SEX DIFFERENCES IN MENTAL ROTATION ABILITY AND THE EFFECTS OF TWO TRAINING METHODS

By

TERESA LYNNE RUSSELL

Bachelor of Science in Arts and Sciences
Oklahoma State University
Stillwater, Oklahoma
1980

Master of Science in Arts and Sciences
Oklahoma State University
Stillwater, Oklahoma
1983

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
December, 1988

SEX DIFFERENCES IN MENTAL ROTATION ABILITY AND THE

EFFECTS OF TWO TRAINING METHODS

Thesis Approved:

William E. Jaynes

Thesis Advisor

Norman N. Durham

Dean of the Graduate College

ACKNOWLEDGMENTS

Over the course of my graduate training, a number of people have provided the kind of support, guidance, and mentoring that enable a student to grow and develop effectively. Obviously, understanding of general principles and knowledge of previous research in different areas of psychology are necessary building blocks for a career in our field. For this reason, I am grateful to the faculty of the OSU psychology department; cumulatively, the efforts of the faculty equipped me with a broad base of educational experiences that I have found to be a solid foundation for dealing with a variety of applied research questions. Specifically, I thank Drs. Helm and Hochhaus for encouraging my participation in their programs and for sharing their understanding of issues and problems. Along the same lines, I thank the Dean of the Graduate College for allowing me to undertake educational experiences outside standard curricula for the purpose of career enhancement.

Dr. William Jaynes encouraged me, during undergraduate training, to seek higher education, and, after I began graduate school, he mentored me, identified the types of developmental experiences I needed, and helped me pursue those experiences. Over the last few years, he has been patient, yet supportive and persistent; needless to say, I am greatly indebted to him.

I thank Dr. Marvin D. Dunnette for giving me an opportunity rarely afforded graduate students to grow and to learn from him and his colleagues at the University of Minnesota and Personnel Decisions Research Institute (PDRI). Specifically, at PDRI, Dr. Norman Peterson and Ms. Janis Houston gave me the opportunity to work on a large-scale project, the Army's Project A; Dr. Leaetta Hough mentored me on writing skills and entrusted me with positions of responsibility on her projects; and Drs. Rod Rosse and Walter Borman have provided guidance and support over the last few years. There is no substitute for the expertise and encouragement offered by these people, and I am grateful for having had the opportunity to share in their research efforts.

Finally, I thank Norm Peterson who has been both my friend and my sounding board, and my father, Stanley Russell, who has given me strength and emotional support in countless times of need. With regard to the study described in this paper, I am also indebted to my dad and my sister, Janet Russell, for spending hours photographing three-dimensional objects and developing film.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$\overline{X}$	Mean Score
$\overline{X}_F$	Females' Mean Score
$\overline{X}_M$	Males' Mean Score
SD	Standard Deviation
df	Degrees of Freedom
E(MS)	Estimated Mean Squares
S_p	Pooled Standard Deviation
d	Standardized Mean Difference

CHAPTER I

INTRODUCTION

Background

The finding of male superiority on tests measuring spatial ability has spawned research in virtually every sector of psychology. Developmental and physiological psychologists and behavior geneticists pursue a biological locus of spatial ability and, in turn, sex differences (e.g., hemispheric lateralization, androgen levels, X-linked recessive gene). Experimentalists and cognitive scientists study sex differences from a process-oriented perspective, and, naturally, education, learning, and social theorists have proposed differential experiences of the sexes as mechanisms for sex differences. Because accumulating evidence suggests that spatial abilities are to some degree trainable, psychometric and individual differences researchers are trying to understand how individuals' spatial test scores change with time, practice, or other interventions and the effect of the change on reliability and validity. Research to date suggests that there will be no single answer to the questions posed; moreover, biological, cognitive, and experiential factors probably influence spatial test performance interactively.

The practical implications of spatial abilities research for applied settings are far reaching. Research on the trainability of spatial ability and the generalizability of training across different spatial ability measures may shed light on the activities and experiences likely to foster performance improvement and thus has implications for educational settings.

In the industrial arena, research has centered around the validity of spatial ability test scores as predictors of job effectiveness in many occupations (e.g., architecture, engineering, technical and mechanical fields). For example, the U. S. Army has sponsored one of the largest-scale research projects to date (i.e., involving over 40,000 soldiers), in part, to investigate the incremental validity provided by spatial, perceptual, and psychomotor test scores over that of the Armed Services Vocational Aptitude Battery (ASVAB). The Navy and civilian organizations are conducting similar investigations of less magnitude. Key issues for the fair and valid use of spatial ability tests in industrial settings revolve around the extent to which spatial abilities are trainable and the effect of performance improvement or change on test reliability and validity.

Overview

This paper describes current understanding about spatial abilities and explains the methods and procedures used in a study of the trainability of spatial ability. It was prepared in partial fulfillment of the requirements of the doctoral program in psychology at Oklahoma State University and is divided into five chapters: I Introduction, II Literature Review, III Methods and Procedures, IV Results, and V Conclusions.

Chapter II provides background information about spatial ability research. The history of spatial ability, its relationship to other constructs such as mathematical ability, research relating to sex differences in spatial ability, and previous attempts to train spatial skills are discussed, and the hypotheses for the current research effort are posed.

Chapters III through V focus on a study designed to examine the effect of training on subjects' performance on mental rotation tests, the performance of females relative to males under different training conditions, and the generalizability of training on one test to performance on a different test. The methods and procedures used in the current study are described in Chapter III, and the results are provided in Chapter IV. Chapter V provides commentary and conclusions.

CHAPTER II

ISSUES AND EVIDENCE RELATED TO SEX DIFFERENCES IN SPATIAL ABILITIES: A LITERATURE REVIEW

Literature reviews and research articles on sex differences in spatial abilities often include an assertion such as "male superiority on tasks requiring these (spatial) abilities is among the most persistent of individual differences in all the abilities literature" (McGee, 1979, p. 895). It is, therefore, not surprising that a recent article by Caplan, MacPherson, and Tobin (1985), attempting to refute sex differences in spatial abilities, drew harsh criticism from a number of spatial abilities researchers (Burnett, 1986; Eliot, 1986; Halpern, 1986; Hisock, 1986; Sanders, Cohen, & Soares, 1986). Indeed, a sex difference in performance on psychometric measures of spatial abilities has been reported in numerous studies (e.g., Samuel, 1983; Sanders, Soares, & D'Aquila, 1982; Wilson & Vandenberg, 1978) and has been addressed in major reviews of the cognitive abilities literature (e.g., Anastasi, 1958; Maccoby and Jacklin, 1974; Tyler, 1965). What issues then led Caplan et al. (1985) to assert that sex differences in spatial ability are virtually nonexistent? In sum, Caplan et al.'s (1985) arguments against the frequently cited sex difference favoring males on tests of spatial abilities are three-fold: 1) questioning the construct validity of spatial ability(ies), the authors argue that it is "premature" and "inappropriate" to investigate sex differences because of definitional problems with the term spatial ability; 2) the authors contend that sex differences are observed infrequently and, when observed, are trivial; and, 3) the authors assert that research does not support biological explanations of spatial ability. In effect Caplan et al. (1985) question the theoretical underpinnings around which spatial abilities research has been conducted and criticize the manner in which research results are reported.

The purpose of the current review is to summarize research related to spatial abilities, to sex differences in spatial abilities, and to potential explanations of sex differences in spatial abilities. It is divided into three major sections. The first section, What is Spatial Ability?, summarizes factor-analytic, content-analytic, and correlational research to explicate spatial variables, and the second section describes research related to sex differences in spatial ability. The third section reviews current hypotheses regarding sex differences in spatial ability and presents the hypotheses addressed in the current study.

What is Spatial Ability?

In their classic article on construct validation, Cronbach and Meehl (1955) outlined guidelines for examining construct validity. In short, construct validity is not a number but is instead founded in a network that explicates the construct's expected and observed relationships with other variables. Factor-analytic research explaining the covariance among specific tests, correlational studies and other types of research examining the relationships between the construct and other variables, all contribute to the construct's nomological network. What factor-analytic evidence bears on the construct validity of spatial abilities, and how do spatial abilities interrelate with other variables?

Factor-Analytic Research

Identification of a Spatial Factor

Many of the early studies attempting to isolate a spatial factor were conducted by British researchers. For example, McFarlane (1925) administered a series of ten tests to 172 boys and 184 girls between the ages of 10 and 11. The test included four wooden construction tests, a puzzle box, a painted cube test, and a plunger test. She factored the scores and found evidence for a group factor, practical ability, in addition to general intelligence (g) for boys. She described practical ability as the ability to be adept at judging concrete spatial relations, but because girls' scores correlated more highly with "g" than did boys' scores, Spearman (1927) claimed that MacFarlane's results could be explained by sex differences in experience with construction tasks and that her tests were actually inadequate measures of "g."

As research cumulated in the 1920s and '30s, support for the existence of a spatial factor grew. Brown and Stephenson (1933) identified a factor underlying several tests -- Code, Code Parts, Mazes, Pattern Perception, and Fitting Shapes -- and called the factor "Perceptual." El Koussey (1935), under the guidance of Stephenson, administered 28 tests to 162 boys. He found a group factor "k" in addition to "g" and described the eight tests loading on "k" as requiring the "ability to obtain and the facility to utilize, spatial imagery" (p. 86). In 1934, I. Macfarlane Smith (reported in Smith, 1948) obtained tests used by El Koussey from Dr. Stephenson, wrote new items, and formulated similar tests. Then, Smith administered nine spatial ability tests, the Otis Intelligence Test, and an unpublished intelligence test to 70 boys and 52 girls, 100 of whom completed all tests. His results were substantially the same as El Koussey's. Five tests, Fitting Shapes, Pattern Perception, Completion, Analogies, and Form Equations, loaded on "k" while the two intelligence tests did not.

Thurstone's research on primary mental abilities (Thurstone, 1938; Thurstone & Thurstone, 1941) was a landmark in abilities testing. Thurstone (1938) administered 56 tests designed to tap a wide range of abilities to 218 subjects. He extracted 13 factors but could only label nine: Perceptual Speed (P), Number (N), Verbal Relations (V), Word Fluency (W), Memory (M), Induction (I), Reasoning (R), Deduction (D), and one called Space (S). Five tests with the highest loadings on the Space factor were Flags, Lozenges B, Cubes, Pursuit, and Surface Development -- all of which require the ability to imagine the transformation of an object or figure in space. In a separate study of eighth grade children, Thurstone and Thurstone (1941) identified seven factors: Perceptual Speed (P), Number (N), Verbal Comprehension (V), Word Fluency (W), Memory (M), Inductive Reasoning (I), and Space (S), with three tests (Flags, Figures, and Cards) loading on the Space factor.

In the 50 years since Thurstone's initial work numerous studies (e.g., Vernon, 1950, 1960) have yielded at least one spatial factor mathematically distinct from verbal ability. In the 1970s and 1980s the debate among researchers (e.g., McGee, 1979; Lohman, 1979) hinges around the number and structure of spatial subabilities rather than the existence of a broad spatial construct.

First-Order Factors of Spatial Ability

Factor analysis is indeed a useful tool for delineating the organization as well as the number of meaningful subabilities. The number and structure of spatial abilities factors obtained across studies is, however, likely to vary with the tests included in the studies and factor-analytic method employed. If no, or too few, tests indexing a specific construct are included in a battery the construct will not emerge factor analytically. Also, the factor-analytic technique employed in the study is often guided by the researcher's implicit theory of the structure of the intellect [e.g., hierarchical (Spearman, 1904; Vernon, 1950) or non-hierarchical (Guilford, 1967)]. In turn, different algorithms can lead researchers to different conclusions about the same data. For example, a number of researchers have reanalyzed Thurstone's (1938) data base of scores. Recall that Thurstone had identified 13 orthogonal factors including one called Space. Reanalyzing Thurstone's data with different factor analytic methods, Spearman (1939) and Eysenck (1939) identified a general factor and subfactors including a spatial ability factor. Zimmerman (1953), on the other hand, found two spatial factors (one called Spatial

Relations and one called Visualization) and no general intelligence factor.

Although variation in factor analytic techniques and theoretical frameworks has led to inconsistencies, more than half-a-century of psychometric testing has provided important clues as to the number and type of spatial subabilities. Over the years, literally hundreds of studies have been conducted, too many to describe in detail in this review. Major studies designed to clarify the nature of abilities, in particular spatial ability, are summarized below.

In 1950, Thurstone reported three spatial factors: S1, S2, and S3. Factor S1 was defined by Figures and Cards tests, tests that require the ability to visualize a rigid configuration when it is moved into different positions. Mechanical Movements and Surface Development tests loaded significantly on S2, and Thurstone interpreted this factor as the ability to visualize a configuration in which there is movement or displacement among the internal parts of the configuration. Thurstone concluded that S3 was poorly defined because there was little apparent similarity between the two tests loading on the factor. He did, however, suggest that S3 might represent the ability to solve spatial problems in which the body orientation of the observer is an essential part of the problem.

Updating Thurstone's PMA work, Pawlik (1966) outlined 19 Primary Mental Abilities, six of which were spatial in content: Spatial Visualization (Vi), Spatial Relations (S), Spatial Orientation (SO), Speed of Closure (Cs), Flexibility of Closure (Cf), and Perceptual Speed (P). Table I lists marker tests for each of these factors.

TABLE I

MARKER TESTS FOR SPATIAL FACTORS FROM PAWLIK (1966)

Factor	Marker Tests
Spatial Visualization	Mechanical Movements
Spatial Relations	Cubes
Spatial Orientation	Complex Instrument Comprehension
Speed of Closure	Gestalt Completion, Speed of Dark Adaptation
Flexibility of Closure	Hidden Figures
Perceptual Speed	Mirror Reading, Identical Forms

As described by Pawlik (1966), Spatial Visualization referred to the ability to imagine the movement or displacement of a configuration or some of its parts, whereas Spatial Relations involved the ability to recognize the identity of an object when it is seen from different angles or in different positions. The ability to solve problems in which the body orientation of the observer is an essential part was termed Spatial Orientation. Flexibility of Closure described tasks requiring the extraction of a figure embedded in a distracting field, and Speed of Closure involved organizing configurations into a structured pattern. Tests loading on the Perceptual Speed factor involved speed in comparing visual configurations.

Guilford and his colleagues (Guilford & Lacey, 1947; Hoffman, Guilford, Hoepfner, and Doherty, 1968; Michael, Guilford, Fruchter, and Zimmerman, 1957) conducted a series of factor-analytic studies using Army Air Force (AAF) tests. They found strong support for two spatial factors, Visualization and Spatial Relations, and some support for two additional factors, S2 and S3. Marker tests for each of these factors are listed in Table II. S3, defined by the Two Hand Coordination Test, was, however, only found in one of many analyses and was dropped from the later discussions (1957, 1968). The Visualization factor was "strongest in tests that present a stimulus either pictorially or verbally, and in which some manipulation or transformation to another visual arrangement is involved" (Guilford and Lacey, 1947, p. 838). Tests loading on this factor included Mechanical Movements and Space Visualization I, a paper folding task. Spatial Relations was defined as requiring the ability to determine relationships between different

TABLE II

MARKER TESTS FOR SPATIAL FACTORS FROM GUILFORD & LACEY (1947)

Factor	Marker Tests
Visualization	Spatial Visualization I, Mechanical Movements
Spatial Relations	Flags, Figures, Cards, Cubes
S2	Hands, Flags
S3	Two Hand Coordination

spatially arranged stimuli and responses and the comprehension of the arrangement of elements within a visual stimulus pattern. Because Thurstone's Flags, Figures, Cards, and Cubes loaded on this factor, Guilford and Lacey believed it to be essentially the same as the one Thurstone called Space (Thurstone, 1938). The weaker factor, S2, was specific to Thurstone's Hands and Flags tests, and appeared to involve the ability to make right hand-left hand discriminations.

Over the last few decades, researchers at the Educational Testing Service (Ekstrom, 1973; Ekstrom, French, Harman, & Derman, 1976; Ekstrom, French, & Harman, 1979; French, Ekstrom, & Price, 1963) have conducted, reviewed, and summarized factor-analytic research in effort to develop an integrated listing of validated tests and constructs. In 1963, ETS published a Kit of Factor-Referenced Tests including psychometric marker tests for 24 constructs. Seven constructs were spatial in character: Flexibility of Closure, Speed of Closure, Spatial Orientation, Visualization, Spatial Scanning, Perceptual Speed and Figural Adaptive Flexibility. Marker tests for each of these factors are listed in Table III. According to Ekstrom (1973) Spatial Scanning and Figural Adaptive Flexibility were poorly defined. Expanding on this work, Ekstrom et al. (1976; 1979) reported refinements to the Kit. A factor was considered "established if the construct underlying it had been found in at least three factor analyses performed in at least two different laboratories or by two different investigators" (p. 3, 1976). On the basis of the research findings, Ekstrom et al. (1976, 1979) reported that 23 cognitive factors, each with two to five marker tests, had been replicated

TABLE III

MARKER TESTS FOR SPATIAL FACTORS FROM EKSTROM ET AL. (1979)

Factor	Marker Tests
Visualization	Paper Folding, Surface Development, Form Board
Spatial Orientation	Card Rotations, Cube Comparisons
Speed of Closure	Gestalt Completion, Concealed Words
Flexibility of Closure	Hidden Figures, Hidden Patterns, Copying
Spatial Scanning	Maze Tracing Speed, Choosing a Path, Map Planning
Perceptual Speed	Finding A's, Number Comparison, Identical Pictures
Figural Flexibility	Toothpicks, Planning Patterns

adequately. The factors listed above were retained; however, the authors indicated that the nature of Figural Adaptive Flexibility had not been clearly demonstrated and that new marker tests were needed to assess this construct.

Ekstrom et al. (1976) note that Spatial Orientation and Visualization have been particularly difficult to define. Comparing the tests that define these two factors, they suggested that Visualization (e.g., paper-folding tests) requires that the figure be mentally restructured into components for manipulation and defined it as "the ability to manipulate or transform the image of spatial patterns into other arrangements" (p. 173). Spatial Orientation tests (e.g., Card Rotations), on the other hand, require that the whole figure be manipulated or rotated, and Orientation was, thus, defined as "the ability to perceive spatial patterns or to maintain orientation with respect to objects in space" (p. 149).

Summarizing factor-analytic research through 1979, McGee (1979) contended that "the available evidence conclusively demonstrates the existence of at least two Spatial factors: Visualization and Orientation" (p. 890). He defines Spatial Visualization as involving "the ability to mentally rotate, manipulate, and twist two- and three-dimensional objects" (p. 896). Orientation is described as the ability to comprehend "the arrangement of elements within a visual stimulus pattern, the aptitude to remain unconfused by the changing orientations in which a spatial configuration may be presented, and the ability to determine spatial orientation with respect to one's body" (p. 897). It is important to note here that McGee's interpretations are at the level of factor definitions; little reference is given to marker tests for factors. His definition of Visualization is, thus, particularly difficult to trace because other authors have referred to the factor frequently formed by Thurstone's Flags, Figures, and Cards (two-dimensional rotation tests) as Spatial Relations (Guilford & Lacey, 1947) or Spatial Orientation (Ekstrom et al., 1976, 1979).

