA. Hired as a data analyst for a dissertation study being conducted by a psychology doctoral student.
- 1992 - 1993 **Research Assistant.** University of Utah - Salt Lake City, UT. Assisted with the design, data collection and data analysis for a study of the use of mnemonic techniques with senior adults.

EVALUATION/TEST DEVELOPMENT EXPERIENCE:

- 1998 **Project Coordinator.** Western Institute for Research and Evaluation - Logan, UT. Developed phone interviews designed to assess the status of the Junior Achievement high school programs throughout the United States. Formed team undergraduates to collect these phone-interview data from each of the approximately 250 Junior Achievement sites. Coded, synthesized, and analyzed the data.
- 1997 - Present **Test Developer.** Utah State University - Logan, UT. Oversaw the computerization of the Trio Measure of Visual Processing Ability.
- 1997 - Present **Evaluation Specialist.** Western Institute for Research and Evaluation - Logan, UT. Assist with research focusing on education in lower socioeconomic populations.
- 1996, Fall **Evaluation Specialist.** Fit Kids of Utah Project, Western Institute for Research and Evaluation - Logan, UT. Designed, conducted and analyzed focus groups and phone surveys for Fit Kids of Utah.
- 1995 - 1996 **Evaluation Specialist.** Junior Achievement Project, Western Institute for Research and Evaluation - Logan, UT. Designed interviews. Organized team of undergraduate and graduate students to synthesize the curriculum content and objectives of the three Junior Achievement programs being evaluated. Developed tests designed to assess the effectiveness of Junior Achievement Curricula.

CLINICAL EXPERIENCE:

- 1992 - 1993 **Case Supervisor.** Youth Support Systems - West Valley City, UT. Designed and implemented home-based, behavioral modification programs for children who had been discharged from inpatient psychiatric settings.
- 1989 - 1990 **Psychiatric Technician.** Charter Canyon Hospital - Orem, UT. Was responsible for tracking patients' psychological and physiological progress. Taught self-reliance, self-monitoring, and listening skills classes.

MANUSCRIPTS IN PROGRESS:

Borman, D., Donaldson, K., & Van Dusen, L. (1998). *An Exemplar Explanation for the Encoding, Storage and Recall of a Highly Discrete Movement*. Manuscript submitted for publication.

Borman, D., & Van Dusen, L. (1998). *Self-Adapted Testing: A Review of the Issues Related to the Use of Self-Adapted Testing as a Practical Alternative to Computerized-Adaptive and Fixed-Item Testing*. Unpublished manuscript.

PRESENTATIONS:

Borman, D., Spach, J. D., & Van Dusen, L. (1999). *A Validation Study of the Trio Measure of Visual Processing Ability*. Poster session to be presented at the annual meeting of the American Psychological Association.

Borman, D., & Donaldson, K. (1997). *Multiple memory systems and their effects on the formation of motor schemata*. Poster session presented at the annual meeting of the Western Psychological Association, Seattle, WA.

Borman, D., Foxley, D., & Hill, R. (1995). *Enhancing older adults' digit-string memory with phonetic mnemonics*. Poster session presented at the annual meeting of the Western Psychological Association, Los Angeles, CA.

Stromberg, H., Henderson, B., & Borman, D. (1994). *Denial, locus of control, and anxiety issues of Type I and Type II diabetics*. Paper presented at the annual meeting of the Northcoast Healthcare Providers, Eureka, CA.

PROFESSIONAL AFFILIATIONS:

- American Psychological Association
- Western Psychological Association

AREAS OF EXPERTISE:

- **Statistical Analysis**
Including: Descriptive and Inferential Statistics, Multivariate Analyses, Factor Analysis, Regression Analysis, General Linear Model, Discriminant Analysis, Cluster Analysis, Loglinear Models and Trend Analysis.
- **Test and Instrument Design**
Including: Item Response Theory, Classical Test Theory, Generalizability Theory, Self Adapted Testing, Computerized Adaptive Testing, and an understanding of general psychometric principles.
- **Quantitative and Qualitative Research Methodology and Design**
Including: Ethnography, Triangulation Methodologies, Meta-analysis, Observation, Group Testing, Focus Group Interviewing Techniques, and Integrative Reviews.

- **Program Evaluation for Education and Administration**
Including: Objective-based and Curriculum-based.
- **Computer Programs**
Including: SPSS, QUALSTAT, MS Office, Quattro Pro, and Basic Programming.
- **Multimedia Packages**
Premiere, Illustrator, Photoshop, Sound Forge, Powerpoint, Liveworks.

SCHOLARSHIP:

- Vaughn L. Heatherington Excellence in Leadership Scholarship, Aetna Life and Casualty Insurance (1987) - Rancho Cordova, CA.

SOCIETIES AND HONORS:

- Pi Gamma Mu
- Psi Chi
- Phi Theta Kappa

VOLUNTEER:

- | | |
|-------------|--|
| 1989 - 1990 | Instructor. Utah State Hospital - Provo, UT. |
| 1989 - 1990 | Paraprofessional. Brigham Young University - Provo, UT. |
| 1988 - 1990 | Rape Crisis Advocate, Utah Valley Rape Crisis Center - Provo, UT. |

RELEVANT UNDERGRADUATE ACTIVITIES:

- | | |
|-------------|--|
| 1987 - 1988 | Copy editor and staff writer, <i>The Daily Universe</i> , Brigham Young |
| 1987 - 1988 | Staff writer, <i>The Student Review</i> - Provo, UT. |