A more thorough review and reanalysis of factor-analytic research on spatial abilities was prepared by Lohman (1979). Lohman reviewed literally dozens of major spatial abilities studies, refactored and reanalyzed results, and compared results of various authors. He concluded that there is strong support for three spatial subabilities--Visualization, Orientation, and Space Relations--and that there is some empirical support for several minor factors. Lohman defined Spatial Relations as the ability to solve mental rotation problems (similar to McGee's Visualization) and suggested that the factor emerges only if Thurstone's Flags, Figures, and Cards, or highly similar tests are included in the battery. He later relabeled this factor Speeded Rotation (Lohman, 1988). Lohman did not define Visualization other than noting that it underlies complex spatial tasks that are relatively unspeeded. He suggested that Spatial Orientation involves reorienting an imagined self; that is, "subjects must imagine they are reoriented in space, and then make some judgment about the situation" (p. 188). According to Lohman, minor factors such as Speed of Closure, Perceptual Speed, and Spatial Scanning are narrow; they relate to very basic cognitive processes in contrast with Visualization tasks that require several transformations and complex processing.

In sum, McGee and Lohman concur on the definition of Spatial Orientation. Unfortunately, authors use different labels for the factor defined primarily by Thurstone's Flags, Figures, and Cards, or highly similar tests. McGee calls this factor Visualization, whereas Guilford and Lacey (1947) named it Spatial Relations, Lohman (1988) refers to it as Speeded Rotation, and others (Ekstrom, et al. 1976) have used the name Spatial Orientation. Even so, there does appear to be consistent support for two to three major spatial subabilities--Spatial Orientation, Speeded Rotation, and Visualization--as defined by Lohman.

Higher Order Factors

The work of Horn and Cattell (1966) and a recent study by Gustaffson (1984) illustrate how spatial subfactors might be related to one another. Horn and Cattell's hierarchical model of intelligence includes five higher order factors of general abilities, fluid intelligence (Gf), crystallized intelligence (Gc), General Visualization (Gv), General Fluency (F), and General Speediness (Gs). Although both Gf and Gc are viewed as aspects of general intelligence, Horn and Cattell suggest that Gf is involved in novel tasks and that Gc is shown in tasks reflecting education and experience. Gv appears in tasks that are spatial or figural in content.

Gustaffson investigated hypothesized relationships among hierarchical models of intelligence (i.e., proposed by Spearman, Vernon, Horn, and Cattell) and non-hierarchical models such as Thurstone and Guilford models. He administered thirteen ability tests (verbal and spatial) to 602 boys and 622 girls in the 6th grade. Ten first-order factors were obtained, five of which were spatial in content: Visualization, Orientation, Flexibility of Closure, Speed of Closure, and Cognition of Figural Relations. The marker test for Visualization was Metal Folding, a test in which the subject must find the three-dimensional object which corresponds to a two-dimensional drawing. The test used to define Orientation was Card Rotations which is very similar to Thurstone's Cards, Figures, and Flags; according to Lohman's classification the factor would, thus, be labeled Speeded Rotation. Disguised Words and Disguised Pictures loaded on Speed of Closure, and the Group Embedded Figures Test and Hidden Patterns loaded on Flexibility of Closure. The Raven Progressive Matrices test, a figural reasoning test, was the marker for Cognition of Figural Relations.

At the second order, Visualization, Orientation (Lohman's Speeded Rotation), Flexibility of Closure, and Mathematical Achievement loaded on Gv (Visualization, see also Horn and Cattell, 1966). Speed of Closure and Cognition of Figural Relations loaded on Gf (Cattell's Fluid Factor) along with Inductive Reasoning and Memory Span Factors. (Gc was related to Verbal Achievement, Vocabulary, and Mathematical Achievement). In turn, Gf was highly related to "g" (in the Spearman sense). Gustaffson concluded that Cattell's Gf is essentially identical with "g." Gustaffson's study suggests that Visualization, Orientation (Speeded Rotation), and Flexibility of Closure measures are tapping a higher-order spatial factor while Cognition of Figural Relations and Speed of Closure may be indexing general reasoning (g) or fluid (Gf) abilities. The finding that figural reasoning tests tend to load on a reasoning factor rather than a spatial factor has been corroborated by other investigators (cf., Horn & Cattell, 1966; Ekstrom et al. 1976).

Content-Analysis of Spatial Ability Tests

In contrast to factor-analytic research, Eliot (1980) identified clusters of spatial ability tests on the basis of perceived similarities of abilities required for task solution. He collected more than 300 spatial tests from commercial, out-of-print and experimental sources and content-analyzed them to develop a classification scheme in terms of the behaviors required for their solution. Two broad categories subsuming 12 more specific groups of tests resulted. One broad category, Matching, described tests for which the subject must match two stimuli and included maze or copying tasks, embedded-figures tasks, figural memory tasks, figural combination tasks, and two-dimensional rotation tasks. The second category contained block tasks, three-dimensional rotation tasks, paper-folding tasks, surface development tasks, perspective tasks, combination tasks, and figural collage tasks--all presumed to require manipulation or transformation as well as matching. Because factor-analytic researchers typically interpret factors on the basis of perceived similarity of types of items on covarying tests, content-derived categories and empirical factors should converge. Indeed, Eliot's categorization scheme is reminiscent of distinctions among spatial factors discussed previously. For instance, embedded figures, copying, and hidden patterns tests are purported to index Flexibility of Closure (e.g., Ekstrom et al., 1976; Pawlik, 1966; Horn & Cattell, 1966) and are categorized similarly as Matching by Eliot. Likewise, Ekstrom et al., (1976) found that surface development and paper-folding load on a visualization factor and Eliot describes these as Manipulation tests. Eliot, however, classifies embedded figures tests together with two-dimensional rotation tasks, a classification that is not supported factor-analytically. Also, some of the figural relations tasks categorized by Eliot as matching tasks appear to be better measures of Gf than Gv. It, therefore, appears that further distinctions may be needed in Eliot's classification scheme before it will be highly useful in categorizing spatial tests.

Summary

Although there has been some disagreement over the specific titles of spatial subfactors, several findings appear rather consistently in spatial abilities research. First, there is clear and consistent evidence for at least one spatial factor accounting for a large portion of the variance among spatial ability tests, distinct from measures of verbal ability. Second, three spatial subabilities (i.e., Speeded Rotation, Spatial Orientation, and Visualization) appear to have strong support, and up to six subabilities (i.e., Speeded Rotation, Spatial Orientation, Visualization, Flexibility of Closure, Spatial Scanning, and Perceptual Speed) have some empirical basis. Speeded Rotation is measured by tests like Thurstone's Flags, Figures, and Cards. Spatial Visualization underlies complex spatial tasks such as paper-folding, and Spatial Orientation is shown in tasks that involve mentally reorienting oneself with respect to an object. Third, spatial ability requires the use of mental imagery to manipulate spatial representations.

Correlates of Spatial Abilities

Over the years, relationships between spatial abilities and a variety of perceptual-cognitive aptitudes have been suggested. In particular, investigators have attempted to delineate association of spatial abilities with mechanical and mathematical abilities and with cognitive style.

Mechanical and Mathematical Abilities

The study of spatial ability is inherently related to investigation of mechanical aptitude. A high correlation between the two has been well established (Bennett, Seashore, and Wesman, 1974), and mechanical ability tests have been shown to load on a spatial visualization factor, along with paper folding and other visualization tasks (Guilford & Lacey, 1947). In recent years, accumulating evidence also suggests a moderate relationship between spatial abilities and mathematical ability. For instance, Hills (1957) found that orientation and visualization test scores were related to grades in college mathematics. Similarly, Bennett, Seashore, and Wesman (1974) reported a correlation of .57 between DAT Spatial Relations and success in school geometry. Johnson (1984) administered the Clocks Test (a test from the Guilford-Zimmerman Aptitude Survey requiring mental rotation of pictures of clocks) to 137 males and 144 females. Clocks test scores correlated .56 and .52 with Mathematics scores on the Scholastic Aptitude Test (SAT) for males and females respectively. Similarly, problem solving performance correlated with Clocks Test scores, .53 (males) and .52 (females).

One hypothesis posed to explain observed relationships between mathematical ability and spatial ability (Hamley, 1935; Smith, 1964) is that high spatial ability enhances performance on higher level and graphic mathematics tests (e.g., those involving geometry) but is not necessarily related to basic arithmetic ability. This hypothesis is, however, based on post hoc observations. In the future, the patterns of correlations expected between different spatial and mathematical tests should be explicated prior to collecting data and interventions that test the hypothesis should be defined.

Recently, research has focused on large sex differences found in mathematical ability. In 1973, researchers at Johns Hopkins began a longitudinal study of Mathematically Precocious Youth. From 1980 to 1982, 19,883 boys and 19,937 girls in the 7th grade took the SAT. Stanley and Benbow (1982, and Benbow and Stanley, 1983) reported a 30 point difference favoring males (males' $\bar{X}$ = 416, SD = 87; females $\bar{X}$ = 386, SD =74) in SAT-M scores and no difference in SAT-V scores. Although a 30 point difference, slightly over one third of a standard deviation, might appear small, Benbow and Stanley illustrate marked effects on the upper end of the distribution. The ratio of boys to girls receiving SAT-M scores greater than 420 is 1.5 to 1. The ratio among those who scored 500 or more is 2.1 (males) to 1 (females). Six hundred forty eight males and one hundred females scored 600 or greater, a ratio of 4.1:1. Although scoring 700 or more on the SAT-M in the 7th grade is rare, one hundred and forty seven boys and only eleven girls did so, a ratio of 13.4 to 1.

Because spatial abilities typically correlate with mathematical abilities, investigators have suggested that the sex difference in mathematical ability may be a manifestation of the sex difference in spatial ability (Burnett, Lane, & Dratt, 1979; Smith, 1964; Sherman, 1967). Before this hypothesis can be assessed with any degree of certainty, the relationship between spatial and mathematical abilities must be clarified without regard to the sex factor; that is, it must first be shown that spatial ability does enhance mathematical skills.

Along the same lines, sex differences in mechanical ability are even larger than those reported for mathematical or spatial ability (discussed later in this review), and given a strong relationship between mechanical and spatial abilities, sex differences in mechanical ability could be, in part, due to the spatial nature of mechanical ability tests. Further research is needed, however, to explicate the spatial - mechanical and spatial - mathematical abilities relationships.

Cognitive Style

Since Witkin and his colleagues initially proposed the concept of field dependence (Witkin, 1950; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962), literally thousands of studies have been conducted in this arena. According to the theory, persons identified as field dependent are sensitive to the stimulus background and cannot disregard it, whereas field independents are less sensitive to context and can extract or separate figure from context. A number of measures have been used to tap field independence/dependence (e.g., Embedded Figures, Rod-and Frame-Test, rotary-match brightness constancy task, body steadiness task), embedded figures tests being used most frequently. Embedded figures and the Rod-and-Frame Test, however, are spatial in nature and typically correlate highly with other spatial ability measures [see McGee (1979), McKenna (1984), and Sherman (1967) for reviews of this literature].

Sex differences in field dependence have been reported frequently (e.g., Witkin et al., 1962; Witkin, Goodenough, and Karp, 1967), and females have thus been characterized as more field dependent than men. Sherman (1967) and McGee (1979) have noted, however, that while significant sex differences are often found in performance on embedded figures and rod-and-frame tasks, non-spatial measures of field dependence (i.e., rotary-match brightness constancy, paper-square-match brightness constancy, and body steadiness) do not yield significant differences (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954). This led McGee (1979) to conclude that "the presence of a spatial component in tests of field dependence - independence seems to be a prerequisite for the appearance of sex differences" (p. 898). Likewise, Harris (1978) contends that sex differences in tasks with a high spatial component are an artifact of the spatial nature of these tests and cannot be generalized to cognitive style. Sex differences must be observed on non-spatial measures of cognitive style before sex differences in field dependence - independence can be shown to be more than an artifact of the sex difference in spatial tasks.

Summary

Over the years, several hypotheses regarding the relationships between spatial and mathematical abilities, spatial and mechanical abilities, and spatial ability and cognitive style have been posed, namely (1) high spatial ability might enhance development of higher mathematical ability, (2) sex differences in mathematical ability might result from sex differences in spatial ability, (3) sex differences in mechanical ability might be due to sex differences in spatial ability, and (4) sex differences in cognitive style might be an artifact of the sex difference in spatial ability. Of these, only the fourth can be inferred directly from existing data. To date, very little is understood about how spatial skills might facilitate mathematical and mechanical aptitude and, in turn, how the observed sex differences in mathematical and mechanical abilities might result from the sex difference in spatial ability. Before it can be shown that the sex difference in spatial ability plays an important role in the observation of sex differences on these other abilities, the nature of the relationship between mathematical and spatial ability and between spatial and mechanical abilities needs to be clarified.

Are Sex Differences in Spatial Abilities Inconsistent and Trivial?

Is male superiority on tests of spatial ability a consistent finding, and, if so, is the magnitude of the effect trivial? The observation of a male advantage on tests of spatial ability dates back to the initial identification of spatial abilities. For example, McFarlane (1925) administered a series of ten tests designed to measure "practical ability" to 172 boys and 184 girls between the ages of 10 and 11. Boys outscored girls on all but two of the tests. Similarly, Smith (reported in Smith, 1948) compared scores of 70 boys ($\overline{X}$ = 60.23, SD = 13.73) and 52 girls ($\overline{X}$ = 52.40, SD = 11.69) on his spatial battery and noted over half of a standard deviation difference between the two means. Boys and girls did not differ, however, on the Otis Intelligence Test. Smith related this finding to differences that had emerged in mechanical ability tests (Earle, 1929). Likewise, Hobson (1947) found about half a standard deviation difference between scores of 720 boys ($\overline{X}$ = 40.58, SD = 20.07) and 716 girls ($\overline{X}$= 31.45, SD = 18.89) on the Primary Mental Abilities Space factor, even though the girls had scored higher than boys on the Kuhlmann-Anderson IQ test. Hobson considered this difference to be "a real one of considerable magnitude" (p. 129) given the expected relationship between Spatial Orientation and performance in vocational fields such as architecture and engineering.

Over the years, a sex difference in performance on measures of spatial ability has been discussed in virtually every major review of the cognitive abilities literature (e.g., Anastasi, 1958; Harris, 1981; Maccoby and Jacklin, 1974; Tyler, 1965). Conclusions drawn in narrative reviews, however, can be problematic for two reasons. First, reviewers sometimes use a box score approach in summarizing data. That is, significant results are tallied against insignificant findings and either "effect" or "no effect" is declared a winner. Second, reviewers almost always summarize findings across different types of spatial ability tests (and sometimes across significant variables such as age). Aggregating results across variables that contribute to variance across studies attenuates and obscures the magnitude of any "true" effects present in the data.

Problems with the box score approach are exemplified by a recent debate over data on the Porteus Maze Test (1965). The Porteus Maze Test, published about the time of World War I, consists of maze problems. In 1965, Porteus reported the results of over 40 years of research with this instrument. In sum, males obtained higher test scores than females in 99 out of 105 studies. Porteus, however, only computed significance tests for 18 of the studies, of which four were significant.

Comparing the number of significant results to insignificant or unreported significance levels, Caplan et al. (1985) conclude that because only four out of 105 studies report significant results, there are no sex differences in Porteus's data. Hisock (1986) and Halpern (1986), on the other hand, point out that the probability of the mean for one sex exceeding that for the other in 99 of 105 instances is 10^{-7} and contend that the Porteus data supports the idea of sex differences. Hisock's and Halpern's conceptualization of the results corresponds with the ideas of proponents of meta-analysis (Hunter, Schmidt, & Jackson, 1982). In short, for a given mean difference and standard deviation, the probability that a statistical test will result in a significant finding is a function of sample size. Large samples may yield trivial differences that are "significant," while small sample studies may fail to detect fairly large mean differences. The existence or magnitude of an effect is, therefore, obscured when "significant" vs. "insignificant" results are simply tallied. Although there are no meta-analytic data regarding Porteus's findings, it appears that there is a consistent sex difference favoring males across studies. The size, or magnitude, and meaningfulness of the effect are the more relevant issues.

Existence and Magnitude of Sex Differences

Because a vast body of literature regarding sex differences in spatial ability has accumulated over the years, reviewers are faced with the problem of identifying and selecting studies to report. Although the review process is intended to ensure the quality of published results, journals may be biased toward publishing positive findings of sex differences (as suggested by Caplan et al., 1985). Other studies must, therefore, be identified to ensure the quality of the review. As noted by Burnett (1986), normative data on psychometric instruments warrants attention because publication of the results does not depend on positive findings of sex differences and because it typically reflects large sample sizes representative of various populations. For instance, in 1977 the American College Testing Program (ACT) reported national norms based on a total sample greater than 15,000. Means of stanine scores on the ACT Space Relations subtest, a block counting test, were 5.39 (SD = 1.95) for males and

4.47 (SD = 1.79) for females, approximately half a standard deviation difference. ACT Mechanical Reasoning yielded over a standard deviation difference in means favoring males (Males $\bar{X}$ = 5.69, SD = 1.86; Females $\bar{X}$ = 3.79, SD = 1.61). In contrast, females outperformed males in Language Usage and Clerical Skills, and both genders performed similarly on Reading Skills and Numerical Skills. Differences between means based on alternate norming samples [part-time students (796 males and 512 females) and blacks (721 males and 745 females)] were virtually identical to those based on the national sample.

Similarly, Bennett, Seashore, and Wesman (1974) reported large sample means for the Differential Aptitude Test (DAT) subtests. With regard to the Spatial Relations subtest, a paper-folding test, the mean for 8th grade boys was 24.1 (SD = 9.9, N = 7000+) compared to 23.0 (SD = 9.0, N = 6900+) for 8th grade girls. While the difference in means was only about one-ninth of a standard deviation at the 8th grade, it increased steadily in the high school years, and at the 12th grade was almost one-third of a standard deviation (i.e., males $\bar{X}$ = 34.3, SD = 13.0, N = 5000+; females $\bar{X}$ = 30.9, SD = 11.9, N = 5350+). For DAT Mechanical Reasoning the difference is greater than one standard deviation favoring males in the 12th grade (i.e., males $\bar{X}$ = 48.3, SD = 9.5; females $\bar{X}$ = 37.4, SD = 8.5; N > 4000). As with the ACT subtests, females obtained higher scores than males in Language Usage and Clerical Speed and Accuracy.

Importantly, these examples illustrate that the magnitude of the sex difference on spatial ability tests may vary with the type of test employed. Only recently have researchers suggested that some types of tests produce larger sex differences than others. Halpern (1986) and McGee (1979) assert that large sex differences favoring males are consistently found in three measures: 1) mental rotation tests, 2) rod-and-frame tests, and 3) Piaget's water level test. The remainder of this review will focus primarily on mental rotation ability because it is of particular importance to the current study.

Mental Rotation

Mental rotation tests require correctly identifying or "matching" an object with a replica that has been rotated in two or three dimensions. Although large sex differences have been reported for both two- and three-dimensional rotation ability, three-dimensional rotation tasks have yielded the larger effect. The results of several studies employing the Mental Rotations Test (Vandenberg & Kuse, 1978), a three-dimensional mental rotation test based on Shepard and Metzler's (1971) cubes, are discussed below.

In 1975, Yen administered a spatial battery including the Mental Rotations Test to over 1200 high school boys and 1200 high school girls. Mean scores are provided in Table IV. Scores for both boys and girls tended to increase through high school, but less so for girls. The effect size ranges from .82 to 1.00, indicating that there was almost a full standard deviation difference between males' and females' scores.

McGee's (1977) doctoral dissertation was a familial study of spatial abilities in which 801 individuals in 200 families participated. As shown in Table IV, within each generation, males

TABLE IV

SEX DIFFERENCES IN MENTAL ROTATION TEST SCORES*

		MALES			FEMALES			
AUTHOR	SAMPLE	MEAN	SD	N	MEAN	SD	N	d**
Yen, 1974	9th Grade	12.6	5.1	414	8.4	4.1	402	.91
	10th Grade	13.0	5.3	350	9.1	4.2	341	.81
	11th Grade	13.5	5.2	256	9.0	4.3	254	.94
	12th Grade	14.1	5.0	199	9.4	4.4	213	1.00
McGee, 1976	Parents	17.0	9.3	155	10.0	6.9	168	.86
	Offspring	23.0	8.8	241	16.0	8.3	237	.82
Wilson & Vandenberg, 1978	Hawaiian Families	19.7	11.7	2502	10.1	10.5	2576	.87
Vandenberg & Kuse, 1978	Undergrads	9.9	4.4	115	7.0	4.0	197	.71
		9.1	4.2	115	6.2	3.5	197	.78
McGee, 1978	Undergrads	4.2	2.7	173	3.4	2.5	174	.31
Drauden, 1980	Undergrads	24.3	8.1	92	18.5	5.6	114	.85
Freedman & Rovengno, 1981	Undergrads	8.4	3.9	40	4.5	2.3	40	1.21
Sanders, Soares, & D'Aquila, 1982	Undergrads	23.4	9.8	359	15.2	8.6	672	.91

* Differences in mean test scores across studies are due to different methods of scoring this test. Different scoring methods do not effect the standardized difference between males' and females' scores.

** The formula for the standardized difference is

$$d = \frac{\bar{X}_m - \bar{X}_F}{S_p} \quad \text{where,} \quad S_p = \sqrt{\frac{(N_m - 1)s^2_m + (N_F - 1)s^2_f}{(N_m - 1) + (N_F - 1)}} .$$

outperformed females by over eight-tenths of a standard deviation on the Mental Rotations Test. Additionally, sons obtained higher scores than fathers, by about two-thirds of a standard deviation, and daughters performed about eight-tenths of a standard deviation better than mothers. McGee (see also Bouchard & McGee, 1977) hypothesized that sex differences might be due to an X-linked recessive gene, but their speculations were not supported by patterns of correlations obtained. McGee also investigated the idea that sex differences might be due to differential response patterns across items for males and females. He, therefore, tabulated the percent passing on each item separately for males and females and rank-ordered the items in terms of their difficulty for each sex. The correlation between the rank-ordered items was .96, indicating that items that are difficult for females are also difficult for males, and on easier items percent passing increases for both sexes.

Wilson and Vandenberg (1978) solicited families living in Hawaii for a large scale study of familial resemblance in cognitive abilities. A battery of 15 cognitive tests was administered to 2,502 males and 2,576 females between the ages of 14 and 60. On the Mental Rotations test (3-D) the mean for males was 19.7 (SD = 11.7) compared to 10.1 (SD = 10.5) for females, about .87 of a standard deviation difference. The difference in means on the Card Rotations Test was .47 of a standard deviation (Males $\bar{X}$ = 105.5, SD = 37.3; Females $\bar{X}$ = 87.5, SD = 37.9). Means were also reported for parents and offspring separately. Their findings replicated McGee's (1977) data; sons ($\bar{X}$ = 24.3) outperformed fathers ($\bar{X}$ = 16.5) and daughters ($\bar{X}$ = 15.1) scored higher than the mothers ($\overline{X}$ = 6.4). Also consistent with McGee's data, the genetic hypothesis investigated by these researchers was largely disconfirmed by the pattern of parent, offspring, and sibling correlations obtained.

Vandenberg and Kuse (1978) administered four spatial tests to 197 female and 115 male undergraduate students at the University of Colorado. The Mental Rotations Test was split in half, and both halves were administered to all the subjects. The results appear in Table IV. On part 1 of the Mental Rotations males outperformed females by .71 of a standard deviation, and similarly, on part 2 there was three-quarters of a standard deviation between male and female means.

Sanders, Soares, and D'Aquila (1982) administered the Mental Rotations Test and the ETS Card Rotations Test to 672 female and 359 male undergraduate students at the University of Connecticut. Males obtained higher scores than females on both tests (see Table IV.). The size of the effect on the Mental Rotations Test was again well over half a standard deviation (i.e., .91 of a standard deviation difference between means), whereas the difference between means on the Card Rotations Test was .29 of a standard deviation. Similarly, Freedman and Rovengno (1981) administered the Mental Rotations test to 80 undergraduates and obtained over a standard deviation difference favoring males.

McGee (1978b) and Drauden (1980) both conducted training studies employing the Mental Rotations Test. Means and standard deviations of males' and females' scores before training are presented in Table IV. In Drauden's study, males performed about eight-tenths of a standard deviation better than females, and McGee found a difference favoring males by about one-third of a standard deviation.

Further evidence regarding sex differences in mental rotation ability comes from the chronometric research arena. Since Shepard and Metzler (1971) reported that subjects' performance on mental rotation tasks can be described in terms of a slope and intercept of latency measures and that the slope increases linearly with the angular departure of the stimulus from a standard position, researchers have attempted to isolate the locus of sex differences in mental rotation. Accumulating research in this realm suggests that adult females rotate objects slower than do adult males (Kail, Carter, and Pellegrino, 1979; Kail, Stevenson, & Black, 1984; Alderton & Pellegrino, 1985; Regian & Pellegrino, 1984; Tapley & Bryden, 1977).

In sum, a sex difference over one-half a standard deviation in magnitude (and approaching a full standard deviation) has been documented for the Mental Rotations test. A recent meta-analysis of spatial ability studies corroborates this finding (Linn & Petersen, 1985).

Meta-Analytic Results

Linn and Petersen (1985) reviewed spatial abilities studies reported between 1974 and 1982 to assess the magnitude of sex differences, to identify spatial ability measures producing sex differences, and to discern the onset (during the lifespan) of sex differences in spatial ability. They identified 200 effect sizes, of which 172 contained enough information to be included in their meta-analysis. They categorized the effect sizes into three groups according to the type of spatial ability measure: spatial perception, mental rotation, and spatial visualization. The spatial perception category included tasks in which the subject must "determine spatial relationships with respect to the orientation of their own bodies, in spite of distracting information" (p. 1482). Examples include the Rod and Frame Test and Piaget's water level task. The mental rotation category included two- and three-dimensional rotation tasks such as Mental Rotations (Vandenberg & Kuse, 1978), and Cards and Flags (French et al., 1963). The spatial visualization category contained tests that Petersen and Linn believed to require complicated, multistep manipulations; these included tests usually thought to load on a Flexibility of Closure factor, such as the Embedded Figures Test and Hidden Patterns, as well as tests such as Paper Folding, Paper Form Board, and Surface Development--tasks thought to measure a Visualization factor (as defined by Lohman, 1979).

Following the procedures suggested by Hedges (1982), they computed mean effect sizes (standardized mean differences) across studies according to three categories of tests and assessed the homogeneity of the effect sizes to discern whether other sources of variance (i.e., age or test specific) should be considered. They found that the effect sizes for the spatial perception and rotation categories were not homogeneous and re-examined the data according to the age of the participants and the test employed in the study. With regard to rotation, age had little or no influence on effect size. The mental rotation test employed did, however, affect the magnitude of the effect size. More specifically, Mental Rotations, which involves rotation in the depth plane produced substantially larger sex differences (i.e., an average effect size of .94 favoring males) than did two-dimensional rotation tasks (i.e., average effect size = .26 favoring males), regardless of age. In contrast, age contributed to the lack of homogeneity in the effect sizes for spatial perception; the type of test did not. In particular, the effect size was slightly over one-third standard deviation in studies of children 18 or under and was about two-thirds standard deviation in studies of individuals over 18. Petersen and Linn suggested the age effect might be due to a cohort effect, which does not appear very likely given that no age effect occurred for test categories other than spatial perception, or sampling biases. No clear explanation emerged. With regard to spatial visualization, the average effect size (.13 favoring males) was homogeneous across age groups and different tests.

Summary

In sum, Petersen and Linn's data corroborate the notions of McGee (1979) and Halpern (1986) regarding effect sizes for particular tests. More specifically, the mean effect size for spatial perception and two-dimensional rotation tests were approximately one-third of a standard deviation, and effect sizes for three-dimensional rotation tasks were large, averaging .94. In contrast, tests in the visualization category (e.g., Embedded Figures) produced a relatively small effect size (.13) favoring males.

Consequences of Sex Effects

Caplan et al. (1985) consider one-half of a standard deviation difference in spatial scores to be trivial. As illustrated by Benbow and Stanley (1982), however, a mean difference of only a third of a standard deviation has substantial effects on the upper end of the distribution. The importance of a sex difference, thus, lies in how spatial abilities might be related to other aptitudes and in how spatial tests are used.

One example of the use of spatial ability tests comes from the employment setting. The Uniform Guidelines for Employee Selection Procedures (1978) outline criteria for legal use of selection tests for employment. Simply stated, if scores on a test are correlated with job performance, they can be used to select people for jobs. Several studies have found scores on spatial ability tests to be related to job performance in specific occupations. For instance, spatial ability scores have yielded correlations with performance in skilled trades jobs (Ghiselli, 1966, 1973) and a number of Military Occupational Specialties (MOS) in the Army (McHenry, Hough, Toquam, Hanson, & Ashworth, 1987; Wise, Campbell, & Peterson, 1987). Spatial ability tests are, therefore, used in industrial settings to select persons for job openings. Likewise, some graduate and undergraduate programs for specialty areas consider such scores in admitting students. Because a mean difference of the size found for mental rotation ability can have dramatic effects on the upper end of the distribution, more men will be selected or admitted than women. Burnett (1986) illustrated the potential effects of a half of a standard deviation favoring males. If

"500 men and 500 women apply for approximately 213 openings in architecture school,..., and a spatial ability test in which men and women differ by half a standard deviation is used as part of the selection battery, then approximately 142 (28.43%) men but only 71 women (14.23%) would be admitted -- twice as many men as women" (p. 1013). Fewer openings, higher cut scores, or a greater sex difference would magnify the differences. In this context, a difference of even half a standard deviation is not trivial.

What Hypotheses Have Been Posed to Explain the Nature of Sex Differences?

Hypotheses proposed to explain male superiority relate spatial ability to a sex-linked gene (Bock & Kolakowski, 1973; Hartlage, 1970; Stafford, 1961; Yen, 1975), hormonal state (Broverman, Klaiber, Kobayashi, & Vogel, 1968), hemispheric lateralization (Buffery & Gray, 1972; Burnett, Lane, & Dratt, 1982; Levy, 1969), differences in cognitive processing (Kail, Carter, & Pellegrino, 1979), or differential experience (Sherman, 1967). This section reviews research related to biological hypotheses briefly and discusses data related to experiential hypotheses in detail because they are most relevant to the current study.

Biological Explanations of Sex Differences

Genetic Hypothesis

A number of researchers have proposed that spatial ability may be enhanced by an X-linked recessive gene. To test this hypothesis, investigators obtain parent-offspring correlations and compare them to those expected by the Mendelian model. The pattern expected in such studies is that sister-sister correlations be the highest because sisters receive the one X chromosome from their father and are expected to receive duplicate X chromosomes from their mother in 50 percent of the cases. Mother-son and father-daughter correlations are expected to be next highest because, in each case, the offspring will receive one X chromosome from the parent. Because fathers transmit no X chromosomes to their sons, father-son correlations are expected to be the lowest, essentially zero.

Initially, research (Bock & Kolakowski, 1973; Hartlage, 1970; Stafford, 1961; Yen, 1975) supported the model. In the latter 70's, though, four large sample studies (Bouchard & McGee, 1977; DeFries, Ashton, Johnson, Kuse, McClearn, Mi, Rashad, Vandenberg, & Wilson, 1976; Park, Johnson, DeFries, McClearn, Mi, Rashad, Vandenberg, & Wilson, 1978; Williams, 1975) did not yield the expected correlations and reviews (Boles, 1980; McGee, 1979) concluded that there is little or no support for the X-linkage of spatial ability. More recently, however, Thomas (1983) questioned the assumptions underlying the pattern of correlations expected by prior researchers. To date, it is unclear whether the X-linked recessive gene hypothesis has been tested adequately.

Hormonal State

In their frequently cited review of abilities research, Maccoby and Jacklin (1974) suggested that reliable sex differences in spatial ability are found for adolescents and adults but not for younger populations. It has, thus, been suggested that sex differences in spatial ability are related to differences in androgen levels (e.g., Wittig & Petersen, 1979). There appear to be two major obstacles to identifying such a linkage. First, a number of more recent studies suggest that male-female differences in spatial ability may emerge prior to adolescence (cf., Newcombe, Bandura, & Taylor, 1983). Indeed, the meta-analytic results reported by Linn and Petersen (1985) suggested that "when sex differences are found, they can be detected across the life span" (p. 1479). Second, methods for assessing androgen levels have relied primarily on imprecise indices of androgynous physical appearance and groups with extremely low androgen levels representing extremes in the normal population range are often studied. Studies relating androgen level to spatial ability have, thus, been problematic. More recently, one study (Shute, Pellegrino, Hubert, & Reynolds, 1983) using radioimmunoassay to determine androgen levels found curvilinear functions significantly relating androgen levels to spatial test scores. In effect, persons with very high or very low androgen levels did not perform as well as those with moderate levels. Research in this area, using newer techniques, has potential for shedding some light on individual differences in spatial ability, if indeed the emergence of sex differences at puberty is a supported finding.

Hemispheric Lateralization

It has been suggested that greater right hemisphere specialization of males enhances their performance on tests of spatial ability (e.g., Ehrlichman, 1972; Lansdell, 1962). Indeed, numerous studies do provide evidence that the right hemisphere is specialized for spatial processing (see McGee, 1979; Harris, 1981; McGlone, 1980 for reviews of this literature). There is, however, less agreement that males are more lateralized than females (e.g., Buffery & Gray, 1972; McGlone, 1980; Bryden, 1979). Newcombe (1982) adds that even if males are more lateralized than females three lines of research are required to demonstrate that sex differences in cerebral organization are causally related to sex differences in spatial ability. First, independent of sex, it must be demonstrated that lateralization is related to spatial ability. Second, the cause-effect relationship between spatial ability and lateralization must be explicated. That is, a high correlation between lateralization and spatial ability might suggest that people who have high spatial ability use strategies that engage the right hemisphere on tests of lateralization (i.e., spatial ability might be "causing" lateralization results) or it may indicate that lateralization has a causal effect on spatial ability; a correlation alone does not answer the cause-effect question. Third, sex-related patterns of lateralization should be shown to developmentally precede sex differences in spatial ability. Her review of existing research indicated that further investigation of each of these issues is needed and that no conclusions regarding the relationship between cerebral organization and spatial ability can be drawn.

Summary

Although some results appear promising, biological theories have not met with considerable success in explaining sex differences in spatial ability. Caplan et al. (1985) correctly asserted that there is no strong evidence supporting a biological theory of sex differences in spatial ability; this fact, however, does not negate the existence of a sex difference (as suggested by Caplan et al.). Indeed, it is probable that no one theory can explain sex differences, and a number of factors, biological and experiential, contribute to spatial performance. Additionally, different hypotheses can be "true" at the same time. For instance, genetic factors might lead individuals to adopt different strategies for accomplishing spatial tasks. Likewise, biological factors might "cause" persons to seek specific types of experiences that enhance natural abilities.

Experiential and Process-Oriented Explanations

The Rotation Process

As mentioned previously, the chronometric studies conducted by Shepard and his colleagues (Cooper & Shepard, 1975; Shepard & Feng, 1972; Shepard & Metzler, 1971) have enhanced our understanding of the mental processes involved in mental rotation. In sum, their research suggests that mental rotation is a Gestalt-like process wherein cognitive process have a one-to-one correspondence with the external rotation of the object. Four processes are involved in the solution of a mental rotation item. First, the individual encodes the identity and orientation of a stimulus in working memory. Then, the subject rotates the mental representation of the comparison stimulus to the orientation of the standard. Third, the person compares the rotated representation of the stimulus with the standard stimulus and responds if the stimuli are the same. If the stimuli are different, additional time is needed to respond. A number of variations of the Shepard et al. algorithm have been proposed (e.g., Carter, Pazak, & Kail, 1983). Moreover, the debate among researchers is whether mental rotation is subject to analytic processing strategies or is analogous to physical rotation. Research to date provides strong support for the Gestalt-like analogue process proposal (Linn & Petersen, 1985).

Accumulating chronometric research suggests that the process of mental rotation does not differ for the sexes, nor does the time required to encode stimuli. It does appear, however, that females rotate objects at a slower rate than do males (Kail, Carter, & Pellegrino, 1979; Kail, Stevenson, & Black, 1984; Alderton & Pellegrino, 1985; Regian & Pellegrino, 1984; Tapley & Bryden, 1977). More specifically, the locus of the sex difference appears to lie in the slope of the reaction time function (i.e., a measure of the rate of rotation). This finding does not eliminate the possibility that males and females might use somewhat different mental rotation strategies. For instance, Kail et al. (1979) suggest that slower subjects may rotate only parts or features of an object at a time and, thus, have to repeat the process until a decision can be made; whereas, those who rotate the entire object can identify a match or mismatch in one rotation.

Interestingly, no hypotheses specifically address the increased magnitude of the sex difference in three-dimensional rotation over that of two-dimensional tasks. In fact, the work of Shepard et al. suggests that the same mental processes are involved in both two- and three-dimensional rotation, the only differences being that the stimulus is three-dimensional and that rotation occurs in the depth plane. Why does mental rotation in the depth plane increase the differences between the sexes? Perhaps differences in the speed of rotation (i.e., females, in general, being slower) are simply accentuated when an object must be rotated in the depth plane. Alternatively, perhaps the image of a three-dimensional object is more difficult to retain than that of a two-dimensional object. Although the time to encode stimuli (i.e., build a mental image) does not appear to differentiate the sexes, one study (Kail et al., 1984) does suggest that individuals who rotate two-dimensional objects slowly may have encoded the stimulus poorly or may experience more rapid stimulus degradation or both. In this context, it is possible that females encode a three-dimensional stimulus poorly or that the mental image of a three-dimensional object deteriorates quickly for females or both. This hypothesis appears to be a plausible explanation of the increased magnitude of sex difference in three-dimensional rotation, because three-dimensional rotation items are typically box drawings of objects, with few cues for depth perception. In turn, few depth perception cues may result in a poorer or rapidly deteriorating mental image of the object.

Experience

Spearman (1927) was actually one of the first researchers to suggest that experiential factors underlie sex differences in spatial (or "practical") ability. In 1967, Julia Sherman expanded upon this proposal. Her differential experience hypothesis claims that males tend to have more experiences that develop spatial ability than do females. If experiential factors are important in fostering high spatial abilities, training and practice received should, over time, result in improved performance on tests of spatial abilities, at least until an asymptotic or ceiling level is achieved. Indeed, a number of researchers have reported increases in subjects' performance on spatial ability tests at retest without a training intervention (Bennett, Seashore, & Wesman, 1974; Dunnette, Corpe, & Toquam, 1987; Kepner & Neimark, 1984; Lohman, 1988), a finding that supports the idea that performance on spatial ability measures is influenced by a learning or experiential component.

In most of the earlier training studies conducted in the 1940's and 50's, researchers attempted to train spatial ability indirectly, through ordinary course offerings in geometry, drafting and blueprint reading, or engineering drawing. Most of these attempts were unsuccessful (Brown, 1954; Churchill, Curtis, Coombs, & Hassell, 1942; Faubian, Cleveland, & Hassell, 1942; Ranucci, 1952), although Blade and Watson (1955) did report a significant increase in students' spatial visualization scores during an engineering course.

Similarly, attempts to train spatial ability indirectly, through tasks dissimilar to the criterion (pre- and post-training) measure have not met with success. For example, Levine, Brahlek, Eisner, and Fleishman (1979) found significant improvement of trained groups over control groups on criterion measures highly related to training, but not on those less directly related to the intervention. Levine, Schulman, Brahlek, and Fleishman (1980) designed training tasks to enhance visualization ability, while at the same time ensuring the training tasks were unlike the criterion measure (Flanagan's Assembly test). Criterion measure scores of trained groups did not improve over that of untrained groups.

More recent efforts in which training was more directly related to the criterion measure have yielded positive results. Brinkmann (1966), for example, provided extensive training in the behaviors thought to underlie spatial visualization (i.e., discrimination, recognition, organization, and orientation) and found significant improvement on a spatial relations criterion test administered before and after training for the trained group but not for an untrained control group. Stringer (1975) attempted to enhance spatial ability using various drawing training procedures and found that trained groups did better than an untrained control group on a test of spatial relations, but only when there was direct similarity between the training and testing materials. Similarly, Embretson (1988) trained subjects using physical analogues of the DAT Space test items, mental folding items; she found a significant increase in trained group scores over that of untrained groups. Lohman (1988) also found that practice solving mental rotation problems like Shepard and Metzler's (1971) led to significant increases in subjects' scores on Thurstone and

Thurstone's (1941) Cards and Figures and ETS Paper Folding and Form Board tests.

Although little research has addressed the extent to which training on one measure generalizes to performance on another measure of spatial ability, the results of training studies described above tend to suggest that training is measure-specific. Transfer of training has been noted in two studies (Lohman, 1988; Sevy, 1983); however, in general, training directly related to the criterion measure has resulted in performance improvement (e.g., Connor, Serbin, & Schackman, 1977), while training indirectly related to the criteria has not (e.g., Levine et al., 1979, 1980). Such results are somewhat counter to the expectations one might draw from factor-analytic research on spatial abilities. Factor-analytic research suggests that spatial abilities are correlated measures of Gv, not highly specific, unique abilities. If, indeed, the skill or ability is trained, training on one measure should result in some improvement on another measure. If not, it may be that performance improvement is simply a consequence of greater familiarity with instructions and the task at hand.

One important, but often overlooked, concept with regard to generalizability of training effects is that transfer is probably a matter of degree. For example, Ferguson (1954, 1956) and Sullivan (1964) proposed that positive transfer will occur if training is closely associated with the measure, and increasing the difference between the material used in the training and that of the test will increase the difficulty of transfer. Moreover, generalizability or transfer is best described as a continuum. One goal of the current study is to examine the degree of transfer systematically by designing criteria with differing degrees of training-relatedness and thus with differing expectations for transfer.

Differential Experience

Although significant training and transfer effects would support the idea that experiential factors influence performance on spatial tasks, Sherman's (1967) differential experience hypothesis is not adequately tested unless the improvement of females relative to males is examined. That is, if males have had more experience with activities that develop spatial ability, they should be closer to the asymptotic level of ability than females (Sherman, 1967). Females, being lower on the learning curve than males, should respond more favorably to training. The few studies that have examined differential response to training by females have produced inconsistent results. Several psychometric studies (Blatter, 1983; Connor, Serbin, & Schackman, 1977; Connor, Schackman, & Serbin, 1978; Goldstein & Chance, 1965; Johnson, Flinn, & Tyer, 1979; Vandenberg, 1975) have reported greater improvement by females than males after training. Similarly, in two chronometric studies of mental rotation females obtained parity with males after practice (Alderton & Pellegrino, 1984; Regian & Pellegrino, 1984) or showed substantial improvement compared to males (Sevy, 1983). But, Drauden (1980), McGee (1978b), and Teegarden (1942) found no differential response to training.

The failure to find consistent support for the differential experience hypothesis could stem from variations in the training techniques employed. As McGee (1979) and Sherman (1967) have noted, there are many unknowns involved in assuming what activities do facilitate higher spatial ability. Indeed, if an activity is highly sex-typed, it is likely that a potent treatment intervention would be required to overcome years of deficits in experience. Training designed to enhance relevant skills may not always be realized. For example, Teegarden (1942) simply increased the time limit on the test, and thus, did not administer training per se. McGee (1978b) "trained" subjects to use a visualization strategy for mental rotation by having them attend a one-hour lecture on how to visualize and rotate an object. It seems unlikely that a one-hour classroom lecture would provide subjects with the experiences needed to overcome deficits in past learning, especially if mental rotation is a highly sex-typed activity.

Drauden's (1980) training was more extensive. She prepared models of items on tests (Card Rotations and Bingham's Cubes), and then during the training sessions administered tests and asked subjects to check their work using the models. Although her results tended to disconfirm the differential experience hypothesis, she was unable to draw strong conclusions because males performed near a ceiling on her post-training measures.

Training studies in which females' performance was increased more than males were designed to provide subjects' with process-related instructions as well as practice and feedback experiences. For instance, the training administered by Connor and her colleagues (1977,

1978) was designed to illustrate the process of extracting an embedded figure. Children were shown overlays of complex figures. As the overlays were successively removed, complex figures were simplified, and the target figure became apparent. Along the same lines, Vandenberg (1975) instructed children to use blocks to build models of the items on the Mental Rotations test to provide children with concrete experience visualizing three-dimensional objects from various angles.

In sum, training that provides practice with skills that are directly linked to the criterion measure (pre- and post-training measure) is likely to result in improved performance on the criterion (e.g., Connor et al., 1977, 1978). Training interventions that rely on lectures or ordinary course offerings are likely to lack the potency needed to facilitate skill development. Previous chronometric research on mental rotation provides some important clues about the rotation process and, in turn, about specific abilities that mental rotation training should address. That is, the studies conducted by Shepard and his colleagues (Cooper & Shepard, 1975; Shepard & Feng, 1972; Shepard & Metzler, 1971) suggest that mental rotation involves a Gestalt-like process of building a mental picture of an object and rotating this mental image. Process-oriented training in mental rotation would, therefore, need to approach mental rotation as a process analogous to physical rotation. Tasks designed to provide experience visualizing a three-dimensional object and rotating it, mentally and physically, would be necessary components of process-oriented mental rotation training.

Hypotheses for Current Study

The study described in subsequent chapters of this paper examines the effect of training on subjects' performance on mental rotation tests, the performance of females relative to males under different training conditions, and the generalizability of training on one test or measure to performance on a different test. It involves three major phases--a pre-training testing session, training sessions, and a post-training testing session. Independent measures are sex (i.e., male and female) and experimental group (i.e., control, feedback training, and mental rotation training). Dependent measures are scores on three mental rotation tests--one test directly related to the training procedures used, one test with items similar to those employed as training tools, and one test unrelated to the training methods.

This study was designed to evaluate several hypotheses suggested by previous literature on sex differences in spatial ability; these hypotheses are described below.

Hypothesis 1: Sex Differences

Before training, males will outperform females on mental rotation tests.

Hypothesis 2: General Training Effects

Trained groups will show greater improvement than untrained groups, post-training. A significant interaction between experimental group (i.e., control, feedback training, and mental rotation training) and the trials factor illustrating greater improvement of trained groups over that of the control group would support this hypothesis.

Hypothesis 3: Specific Training Effects

Training designed to provide direct experience rotating three-dimensional objects will result in greater performance improvement than training that relies only on performance feedback. More specifically, greater improvement should be apparent for subjects receiving mental rotation training than for those in the feedback training group.

Hypothesis 4: Differential Learning

If males are functioning close to the asymptote of their spatial ability whereas females are less experienced, females should respond more favorably to practice than males. Scores of females should, therefore, show greater performance improvement pre- to post-training than those of males. A significant interaction between sex and trials (i.e., pre- to post-training) would indicate that females respond more favorably to practice than do males across testing sessions, regardless of experimental group (i.e., control, feedback training, or mental rotation training) and would thus support the differential experience hypothesis. However, previous research suggests that females do not improve relative to males at retest, without training. A potent treatment intervention is more likely to facilitate learning and improvement of females relative to males. A significant interaction between sex, experimental group, and trials indicating that females in the mental rotation training group improve relative to males while those in the control group and feedback training group do not would support this hypothesis.

Hypothesis 5: Transfer of Training

Training on one measure of mental rotation should generalize to other, highly similar, measures of mental rotation. If so, trained groups should outperform the control group on dependent measures that are not directly related to the training program as well as those that are training related. Significant improvement of trained groups over that of the control on dependent measures unrelated to training would support this hypothesis. If training does not transfer beyond similar types of items, a training effect will be observed only on the dependent measure similar to that used in training.

CHAPTER III

METHODS AND PROCEDURES

This study examined the effect of training on subjects' performance on mental rotation tests, the performance of females relative to males under different training conditions, and the generalizability of training on one test or measure to performance on a different test. It involved three major phases--a pre-training testing session, training sessions, and a post-training testing session. The subjects, male and female Oklahoma State University undergraduates, were divided among three experimental groups--control, feedback training, and mental rotation training.

This chapter provides details about the methods and materials used. It is divided into five major parts. The first part provides an overview of the experimental framework, and the second part describes the subjects. The third part explains the tests used to measure spatial ability. The fourth part provides information about the training materials and procedures, and the fifth part reviews the experimental framework and procedures.

Experimental Framework

As mentioned, this study involved three major phases--a pre-training testing session, training sessions, and a post-training testing session--and three experimental groups--control, feedback training, and mental rotation training. Figure 1 illustrates the experimental design. During pre-training testing, three mental rotation tests were administered to all subjects, and post-training testing involved administering parallel forms of the three tests given pre-training. Male and female subjects were recruited for each of the three experimental groups. Moreover, the study involved two independent variables, sex and experimental method; dependent measures were mental rotation test scores. Sex and experimental method were between-subjects factors, and equal numbers of subjects were sought to fill each sex by method cell.

The data collection framework was also designed to control for two factors that could potentially confound or contaminate the results: order of test administration and experimenter bias. The order of test administration is important because learning from taking one test (or perhaps relaxing thereafter) could result in enhanced performance on subsequent tests. One way to control for order effects is to use a Latin squares design to balance the sequence with which tests are administered across testing sessions. If three tests are to be administered, six testing sessions are needed to ensure proper balance (see Table V). Six pre-training and six post-training sessions (with the three tests balanced according to the Latin squares design) were, therefore, included in the design of the study.

Subtle inflections or variance in instructions given by the experimenter could also impact subjects' test-taking performance. If, for example, control group participants were given somewhat different motivational cues than those in a training group, enhanced performance

Phases

Experimental Groups	Pre-Training Testing Session	Training Sessions 1	Training Sessions 2	Post-Training Testing Session
Control	$S_{1,1}$ $S_{1,2}$ . . . $S_{1,N}$			
Feedback	$S_{2,1}$ $S_{2,2}$ . . . $S_{2,N}$			
Mental Rotation	$S_{3,1}$ $S_{3,2}$ . . . $S_{3,N}$			

Figure 1. Experimental Design

TABLE V

LATIN SQUARES DESIGN FOR ADMINISTERING THREE TESTS (A, B, AND C)

Testing Session	Order of Administration		
1	A	B	C
2	B	C	A
3	C	A	B
4	A	C	B
5	B	A	C
6	C	B	A

by training group participants could be attributable, in part, to the motivational cues given by the experimenter during test taking. Two precautions were undertaken to minimize experimenter effects. First, testing procedures and instructions were standardized. Second, and perhaps more important, testing sessions were set up such that each pre- and post-training testing session included individuals representing each experimental group (i.e., control, feedback training, mental rotation training). This design is depicted in Figure 2. The pre- and post-training testing sessions are composed of representatives from all three experimental groups, and the potential for an experimenter biasing effect is diminished.

In sum, 18 groups of subjects (with equal numbers of males and females) were needed to ensure adequate control for order and experimenter effects (i.e., six groups for testing purposes and three groups within each of the larger six--one for each experimental treatment). Subject recruitment materials and room accommodations were designed with this in mind.

Subjects

One hundred and thirty-five Oklahoma State University undergraduate psychology students between the ages 18 and 49 participated in the study. Sixty-five subjects were females, and seventy subjects were males. Subjects received one to four extra credit points for their participation in the study. (Each subject was asked to attend four sessions and was given extra credit points corresponding to the number of sessions completed.)

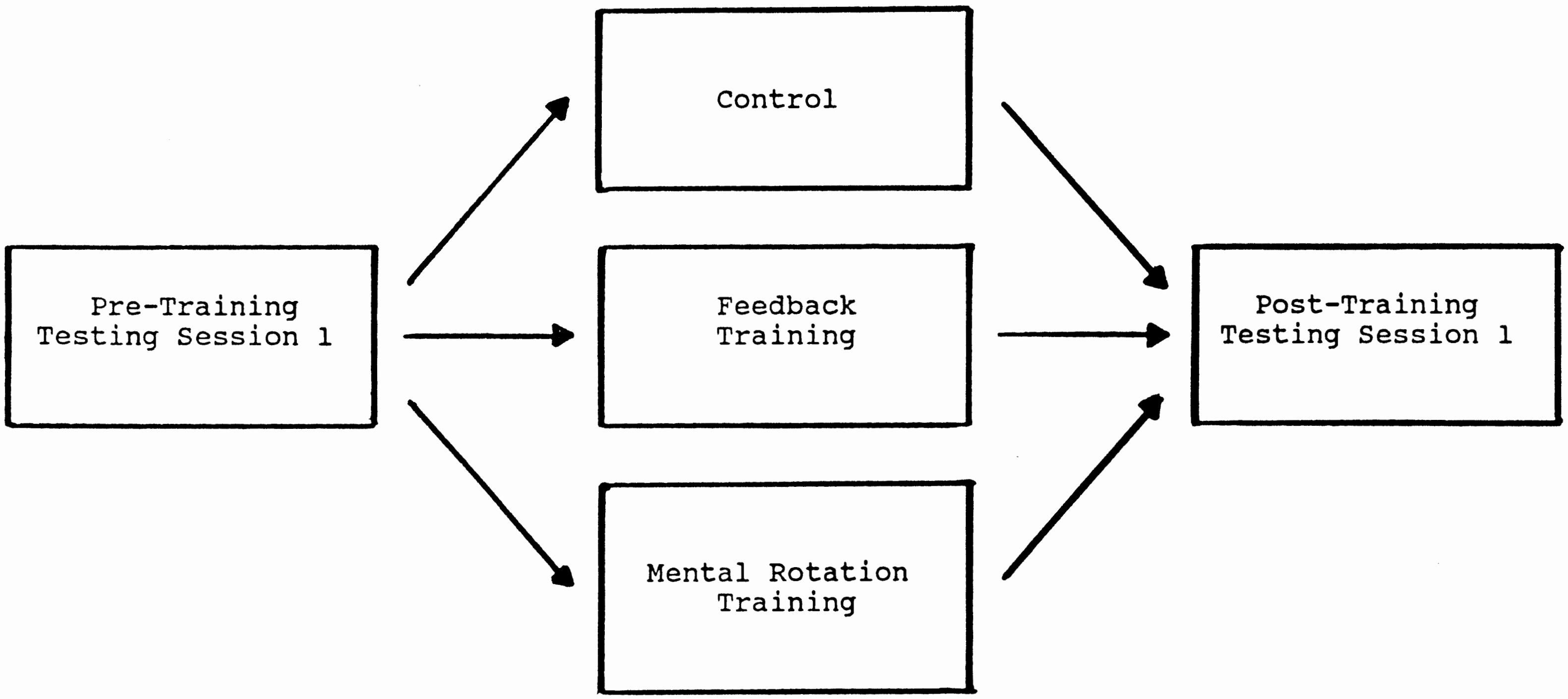


Figure 2. Design of Pre- and Post-Training Testing Session

Each subject "signed-up" for one of 18 blocks of sessions; each block included four one-hour sessions (i.e., a one-hour pre-training session, two one-hour training sessions, and another one-hour post-training session) conducted over the course of one week. Sign-up for each of the eighteen blocks of sessions was limited to eight participants (i.e., four males and four females). The eighteen blocks of sessions were assigned randomly to experimental groups prior to subject recruitment (with the constraint that six blocks of sessions be assigned to each experimental group and that three blocks of sessions be assigned to each testing session).

As mentioned, a total of 135 subjects participated in the study. Not all subjects attended all four sessions (i.e., one pre-training, two training, and one post-training session). As shown in Table VI, 134 subjects attended pre-training sessions; one subject, the 135th, did not attend pre-training or training sessions but did take tests given post-training. Sixty-nine were males, and 65 were females. Of these 134, 109 participated in all phases of the study (see Table VII).

Data from three subjects, who were over the age of 30, were excluded from several analyses (described in Chapter IV) to avoid potentially confounding effects of age. The edited sample size is, therefore, 106 (i.e., 109 - 3). The numbers of males and females in the edited sample are provided in Table VIII. As shown, 53 participants were male, and 53 were female. Their ages ranged from 18 to 29 years with a median of 19 years and an average of 19.8 years (see Table IX).

TABLE VI

NUMBERS OF PARTICIPANTS IN THE PRE-TRAINING SESSIONS

	Number of Participants		
Experimental Group	Males	Females	Total
Control	19	20	39
Feedback	30	21	51
Mental Rotation	20	24	44
Total	69	65	134

TABLE VII

NUMBERS OF PARTICIPANTS WHO COMPLETED ALL PHASES OF THE STUDY

	Numbers of Participants		
Experimental Group	Males	Females	Total
Control	13	18	31
Feedback	26	16	42
Mental Rotation	15	21	36
Total	54	55	109

TABLE VIII

SAMPLE SIZE AFTER EDITING FOR AGE AND ATTRITION

	Number of Participants		
Experimental Group	Males	Females	Total
Control	13	18	31
Feedback	26	16	42
Mental Rotation	14	19	33
Total	53	53	106

TABLE IX

AGE OF PARTICIPANTS IN THE EDITED SAMPLE

		Age in Years		
	N	Range	Median	Mean
Control	31	18 to 28	19	19.9
Males	13	18 to 22	19	18.8
Females	18	18 to 28	20	20.7
Feedback	42	18 to 26	19	19.6
Males	26	18 to 24	19	19.7
Females	16	18 to 26	18	19.5
Mental Rotation	33	18 to 29	19	19.9
Males	14	18 to 22	19	19.6
Females	19	18 to 29	19	20.1
Total Sample	106	18 to 29	19	19.8

Testing Materials and Procedures

One important issue with regard to any type of training is the extent to which training or practice on one specific measure of an ability generalizes to performance on other measures of the same ability construct. That is, if subjects were pretested, trained, and then evaluated again using the same test each time, or even a parallel form thereof, it would be impossible to discern whether any performance improvement was actually indicative of an increase in ability or whether performance improvement was test specific--due to increased familiarity with test instructions and item types.

Three different measures of mental rotation were employed in this study to ensure that the generalizability of training effects could be assessed. Each test represented a different degree of "training relatedness." One test, Mental Rotations (MR), was a three-dimensional rotation test based on Shepard and Metzler's (1971) objects (Vandenberg & Kuse, 1978). It is a measure of Catell's Gv and is, more specifically, reported as a measure of Visualization (Lohman, 1988) rather than Speeded Rotation because it is more complex and less speeded than typical Speeded Rotation tests (e.g., Thurstone's Flags, Figures, and Cards). Mental Rotations had no direct relationship with the training methods employed. Another test, the Depth Plane Object Rotation Test (DPORT), was designed to have a direct relationship with the training methods and was used as a training tool. The third test, the Rotating Three-Dimensional Objects Test (RTDOT), was indirectly related to the training materials. It contained items visually similar to DPORT items, but involved instructions quite different from those for the DPORT and was not employed during training sessions. Both the DPORT and the RTDOT were designed by the author.

Parallel forms of each test (MR, DPORT, and RTDOT) were constructed to eliminate practice effects associated with having seen exactly the same item previously. Form A of Mental Rotations, DPORT Form C, and RTDOT Form A were administered in the pre-training testing sessions to all subjects; Mental Rotations Form B, DPORT Form E, and RTDOT Form B were administered in the post-training testing sessions. Each test is described in greater detail below.

Mental Rotations

Mental Rotations, developed by Vandenberg and Kuse (1978), is based on the three-dimensional figures originally used in Shepard and Metzler's (1971) study of mental rotation. It consists of drawings that show combinations of 10 blocks in various orientations. For each of 20 items the subject is required to determine which two of four alternatives show the same set of blocks as the stimulus does, but in different orientations. Items are scored "correct" only if both correct alternatives are identified. The test has a 10 minute time limit. Test-retest correlations reported are .845 for 21 sixth grade boys and .635 for 30 sixth grade girls (Vandenberg, 1975). A sample object from this task appears in Figure 3.

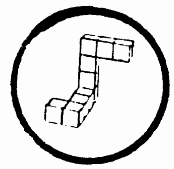


Figure 3. Sample Mental Rotations Object

Construction of Alternate Forms

Alternate forms of this test were constructed by simply splitting the test in half, as other researchers have done (McGee, 1979; Vandenberg, 1978). Each half of the test has ten items and a five minute time limit.

The Rotating Three Dimensional Objects Test

The Rotating Three-Dimensional Objects Test (RTDOT) has two forms, A and B, and each form has 40 items. Each item consists of a pair of pictures that show objects rotated in the depth plane. The subject is asked to decide whether the pictures are of the same object or of its mirror image. On twenty of the 40 items, the pairs of pictures are different views of the same object. The remaining 20 items have pairs of objects that are different (i.e., mirror images). Each form of the RTDOT has a seven minute time limit. A sample problem appears in Figure 4.

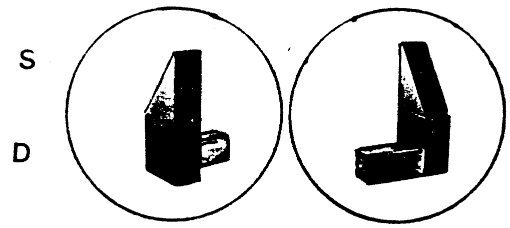


Figure 4. Sample Rotating Three-Dimensional Objects Item

The RTDOT was designed to rely more heavily on "power" than "speed." Most researchers think of power and speed as a continuum; on one end, "pure" power tests allow enough time for all persons to attempt all items while, on the other end, "pure" speeded test are administered such that few, if any, subjects have time to complete the test (Toquam, Dunnette, Corpe, Houston, Peterson, Russell, & Hanson, 1986). The time limit for the RTDOT was set such that subjects would be expected to attempt most items, and the difficulty of individual items was expected to be related to the characteristics of the items rather than their position on the test (i.e., beginning, middle, end).

Construction of Parallel Forms

Constructing parallel forms of a new test requires careful attention to parameters likely to influence the difficulty of individual test items. Three such factors are relevant to the RTDOT: 1) the angular difference between the objects forming each item, 2) the number of "same" and "different" response alternatives that are scored "correct," and 3) the type of object pictured. Of these, the angular difference between the two objects is, perhaps, the most important. Previous mental rotation research suggests that response time and accuracy are significantly related to the angular difference between two objects, regardless of the type, or complexity, of the objects (Shepard & Metzler, 1971). That is, item difficulty is likely to increase as the angular difference between the two objects greatens. Also, on tests for which the subject is asked to respond "same" or "different," it is important to include equal numbers of items that picture the same object (viewed from different orientations) and that picture objects that are different from one another (i.e., mirror-images) because it is possible that "different" items may be more difficult than "same" items. Therefore, to enhance parallelism of RTDOT forms, the types of objects pictured, the angular differences between objects, and the number of same/different responses were balanced within and across forms.

Types of Objects. The objects pictured in RTDOT items were constructed from wooden blocks and styrofoam balls. Ten asymmetrical objects were constructed, five pairs of matching, but unidentical objects. An example of a pair of matching but unidentical objects is provided in Figure 5. One object is constructed entirely from styrofoam balls; the other is constructed from styrofoam balls and a wooden rectangle. The pairs of matching objects were divided into two sets, one for each RTDOT form, such that the forms would be composed from similar objects.

Angular Differences. Each of the 10 objects (i.e., five for each of the two forms) was photographed in positions varying by increments

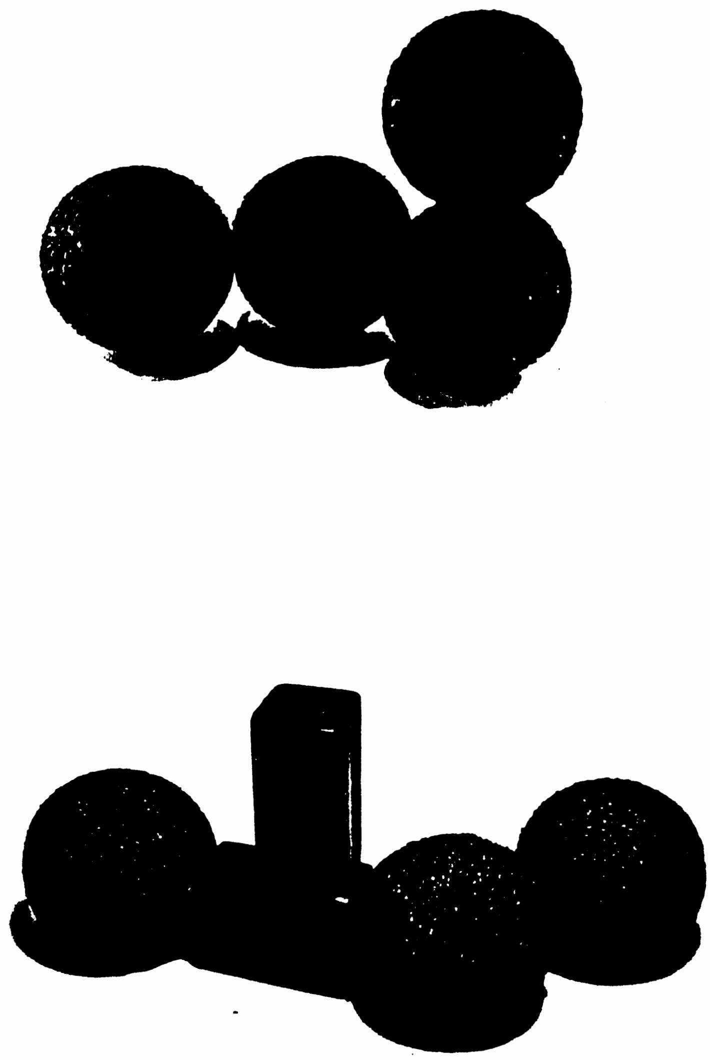


Figure 5. An Example of Matching Objects

of 30 degrees in the depth plane (e.g., 30, 60, 90, 120, 150, and 180 angular degree departures from a standard orientation). Pictures were organized by object and angle and were selected to fill the cells of an object by angle matrix. For example, six pairs of pictures were identified for an object. In the first pair of pictures, the orientation of the object differed by 30 degrees in the depth plane; in the second pair the orientation differed by 60 degrees, and so on. The negatives were then inverted and redeveloped to form mirror-image pictures, and mirror-image pictures were organized into an object by angle matrix. As a result, a large pool of items balanced by object and angle was prepared for each RTDOT form.

Number of Same/Different Correct Responses. Mirror-image pictures were paired with pictures of the original objects to form items that would be scored correct if the subject responded "different," and pictures of original objects (taken from different views) were paired to form "same" items. Equal numbers of "same" and "different" items were selected for various angle by object combinations. As a result, each form of the RTDOT contains 20 "same" items and 20 "different" items, balanced according to angle and object.

Summary. Items within and across the two RTDOT forms were balanced on the basis of the type of object pictured, the angular difference between pictured objects, and the number of "same" and "different" correct responses to enhance the parallel nature of the forms. Information about the test items for each for is summarized in Table X. As shown, both forms contain equal numbers of items where the angular difference between object is 30, 60, 90, 120, 150, and 180

TABLE X

COMPARISON OF ITEMS ON FORMS A AND B
OF THE RTDOT

Angular Difference Between Objects	Number of Items	
	Form A	Form B
30	7	7
60	7	7
90	6	6
120	6	6
150	7	7
180	7	7
Total	40	40

Matched Objects	Number of Items	
	Form A	Form B
Styrofoam Balls	8	8
Triangles	8	8
Cylinder on a Block	8	8
Suspended Oblong on Blank	8	8
Stacked Blocks	8	8
Total	40	40

Correct Response	Number of Items	
	Form A	Form B
Same	20	20
Different	20	20
Total	40	40

degrees. Forms are also parallel in terms of the number of same/different correct responses and the types of objects pictured in items.

Pilot Work

The two forms of the RTDOT were pilot tested to assess the time limit (a tentative time limit had been set at ten minutes), to ensure clarity of the instructions, and to identify any problems with individual items. Four Oklahoma State University graduate students, two undergraduates, and two volunteers who were not students independently reviewed each form. Each pretest participant was instructed to follow this procedure:

1. Obtain a clock with a second hand. Before reading the instructions, record the time on the upper right corner of the test booklet.

2. Read the instructions and work the sample problems. When you are finished, record the time at the bottom of the page.

3. Go back through the instructions and write down any questions, problems, or wording changes.

4. Write the time on the third page of the booklet, and begin working problems as quickly and accurately as possible. When you are finished, write the time on the last page of the test booklet.

5. Go back through the test and identify any items you feel are problematic, and write down your concerns.

6. Repeat this process for each form.

Several changes were made in the tests in response to pretest information. In accordance with suggestions and comments of the reviewers, minor modifications were made in the wording of the instructions, and the time limit was reduced to seven minutes. Several items that had not duplicated clearly were replaced with clearer copies. The final versions of RTDOT forms A and B appear in Appendix A.

The Depth Plane Object Rotation Test

There are three forms of the Depth Plane Object Rotation Test (DPORT), and each form has 30 items. As mentioned previously, DPORT items are visually similar to RTDOT items. Each DPORT item consists of three pictures of three-dimensional objects rotated in the depth plane. The pictures are labeled A, B, and C. Two pictures are different views of the same object; the third picture shows an object that is a mirror-image of the other object. For each item, subjects are asked to identify the object that is different (i.e., a mirror-image) from the other two. Correct answers are balanced across response alternatives: A, B, and C. The DPORT has a six minute time limit, and like the RTDOT, is expected to fall on the power side of the power/speed continuum. A sample item appears in Figure 6.

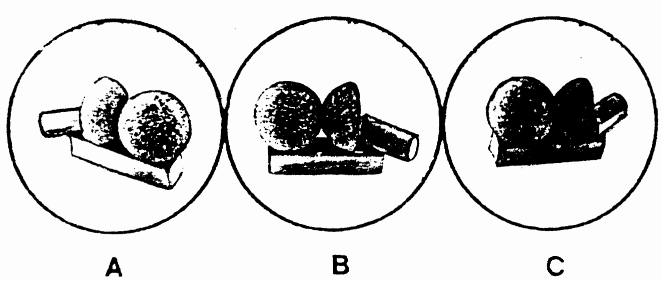


Figure 6. Sample Depth Plan Object Rotation Test Item

Construction of Parallel Forms

The three forms of the Depth Plane Object Rotation Test (DPORT) were constructed from pictures of wooden and styrofoam objects, balanced for angular difference between objects, matched objects, and correct response, as were those for the RTDOT. Figure 7 shows a sample set of matched objects, and Table XI summarizes information about the different forms.

Although the RTDOT and the DPORT were constructed through similar procedures and their items are visually similar to each other, the tests differ in two ways. First, each form of these tests was constructed using its own separate set of objects. None of the items across any two tests are redundant. Second, the tests are formatted differently. For each DPORT item, three pictures are given and subjects are asked to identify the one that is different from the other two (i.e., the mirror image). The DPORT, therefore, involves comparing at least two pairs of pictures to identify the different object.

Pilot Work

The three forms of the DPORT were reviewed by the same individuals who examined the RTDOT, and the same instructions were given. Three types of changes were made in response to reviewers' comments. First, minor wording changes were made in the instructions. Second, the time limit was set at six minutes, and, third, some items were replaced with clearer pictures. The final versions of the three forms of the DPORT appear in Appendix A, along with the RTDOT.

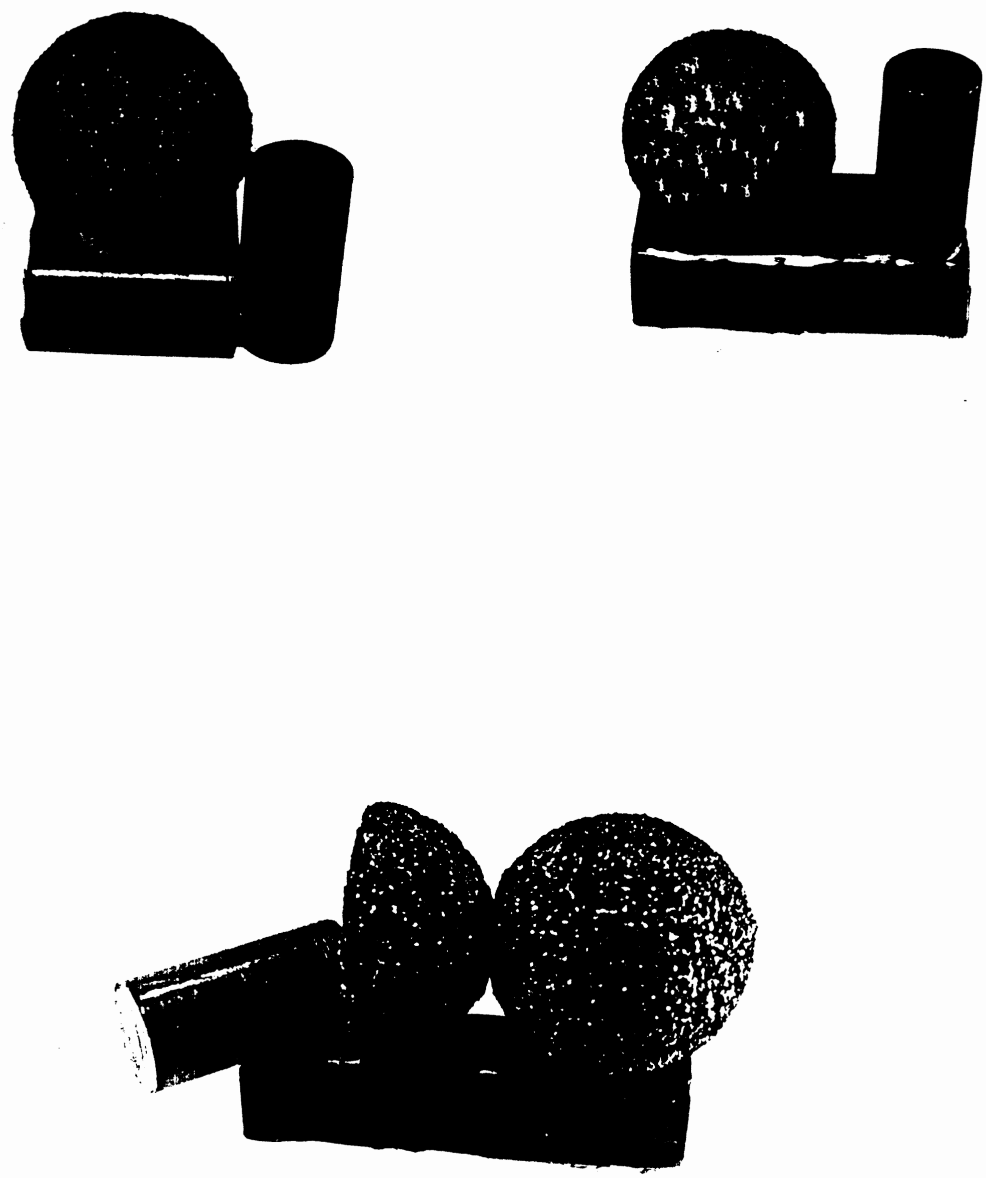


Figure 7. An Example of Matching Objects

TABLE XI

COMPARISON OF ITEMS ON FORMS C, D, AND E OF THE DPORT

	Number of Items		
Angular Difference[1]	Form C	Form D	Form E
30	5	5	5
60	5	5	5
90	5	5	5
120	5	5	5
150	5	5	5
180	5	5	5
Total	30	30	30

	Number of Items		
Matched Objects	Form C	Form D	Form E
Styrofoam Balls	6	6	6
Triangles	6	6	6
Cylinder on a Block	6	6	6
Suspended Oblong on Blank	6	6	6
Stacked Blocks	6	6	6
Total	30	30	30

	Number of Items		
Correct Response	Form C	Form D	Form E
A	10	10	10
B	10	10	10
C	10	10	10
Total	30	30	30

[1] Three pictures appear in each DPORT item, two of one object and one of its mirror image. Angular differences were balanced according to the angular difference between the two pictures of the same object and the angular difference between the mirror image and one of the other two objects.

Training Materials and Procedures

As mentioned, the subjects, male and female Oklahoma State University undergraduates, were divided among three experimental groups--control, feedback training, and mental rotation training. Mental rotation training was designed specifically to provide subjects with experience visualizing and rotating a three-dimensional object in the depth plane. Feedback training included practice and feedback on a mental rotation test. A control group was included to ensure that the training methods could be evaluated adequately. Procedures in each experimental group are described below.

Feedback Training

During each of the two one-hour training sessions, feedback training participants practiced taking mental rotation tests. Forms C and D of the DPORT were administered and timed in each session. After each test, the experimenter walked through the correct answers with participants while they scored their own work.

Mental Rotation Training

Mental rotation training was designed specifically to give subjects practice rotating a three-dimensional object. Each training session consisted of two major components--a picture-object matching task that took about 40 minutes and a testing/feedback session that lasted about 20 minutes.

Picture-Object Matching Task

Subjects spent the first 40 minutes of the session working on a picture-object matching task. Five tables were set up in the room, and, near the center of each table, a pair of objects used to develop DPORT Form C was placed on a rotatable stand. One object of the pair was labeled "Object 1" and the other, "Object 2." Four to five subjects were seated around each table. Each subject was then given a stack of 24 pictures, 12 pictures of each object that was placed on his/her table.

The picture-object matching task consisted of three steps: (1) sorting the pictures into two piles, one for each object, (2) organizing the pictures as though the objects were rotating clockwise in space, and (3) comparing the sorted, organized pictures to an answer sheet which had the pictures taped on a page in clockwise order. Subjects were told that they could move, rotate, or pick up the actual object if they wished to do so.[1] During the task, the experimenter

---

[1] When this experiment was originally proposed, I planned to have subjects sort the pictures into two piles (i.e., without organizing the pictures in accordance with clockwise rotation of the object and comparing answers to an answer sheet). However, I piloted this procedure, informally, on three graduate students and found that adding the process of organizing the pictures as though the object were rotating enhanced the training method without making the task unreasonable in terms of time constraints.

walked through the room helping subjects compare their sorted, organized pictures to the answer sheet. A sample answer sheet is provided in Figure 8.

After completing the picture-object matching task at one table, subjects exchanged tables and repeated the picture-object matching task for a different set of objects.

Testing/Feedback

After all subjects had completed the picture-object matching task at all five tables, all of the objects were removed from the tables and Form C of the DPORT was administered. The experimenter then walked through the correct answers with the group. Procedures used in this portion of the mental rotation training session were, therefore, virtually identical to the procedures used in the feedback training group.

During the second training session the procedures for the first training day were repeated. This time, though, objects used to develop DPORT-D were used in the picture-matching task, and DPORT-D was administered in the testing portion.

Control

During both training sessions, participants in the control group completed four verbal ability measures--Vocabulary: Synonyms, Vocabulary: Antonyms, Definitions, and Word Classification. These tests were pilot measures being developed by Dr. William Jaynes (unpublished research, 1985).

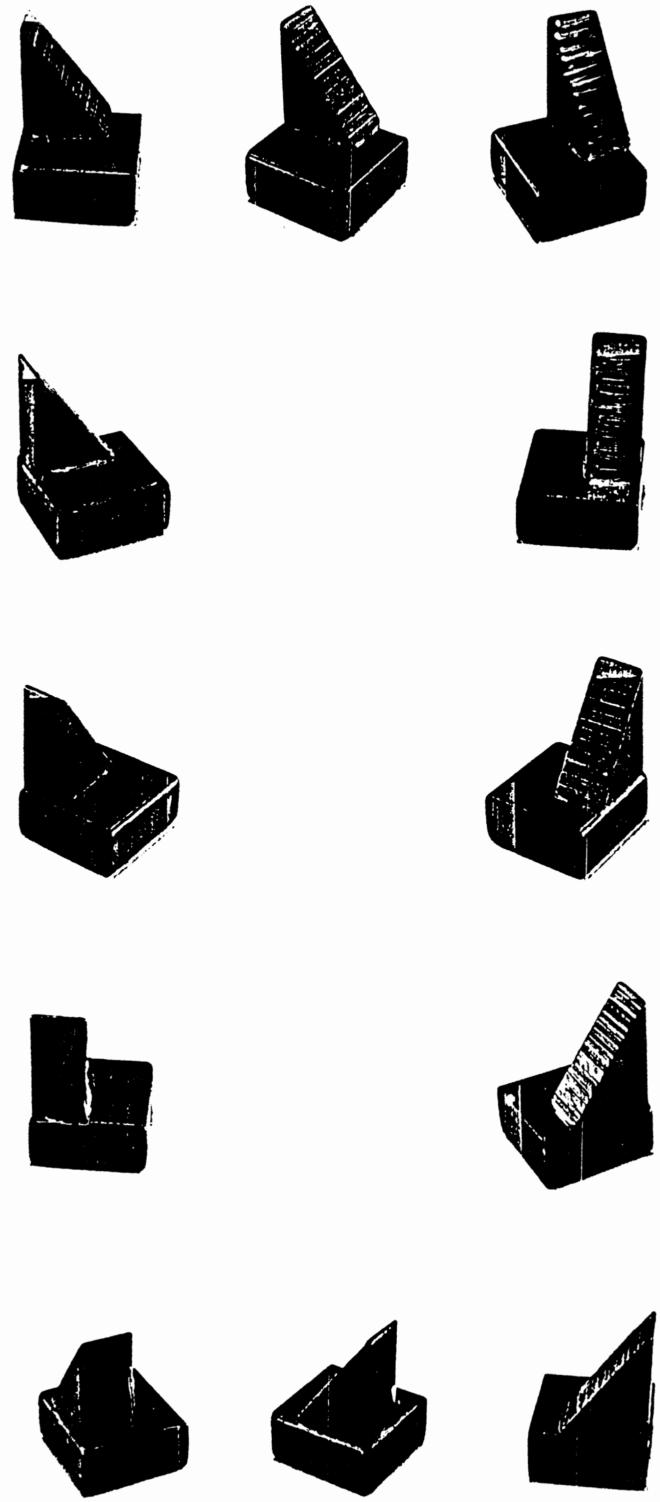


Figure 8. Clockwise Rotation of an Object

Review of Experimental Framework and Procedures

Figure 9 depicts the experimental framework, and key aspects of each part of the study are summarized below.

Pre-Training Testing

Three tests were administered pre-training: Mental Rotations (Form A), the Rotating Three-Dimensional Objects Test (Form A), and the Depth Plane Object Rotation Test (Form C). The sequence with which tests were administered was balanced according to the Latin squares design shown in Table V. Six testing sessions were held, and each included individuals who would later participate in the control, feedback, and mental rotation training groups. In each testing session, the experimenter read the test instructions aloud to the group, allowed time to complete sample problems, answered questions, and timed the test. A total of 134 subjects, 69 males and 65 females were tested.

Training

Feedback Training

Feedback training participants were tested on Forms C and D of the DPORT. After each test, the experimenter walked through the correct answers with participants while they scored their own work.

Control

Control group participants were tested on several verbal ability tests.

	Phases		
Experimental Groups	<u>Pre-Training Testing</u>: . Mental Rotation Test (Form A) . Depth Plane Object Rotation Test (Form C) . Rotating Three Dimensional Objects Test (Form A)	<u>Training</u> Day 1 Day 2	<u>Post-Training Testing</u>: . Mental Rotations Test (Form B) . Depth Plane Object Rotation Test (Form E) . Rotating Three Dimensional Objects Test (Form B)
Control	19 males 20 females	Testing on Verbal Measures	13 males 18 females
Feedback	30 males 21 females	Testing and Feedback on Forms C and D of the DPORT	26 males 16 females
Mental Rotation	20 males 24 females	Picture-Object Matching Task and Testing and Feedback on the DPORT (Forms C and D)	15 males 21 females

Figure 9. Summary of Experimental Framework

Mental Rotation Training

Mental rotation training sessions consisted of two major components--a picture-object matching task that took about 40 minutes and a testing/feedback session that lasted about 20 minutes. Participants were tested on Form C of the DPORT during the first session and on Form D of the DPORT during the second session.

Post-Training Testing

Three tests were administered post-training: Mental Rotations (Form B), the Rotating Three-Dimensional Objects Test (Form B), and the Depth Plane Object Rotation Test (Form E). The sequence with which tests were administered was balanced according to the Latin squares design shown in Table V. Six testing sessions were held, and each included individuals who had participated in the control, feedback, and mental rotation training groups. In each testing session, the experimenter read the test instructions aloud to the group, allowed time to complete sample problems, answered questions, and timed the test. A total of 109 subjects, 54 males and 55 females were tested.

CHAPTER IV

RESULTS

As described previously, the purpose of this study was to examine the effect of training on subjects' performance on mental rotation tests, the performance of females relative to males under different training conditions, and the generalizability of training on one test or measure to performance on a different test. Independent variables were sex and experimental group (i.e., control, feedback, mental rotation); dependent measures were mental rotation test scores.

This chapter describes the results obtained. It is divided into two major sections. The first section describes the psychometric properties of the dependent measures, and the second section describes comparisons made between experimental units.

Psychometric Properties of Dependent Measures

Information about the quality of a test comes from several sources. In particular, we must ascertain how well individual test items appear to be "working" (i.e., how difficult they are, how well they correlate with the total score), how reliable or internally consistent the test is as a whole, and how the test correlates with other tests purported to measure the same ability.

Pre-Training Dependent Measures

Test Score Statistics

Item-level statistics and reliability estimates for the three pre-training measures--Rotating Three Dimensional Objects Test Form A (RTDOT-A), the Depth Plane Object Rotation Test Form C (DPORT-C), and Mental Rotations Form A (MR-A)--are provided in Table XII. Data from all subjects, regardless of whether they completed the experiment, were included in the computations to make best use of all available information about the quality of the tests.

"Difficulty" reported in Table XII is the percent of subjects getting each item correct. Higher numbers, therefore, reflect easier items (larger percentages of subjects responding correctly), and lower numbers reflect greater difficulty. Ideally, the estimated difficulty levels for each test would be .75 for the RTDOT, .50 for the DPORT, and .25 for Mental Rotations (i.e., p(correct)/2 + p(correct)) because each test has a different probability associated with responding correctly by chance (Kaplan & Saccuzzo, 1982). For example, each RTDOT item has two response alternatives, and the probability of responding correctly by chance is, thus, .50. In contrast, each DPORT item has three possible responses (A, B, and C), and the probability of a correct response is .33. As shown in Table XII, the RTDOT and Mental Rotations items were somewhat less difficult than might be desired.

Also reported in Table XII, are the mean and range of item-total score correlations. These are the point biserial correlations between

TABLE XII

TEST SCORE STATISTICS FOR MEASURES ADMINISTERED BEFORE TRAINING (TOTAL SAMPLE)

			Number Correct		Range of Scores		Item Difficulty			Item-Total Score Correlations			Reliability Estimates	
Test	K	N	Mean	S.D.	Max.	Min.	Mean	Range		Mean	Range		R1	Alpha
RTDOT-A	40	134	33.68	5.55	40.00	16.00	.84	.57	.99	.40	.12	.69	.85	.86
DPORT-C	30	133	19.74	6.82	30.00	1.00	.66	.36	.88	.50	.21	.71	.94	.90
MR-A	10	134	5.82	2.55	10.00	0.00	.58	.23	.78	.55	.29	.68	.77	.75

K = Number of Items

N = Number of Subjects

R1 = Odd-Even Correlation, Spearman-Brown Corrected subjects' scores on items (i.e., correct or incorrect) and scores on the whole test, and they indicate the extent to which an item differentiates, or discriminates, between high and low performers. Because point biserial correlations are influenced by the amount of variance on individual items, items in the middle range of difficulty, having more variance, will generally result in higher point-biserials and vice versa. As shown in Table XII, for example, the item difficulties for Mental Rotations items were moderate (i.e., averaging .58 in difficulty) compared to those of the DPORT and RTDOT, and the average point biserial correlation for MR items is high relative to those for the DPORT and RTDOT. Nevertheless, the item-total score correlations for each test are adequate.

Reliability Estimates

Also reported in Table XII are two reliability estimates--the odd-even split halves reliability adjusted for total test length (via the Spearman-Brown formula, $R = 2r/(1 + r)$) and coefficient alpha (Cronbach, 1951). Although both coefficients are measures of internal consistency, alpha, or its mathematical equivalent KR_{20}, represents a lower bound on the reliability because it simultaneously considers all possible ways of splitting a test. Internal consistency estimates are usually thought of as measures of the homogeneity of test items in terms of the construct being measured. When a test is highly speeded, however, high internal consistency estimates may be a consequence of subjects' failure to complete the test. In an extreme case where most subjects only complete half of the items, odd-even split halves reliability would be high, and the front-back split halves coefficient would probably be low; neither coefficient would adequately describe the homogeneity of test items.

Percent completion rates and item difficulty levels were examined to assess the idea that the RTDOT and DPORT are on the power side of the power/speed continuum and to ensure that the odd-even split halves estimate is useful as a measure of internal consistency for the RTDOT and DPORT. Previous studies have used a criterion of 80 percent or more completion to define a power test (Toquam et. al, 1987); that is, if subjects attempt 80 percent or more of the items, the test is referred to as a power test. On the average, subjects completed 99 percent of the RTDOT items (39.5 items out of 40) and 85 percent of the DPORT items (25.3 items out of 30). These data suggest that the RTDOT and DPORT rely more heavily on power than speed.

On speeded tests, item difficulties are a function of items' placement on the test; that is, difficulties decrease as the item number increases. For RTDOT items this trend is virtually nonexistent. The pearson product moment correlation between item number and item difficulty was -.07, indicating only a slight trend for difficulties to drop with position on the test. Examination of the DPORT item difficulties, however, indicated a trend toward lower item difficulties after the twenty-fourth item. This, together with the smaller percent completion rate for the DPORT, suggests that the DPORT is somewhat more speeded than the RTDOT and that the internal consistency estimates provided in Table XII for the DPORT were inflated by the inclusion of the last five or six items. Internal consistency estimates were, therefore, recomputed for the DPORT. When only the first 20 items are included, alpha is .83 and the corrected split halves estimate is .86. These estimates, still adequate in magnitude, more accurately reflect the internal consistency of the DPORT than those in Table XII.

Similarly, item difficulties for Mental Rotations drop substantially after the eighth item (out of ten items). Internal consistency estimates computed across only the first eight were somewhat lower than those provided in Table XII--alpha = .71 and corrected split halves = .71.

Post-Training Dependent Measures

Three tests were administered post-training--Rotating Three Dimensional Objects Test Form B (RTDOT-B), the Depth Plane Object Rotation Test Form E (DPORT-E), and Mental Rotations Form B (MR-B). Psychometric properties of these measures are described below.

Test Score Statistics and Reliability Estimates

Item-level statistics and reliability estimates for the three post-training measures are provided in Table XIII. The range of scores and item difficulty levels suggest that the post-training measures were somewhat "easier" than their pre-training counterparts (see Table XII). This observation is, naturally, expected when post-training performance improvement occurs. Similarly, the item-total score correlations and reliability estimates are somewhat lower for post-training measures than for pre-training measures because there was less variance in post-training performance. Even so, the alpha and split halves reliability

TABLE XIII

TEST SCORE STATISTICS FOR MEASURES ADMINISTERED AFTER TRAINING (TOTAL SAMPLE)

			Number Correct		Range of Scores		Item Difficulty			Item-Total Score Correlations			Reliability Estimates	
Test	K	N	Mean	S.D.	Max.	Min.	Mean	Range		Mean	Range		R1	Alpha
RTDOT-B	40	110	35.04	3.97	40.00	23.00	.88	.57	.99	.32	.00	.57	.76	.78
DPORT-E	30	110	26.97	3.46	30.00	14.00	.90	.59	.99	.41	-.03	.70	.87	.83
MR-B	10	110	5.88	2.43	10.00	0.00	.59	.29	.77	.51	.26	.65	.63	.69

K = Number of Items

N = Number of Subjects

R1 = Odd-Even Correlation, Spearman-Brown Corrected estimates suggest that these measures were adequate with regard to internal consistency.

On the average, subjects completed over 99 percent of the RTDOT-B items (39.9 items out of 40), and likewise an average of 99 percent of the DPORT-E items (29.8 items out of 30) were attempted. Similarly, post-training RTDOT and DPORT item difficulties did not decrease with position on the test. Mental Rotations (Form B) item difficulties did drop somewhat for the last two items; the effect, however, was not as pronounced as the drop in Form A difficulties reported earlier. Internal consistency estimates based on only the first eight items were .61 (alpha) and .54 (corrected split halves).

Test Score Correlations and Test-Retest Reliability

Correlations between scores on pre- and post-training measures are reported in Table XIV. Correlations between pre-training measures ranged from .50 to .57, and post-training measures' correlations ranged from .38 for Mental Rotations with the DPORT to .64 for the DPORT and RTDOT. The correlations between parallel forms provided in Table XIV are test-retest parallel forms reliability estimates. Because three sources of error attenuate test-retest parallel forms estimates--random error, error due to time between administrations, and error due to differences between forms--they are a rigorous test of reliability. For this reason, Cureton (1971) and Campbell (1976) suggest using Spearman's correction for attenuation to derive a "stability coefficient" (r_s), the parallel forms reliability estimate corrected for lack of internal consistency in individual forms.

TABLE XIV

TEST SCORE CORRELATIONS*

	Pre-Training Measures			Post-Training Measures		
	MR-A	RTDOT-A	DPORT-C	MR-B	RTDOT-B	DPORT-E
MR-A						
RTDOT-A	.50 (133)					
DPORT-C	.57 (133)	.50 (133)				
MR-B	[.64 (109)]	.38 (109)	.52 (109)			
RTDOT-B	.53 (109)	[.55 (109)]	.49 (109)	.45 (110)		
DPORT-E	.46 (109)	.48 (109)	[.39 (109)]	.38 (110)	.64 (110)	

* Sample size is provided in parentheses beneath the correlation coefficient.

[ ] indicates correlations between scores on parallel forms.

The formula for r_s is:

$$r_s = \frac{r_{12}}{\sqrt{r_{11} r_{22}}}$$

where r_{12} is the Pearson product moment correlation between the two forms and r_{11} and r_{22} are internal consistency estimates for each form. By controlling for lack of internal consistency, this coefficient permits a more accurate assessment of instability of the trait over time.

Stability coefficients, computed using alpha coefficients as measures of internal consistency, were .97 for Mental Rotations (Forms A and B), .67 for RTDOT (Forms A and B), and .47 for DPORT (Forms C and E). [Recall that two alpha coefficients have been presented for three of the tests, one based on a subset of items and one based on the full set of items. Smaller, more conservative, alpha estimates were used in this computation.] These data suggest that Mental Rotations test performance was highly stable over time and that DPORT scores were less stable than those for other tests.

Construct Measurement

The correlations between tests (reported in Table XIV) were not as high as might be expected, given that the tests should measure the same construct, using different methods. More specifically, correlations in the range of .65 to about .75 were expected. To examine the impact of different attenuating sources of variance on the correlations, the author sorted the 15 correlations provided in Table XIV into three groups according to the type of variance expected to attenuate the correlation and corrected the correlations for attenuation (using Spearman's correction). These data appear in Table XV.

The correlations between parallel forms administered a week apart provide estimates of the test-retest parallel forms reliability as discussed in the section above. All of the other correlations reported in Table XV address the construct validity of the tests; that is, because they are correlations between scores on different tests, the correlation can be interpreted as the extent to which the two tests measured the same construct. As shown in Table XV, correlations between different tests administered the same day are attenuated by three major sources of variance: 1) lack of internal consistency, 2) the method of measurement, and 3) random error. Correlations between different tests administered a week apart are attenuated by instability over time as well as these three sources of variance. [Also, all correlations between the DPORT-E and other tests are likely to have been attenuated by the ceiling effect on this test. This is discussed later in greater detail.] One would, therefore, expect correlations between different tests given a week apart to be somewhat lower than those between different test given the same day. As shown in Table XV, there is a slight trend in this direction.

The corrected correlations in Table XV provide an estimate of the correlation between the two tests, had each test been perfectly internally consistent. Other attenuating sources of variance such as the method of measurement are not part of the correction. The corrected correlations between different tests do, therefore, appear to

Table XV

SOURCES OF VARIANCE AND CORRECTED CORRELATIONS BETWEEN TESTS

Tests	Sources of Attenuation in Correlation: Lack of Internal Consistency	Instability Over Time	Method of Measurement	Error	Correlation: Uncorrected	Corrected
Parallel Forms Administered A Week Apart:						
MR-A, MR-B	X	X		X	.64	.97
RTDOT-A, RTDOT-B	X	X		X	.55	.67
DPORT-C, DPORT-E	X	X		X	.39	.47
Mean					.53	.70
Different Tests Administered The Same Day:						
MR-A, RTDOT-A	X		X	X	.50	.63
MR-A, DPORT-C	X		X	X	.57	.74
RTDOT-A, DPORT-C	X		X	X	.50	.59
MR-B, RTDOT-B	X		X	X	.45	.65
MR-B, DPORT-E	X		X	X	.38	.53
RTDOT-B, DPORT-E	X		X	X	.64	.80
Mean					.51	.66
Different Tests Administered A Week Apart:						
MR-A, RTDOT-B	X	X	X	X	.53	.71
MR-A, DPORT-E	X	X	X	X	.46	.60
RTDOT-A, MR-B	X	X	X	X	.38	.52
RTDOT-A, DPORT-E	X	X	X	X	.48	.57
DPORT-C, RTDOT-B	X	X	X	X	.49	.61
DPORT-C, MR-B	X	X	X	X	.52	.73
Mean					.48	.62

be reasonable (i.e., ranging from .52 to .80 with a mean of .64) given that they are unadjusted for several other attenuating sources of variance.

Between Group Comparisons

As mentioned previously, the data reported in preceding portions of this chapter included all possible subjects, regardless of whether they completed the experiment, because the psychometric properties of the tests were of primary interest. This portion of the chapter focuses on the experimental treatment groups and performance of females relative to males on the pre-training dependent measures. Subjects who did not complete the study were, therefore, eliminated. Subjects over the age of 30 were also deleted to avoid potentially confounding effects due to age.

Effect Size

Means and standard deviations of pre-training test scores for males, females, and sex by experimental group combinations are provided in Table XVI, and Table XVII provides means of post-training test scores. The magnitude of the sex difference as measured by the standardized difference between male and female mean scores,

$$d = \frac{\bar{X}_m - \bar{X}_F}{S_p}$$

where,

$$S_p = \sqrt{\frac{(N_m - 1)s^2_m + (N_F - 1)s^2_f}{(N_m - 1) + (N_F - 1)}}$$

TABLE XVI

MEANS AND STANDARD DEVIATIONS OF PRE-TRAINING TEST SCORES

	Males			Females			
Test	N	Mean	S.D.	N	Mean	S.D.	d*
MR-A	53	6.57	2.56	53	5.09	2.28	.61
RTDOT-A	53	35.43	4.37	53	32.62	6.02	.53
DPORT-C	53	20.87	7.01	53	18.81	5.80	.32

* d is the standardized mean difference between males' and females' scores. A positive value indicates superior performance by males, and a negative d indicates higher performance by females.

$$d = \frac{\overline{X}_m - \overline{X}_F}{S_p} \quad \text{where,} \quad Sp = \sqrt{\frac{(N_m - 1)s^2{}_m + (N_F - 1)s^2{}_f}{(N_m - 1) + (N_F - 1)}}$$

TABLE XVII

MEANS AND STANDARD DEVIATIONS OF POST-TRAINING TEST SCORES

	Males			Females			
Test	N	Mean	S.D.	N	Mean	S.D.	d*
MR-B	53	6.11	2.32	53	5.72	2.57	.16
RTDOT-B	53	35.89	3.76	53	34.34	3.80	.41
DPORT-E	53	27.53	2.76	53	26.55	3.70	.30

* d is the standardized mean difference between males' and females' scores. A positive value indicates superior performance by males, and a negative d indicates higher performance by females.

$$d = \frac{\overline{X}_m - \overline{X}_F}{S_p} \quad \text{where,} \quad Sp = \sqrt{\frac{(N_m - 1)s^2{}_m + (N_F - 1)s^2{}_f}{(N_m - 1) + (N_F - 1)}}$$

was .61 for Mental Rotations (Form A), .53 for RTDOT (Form A), and .32 for DPORT (Form C). Males, therefore, outperformed females by one-third to two-thirds of a standard deviation on the pre-training measures. Post-training, the standardized mean difference between males and females scores was .16 for Mental Rotations (Form B), .41 for RTDOT (Form B), and .30 for DPORT (Form E), and males outperformed females by .16 to .30 of a standard deviation on the post-training measures.

Item Difficulties

One concern embedded in the finding of group (e.g., sex or race) differences on a test is whether the test items work in the same way for the different groups (Cole, 1981). That is, if items that are difficult for males are easy for females and vice versa, there may be some bias in individual items. Item difficulties, for each test, were, thus, examined by sex to discern whether individual items appear to work in the same way for males and females. For each item on each test, item difficulties were computed separately for males and females. For each test, the correlation between the two vectors (male and female item difficulties) was computed; these correlations (reported in Table XVIII) ranged from .83 to .94 across the three pre-training tests, indicating similar patterns of item difficulties for males and females. Likewise, the correlations between item difficulties on the post-training measures ranged from .88 to .89 across the three tests, indicating similar patterns of item difficulties for males and females. These patterns are illustrated for Mental Rotations (Form A) in Figure

TABLE XVIII

ITEM DIFFICULTIES BY SEX

Measure	K	Mean Item Difficulty Across Items: Males	Mean Item Difficulty Across Items: Females	Mean Difference	r
MR-A	10	.66	.50	.16	.91
RTDOT-A	40	.88	.82	.07	.83
DPORT-C	30	.69	.62	.07	.94
MR-B	10	.61	.57	.04	.89
RTDOT-B	40	.89	.86	.03	.89
DPORT-E	30	.91	.89	.02	.88

r = Pearson product moment correlation between male and female item difficulties.

10 and for Mental Rotations (Form B) in Figure 11. As shown in both figures, the profiles of item difficulty are highly similar for males and females.

Stability of Individual Differences

As described previously, stability coefficients indicated that performance on Mental Rotations was highly stable across testing sessions and that performance on the DPORT was less stable than that on the other tests. To further explore the stability of individual performance on these measures, Pearson product moment correlation coefficients were computed for each experimental unit (i.e., each experimental group and males and females in each experimental group). These coefficients are reported in Tables XIX, XX, and XXI.

As shown in Table XIX, correlations between pre- and post-training Mental Rotations scores were significant for all experimental units. That is, the rank order of individual scores was highly similar across the two testing sessions. Similarly, some of the correlations between pre- and post-training RTDOT scores are quite high (see Table XX), indicating, with a few exceptions, little change in pre- and post-training ordering of scores within experimental groups.

The correlations reported in Table XXI (for the DPORT) are notably lower than those in Tables XIX and XX; several did not reach significance. At first glance, these correlations appear to suggest change in the rank ordering of individuals' scores; however, further examination of DPORT-E data suggests that the low correlations are probably due to lack of variance in scores on this test. For example, males in the

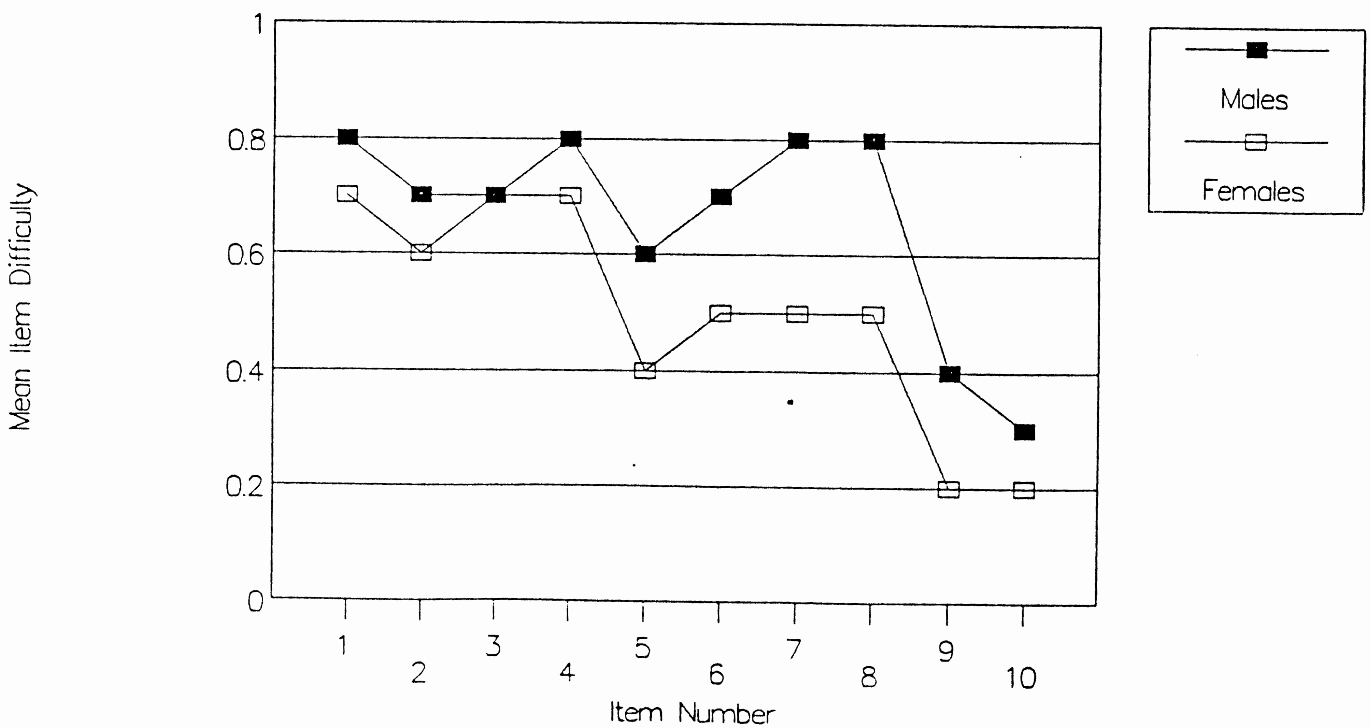


Figure 10. Mental Rotations (Form A) Item Difficulties by Sex

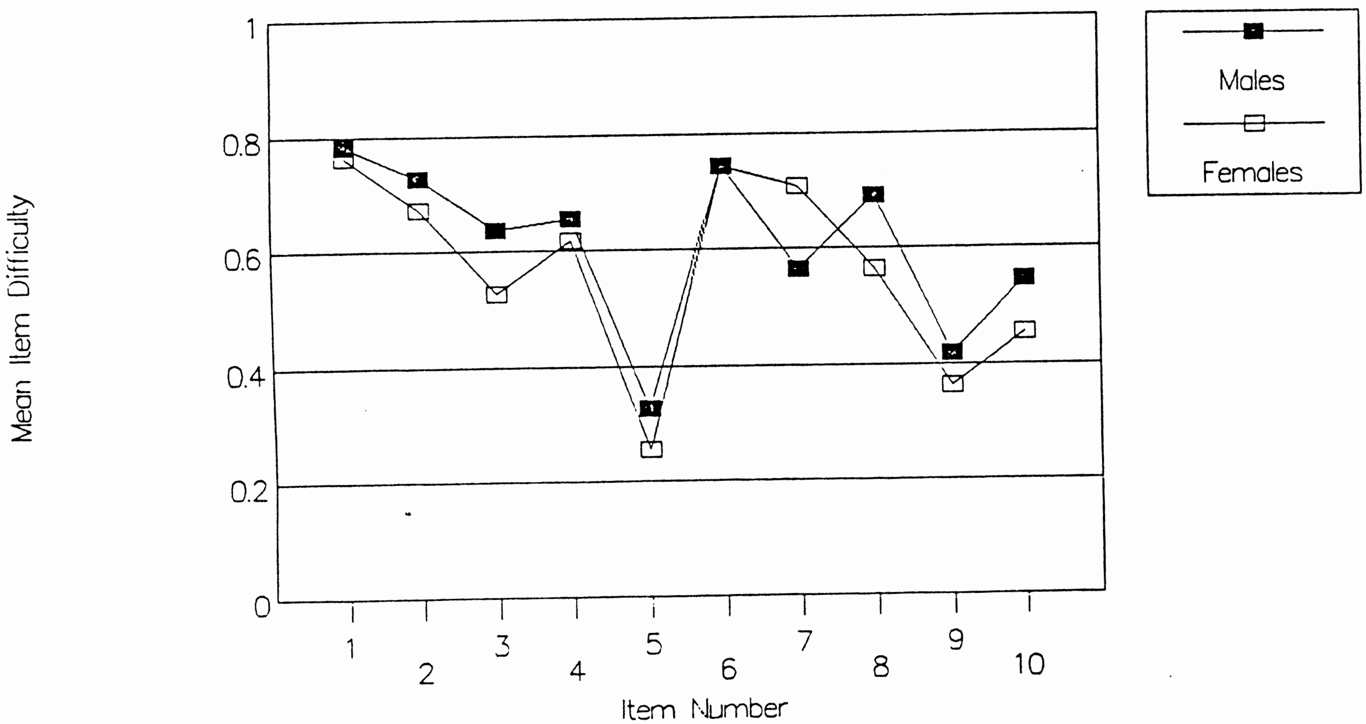


Figure 11. Mental Rotations (Form B) Item Difficulties by Sex

TABLE XIX

MEANS, STANDARD DEVIATIONS, AND CORRELATIONS OF PRE- AND POST-TRAINING MENTAL ROTATIONS TEST SCORES

Experimental Group	N	MR-A Mean	MR-A S.D.	MR-B Mean	MR-B S.D.	Pearson r
All Subjects	106	5.83	2.52	5.92	2.44	.64**
Males	53	6.57	2.56	6.11	2.32	.62**
Females	53	5.09	2.28	5.72	2.57	.67**
Control	31	5.19	2.47	5.58	2.57	.65**
Males	13	6.08	2.81	5.54	2.15	
Females	18	4.56	2.04	5.61	2.89	
Feedback	42	5.81	2.65	6.10	2.40	.55**
Males	26	6.15	2.65	6.04	2.58	
Females	16	5.25	2.65	6.19	2.14	
Mental Rotation	33	6.46	2.32	6.00	2.42	.76**
Males	14	7.79	1.81	6.79	1.89	
Females	19	5.47	2.20	5.42	2.65	

* $p < .05$

** $p < .01$

TABLE XX

MEANS, STANDARD DEVIATIONS, AND CORRELATIONS OF PRE- AND POST-TRAINING ROTATING THREE-DIMENSIONAL OBJECTS TEST SCORES

Experimental Group	N	RTDOT-A		RTDOT-B		Pearson
		Mean	S.D.	Mean	S.D.	r
All Subjects	106	34.03	5.42	35.11	3.84	.54**
Males	53	35.43	4.37	35.89	3.76	.48**
Females	53	32.62	6.02	34.34	3.80	.56**
Control	31	33.74	5.11	32.94	3.96	.50**
Males	13	35.00	3.22	33.62	4.15	
Females	18	32.83	6.06	32.44	3.85	
Feedback	42	34.45	5.27	35.83	3.21	.54**
Males	26	35.15	5.61	36.39	3.53	
Females	16	33.31	4.63	34.94	2.44	
Mental Rotation	33	33.76	5.99	36.24	3.73	.68**
Males	14	36.36	2.27	37.07	3.08	
Females	19	31.84	7.14	35.63	4.13	

* $p < .05$

** $p < .01$

TABLE XXI

MEANS, STANDARD DEVIATIONS, AND CORRELATIONS OF PRE- AND POST-TRAINING DEPTH PLANE OBJECT ROTATION TEST SCORES

Experimental Group	N	DPORT-C Mean	DPORT-C S.D.	DPORT-E Mean	DPORT-E S.D.	Pearson r
All Subjects	106	19.84	6.48	27.04	3.29	.39**
Males	53	20.87	7.01	27.53	2.76	.29*
Females	53	18.81	5.80	26.55	3.70	.47**
Control	31	19.61	5.45	25.29	3.57	.41*
Males	13	19.00	6.53	25.69	3.30	
Females	18	20.06	4.67	25.00	3.82	
Feedback	42	19.98	6.67	27.60	2.31	.28*
Males	26	21.00	7.03	27.35	2.48	
Females	16	18.31	5.86	28.00	2.00	
Mental Rotation	33	19.88	7.28	27.97	3.52	.53*
Males	14	22.36	7.51	29.57	0.85	
Females	19	18.05	6.71	26.79	4.25	

* $p < .05$

** $p < .01$

mental rotation training group performed so well on the test (X = 29.57, out of 30 items) that there is little room for variation in their scores (SD = .85). The correlation between their pre- and post-training scores is zero--an artifact of the low variance in DPORT-E scores--and does not necessarily mean that the rank ordering of males in the mental rotation training group changed from pre- to post-training. In sum, the lack of variation in DPORT-E scores prohibits strong conclusions about the rank ordering of individuals' scores on this test.

Repeated Measures Analysis of Variance (ANOVA)

The Model and Considerations

Repeated measures ANOVAs were conducted, on each dependent variable, to assess the hypotheses posed in this study. Three fixed factors were considered in the model: sex (two levels: male and female), experimental method (three levels: control, feedback training, and mental rotation training), and trials (two levels: pre- and post-training). In sum, the design was 2 x 3 x 2 with repeated measures on the third factor. The structural model has the form:

$$Y_{ijkm} = u + \alpha_i + \beta_j + \alpha\beta_{ij} + \pi_{m(ij)} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \gamma\pi_{km(ij)} + e_{(ijkm)},$$

where Y_{ijkm} denotes a measurement on subject _m_ in the i^{th} gender group, in the j^{th} experimental group, and on the k^{th} trial (i.e., pre- or post-training). α_i is the effect of being in the i^{th} gender group. β_j is the effect of receiving the j^{th} training method, and γ_k is the term associated with the two different "trials" (pre- and post-training). A subject within group G_{ij} is identified by the subscript "m(ij)" to indicate that the subject effect is nested under both the sex and the method factors. Table XXII provides the expected mean squares for the general case of three fixed factors (one of which is repeated) assuming there are equal numbers of subjects within each group (i.e., sex by method combination).

In the current study, however, the numbers of subjects in the experimental groups and in sex by experimental group combinations were unequal, making the design unbalanced. In such a case, variations due to overall main effects and interactions are not additive (i.e., $SS_{a+b} \neq SS_a + SS_b$). As a result, the maximum likelihood method of adding up sums of squared deviations from cell means and the grand means will not yield maximum likelihood estimates of effects. Computationally, this problem can be overcome by using least squares, weighted squares of means, or unweighted means computational procedures instead of the maximum likelihood method. Such procedures involve adjusting the sums of squares for one factor for each other factor and its interactions.

Even though the computational problem associated with nonadditivity can be overcome rather simply, interpretive problems still exist. Because the estimates of effects are not statistically independent across factors and their interactions, the factors are no longer orthogonal. For this reason, some statisticians suggest conducting a series of hierarchical tests wherein higher-order models are computed first, insignificant interactions are dropped, and the model is recomputed. This approach, however, increases the experiment

TABLE XXII

EXPECTED MEAN SQUARES FOR A BALANCED THREE FIXED FACTOR DESIGN WITH REPEATED MEASURES ON ONE FACTOR

Source	df	E(MS)
Between Subjects		
Sex	p-1	$\sigma_\varepsilon^2 + r\sigma_\pi^2 + nqr\sigma_\alpha^2$
Method	q-1	$\sigma_\varepsilon^2 + r\sigma_\pi^2 + npr\sigma_\beta^2$
Sex*Method	(p-1) (q-1)	$\sigma_\varepsilon^2 + r\sigma_\pi^2 + nr\sigma_{\alpha\beta}^2$
Error	pq (n-1)	$\sigma_\varepsilon^2 + r\sigma_\pi^2$
Within Subjects		
Trials	r-1	$\sigma_\varepsilon^2 + \sigma_{\gamma\pi}^2 + npq\sigma_\gamma^2$
Trials*Sex	(p-1) (r-1)	$\sigma_\varepsilon^2 + \sigma_{\gamma\pi}^2 + nq\sigma_{\alpha\gamma}^2$
Trials*Method	(q-1) (r-1)	$\sigma_\varepsilon^2 + \sigma_{\gamma\pi}^2 + np\sigma_{\beta\gamma}^2$
Trials*Sex*Method	(p-1) (q-1) (r-1)	$\sigma_\varepsilon^2 + \sigma_{\gamma\pi}^2 + n\sigma_{\alpha\beta\gamma}^2$
Error	pq (n-1) (r-1)	$\sigma_\varepsilon^2 + \sigma_{\gamma\pi}^2$

error rate and, in hypothesis testing, requires some form of adjustment for the number of models considered in the design. Another approach is to select one model to examine, based on theoretical expectations.

Choice of an approach appears to be dependent upon the purpose of the analysis. In studies that are purely descriptive, it may make sense to disregard insignificant interactions and examine a variety of models. In contrast, it is reasonable to examine only a selected model when specific hypotheses are under consideration, especially if those hypotheses involve expectations about interaction terms.

The model described previously was the only one examined in the current study because interaction terms were very important to the hypotheses posed. For example, significant sex by method by trials or sex by trials interactions might (depending upon the direction of the simple effects) support the hypothesis of differential improvement of females over males. Also, a significant method by trials interaction might suggest performance improvement by subjects in training groups over those in the control group. Moreover, the within-subjects and interaction term sources of variation were of primary interest in the current work.

Homogeneity of Error Variance

One of the basic assumptions underlying analysis of variance models is that the variances due to experimental error within treatments are homogeneous. In ANOVA designs without repeated measures, the F statistic is robust to moderate departures from homogeneity. In repeated measures designs, however, the assumption of compound symmetry (i.e., equal variance-covariance matrices) is highly restrictive, and the *F* test will be biased when this assumption does not hold (Box, 1954). In learning experiments such as the one reported here, assumptions of compound symmetry are rarely met. Indeed, one expectation is that patterns of variance-covariance matrices will differ across treatment groups after the intervention. Fortunately, multivariate test statistics such as Hotelling's T^2 are robust to such violations (Winer, 1971). Therefore, in the analyses reported below, multivariate test statistics were computed for each within subjects effect and were compared to univariate *F* statistics. Any discrepancies in conclusions based on univariate vs. multivariate statistics are reported.

Results

Three least squares, 2 (sex) x 3 (method) x 2 (trials) repeated measures ANOVAs were computed, one for each dependent variable. Although significant main effects for sex were expected on each dependent measure, within subjects effects were of primary interest.

For Mental Rotations, a significant main effect for sex [F (1, 100) = 4.984, $p < .05$] and a significant interaction between trials (pre- and post-training) and sex [F (1, 100) = 8.589, $p < .01$] were observed. Table XXIII provides the summary ANOVA. Analysis of simple effects for the interaction between trials and sex yielded a significant sex effect at pretest [F (1, 100) = 10.92, $p < .001$] and an insignificant sex effect at posttest [F (1, 100) = .60, $p > .05$]. This interaction is illustrated in Figure 12. It is also worth noting here

TABLE XXIII

REPEATED MEASURES ANALYSIS OF VARIANCE ON PRE- AND POST-TRAINING MENTAL ROTATION TEST SCORES

Source	Sum of Squares	DF	Mean Square	F-Ratio	P
Between Subjects					
Sex	48.359	1	48.359	4.984	.028
Method	26.463	2	13.231	1.364	.260
Sex*Method	19.982	2	9.991	1.030	.361
Error	970.385	100	9.704		
Within Subjects					
Trials	.115	1	.115	.055	.816
Trials*Sex	18.068	1	18.068	8.589	.004
Trials*Method	8.574	2	4.287	2.038	.136
Trials*Sex*Method	.945	2	.073	.225	.799
Error	210.357	100	2.104		

N = 106

Squared Multiple R = .14 on MR-A scores and .03 on MR-B scores.

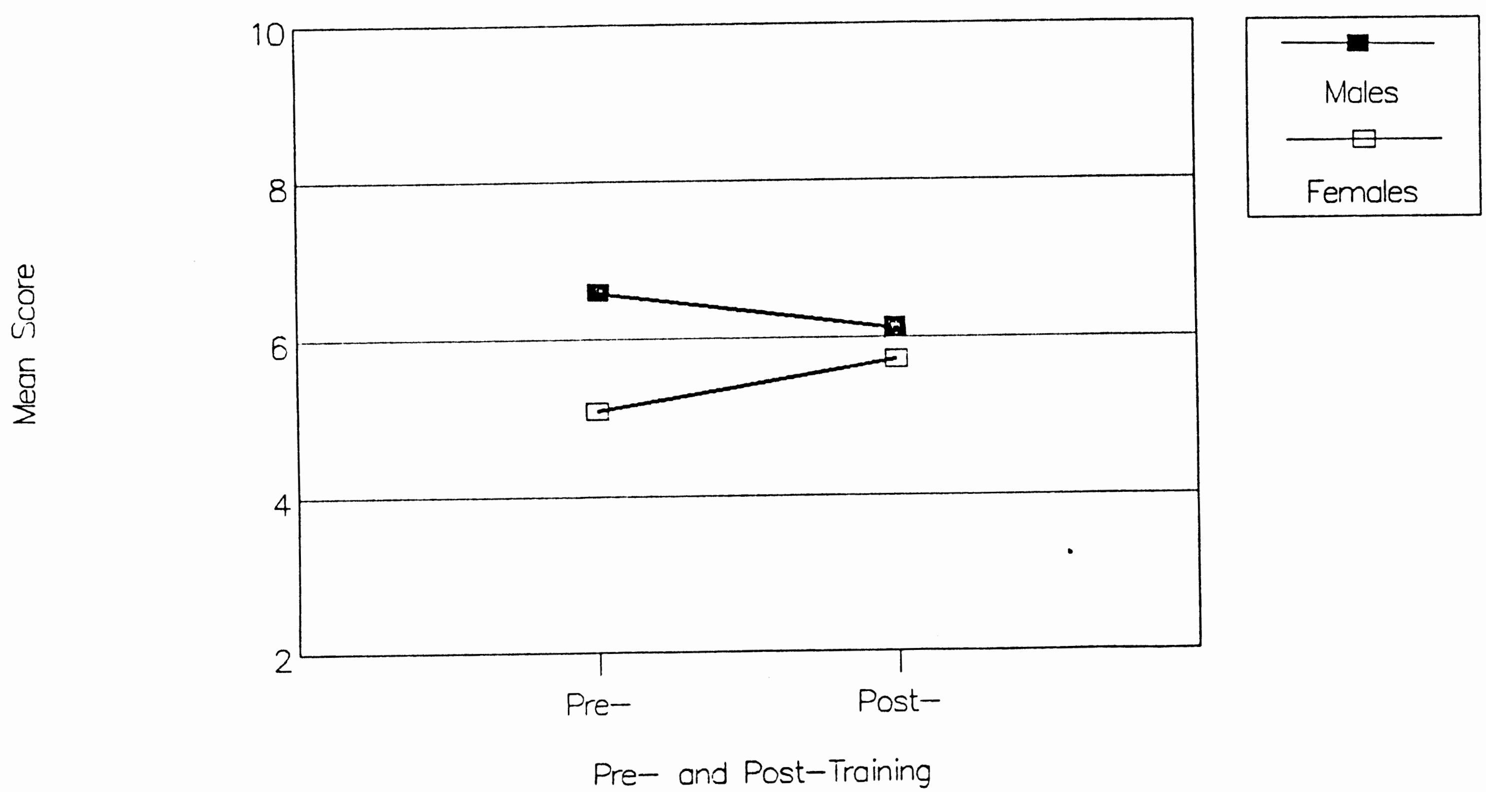


Figure 12. Pre- and Post-Training Mental Rotations Test Scores by Sex

that males post-training scores were slightly less than their pre-training scores; this finding in itself is not statistically significant. It may, however, suggest that the Mental Rotations Test is not highly susceptible to a learning effect.

A significant main effect for sex was also observed on RTDOT scores; Table XXIV provides the summary ANOVA. As shown, the main effect for trials and the interaction between trials and method were significant [F (1, 100) = 4.382, p < .05, and F (2, 100) = 4.149, p < .05, respectively]. Although the univariate F statistic for the sex by trials interaction did not reach significance [F (1, 100) = 2.80, p > .05], the multivariate statistics for this interaction were significant [Rao's F statistic for Wilks' Lambda (2, 99) = 3.66, p < .05; Pillai's F approximation (3, 99) = 3.66, p < .05; Hotelling-Lawley F (2, 99) = 3.66, p <.05]. Analysis of simple effects indicated a significant sex effect pre-training [F (1, 100) = 7.16, p < .05] and an insignificant post-training sex effect [F (1, 100) = 3.57, p > .05]. This interaction is illustrated in Figure 13.

Scheffe's S multiple comparison procedure was used to further assess the significant method by trials interaction. Three contrasts were of interest for each level of the trials factor: control and feedback training, control and mental rotation training, and feedback and mental rotation training. For the first level of the trials factor (i.e., pre-training RTDOT scores) differences between treatment group means were insignificant. For post-training RTDOT scores, the feedback

TABLE XXIV

REPEATED MEASURES ANALYSIS OF VARIANCE ON PRE- AND POST-TRAINING RTDOT SCORES

Source	Sum of Squares	DF	Mean Square	F-Ratio	P
Between Subjects					
Sex	221.371	1	221.371	7.065	.009
Method	111.442	2	55.721	1.778	.174
Sex*Method	19.296	2	9.648	.308	.736
Error	3133.157	100	31.332		
Within Subjects					
Trials	43.641	1	43.641	4.382	.039
Trials*Sex	27.885	1	27.885	2.800	.097
Trials*Method	82.633	2	41.317	4.149	.019
Trials*Sex*Method	16.932	2	8.466	.850	.430
Error	995.868	100	9.959		

N = 106

Squared Multiple R = .08 on RTDOT-A scores and .17 on RTDOT-B scores.

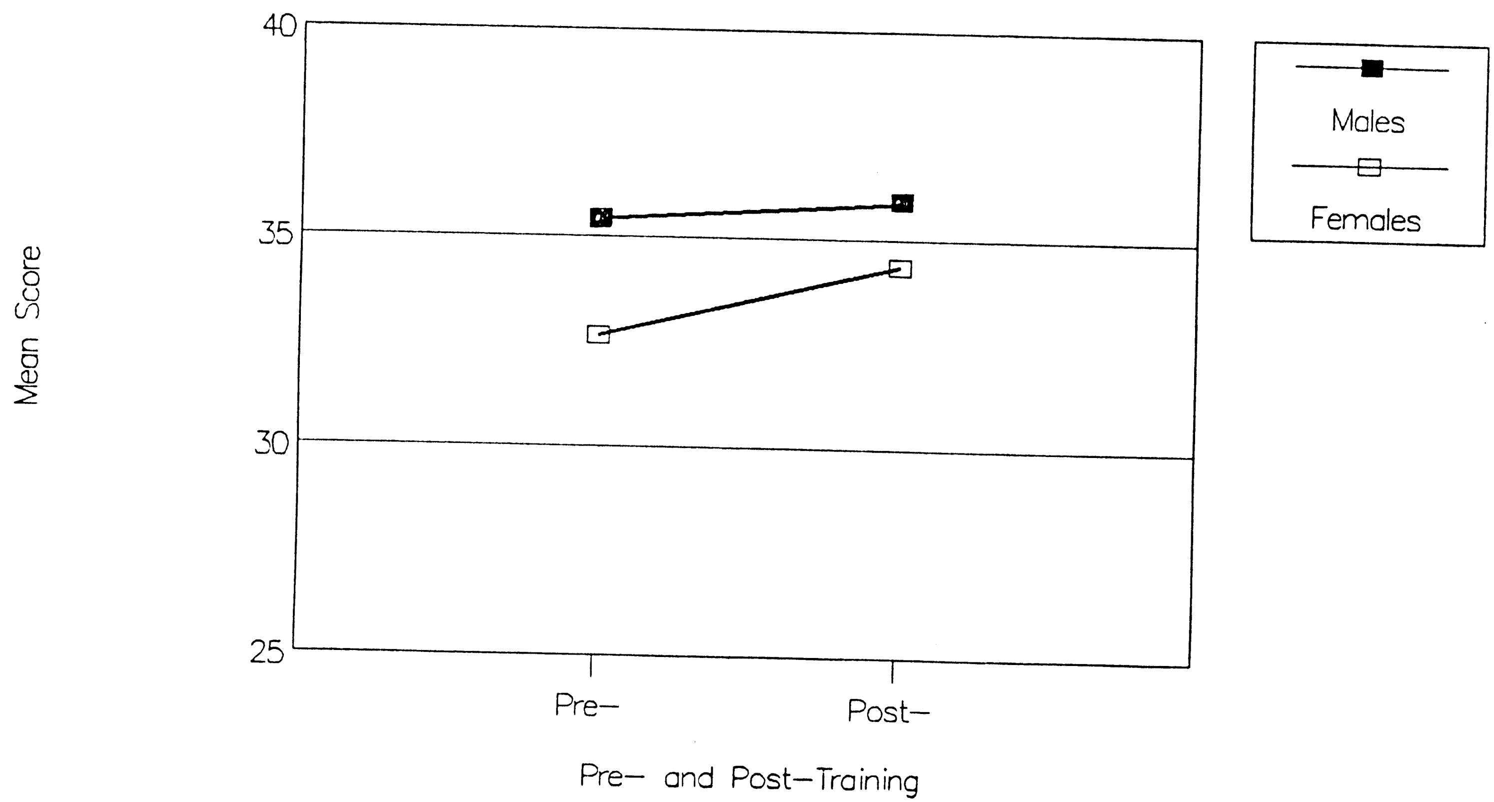


Figure 13. Pre- and Post-Training RTDOT Scores by Sex

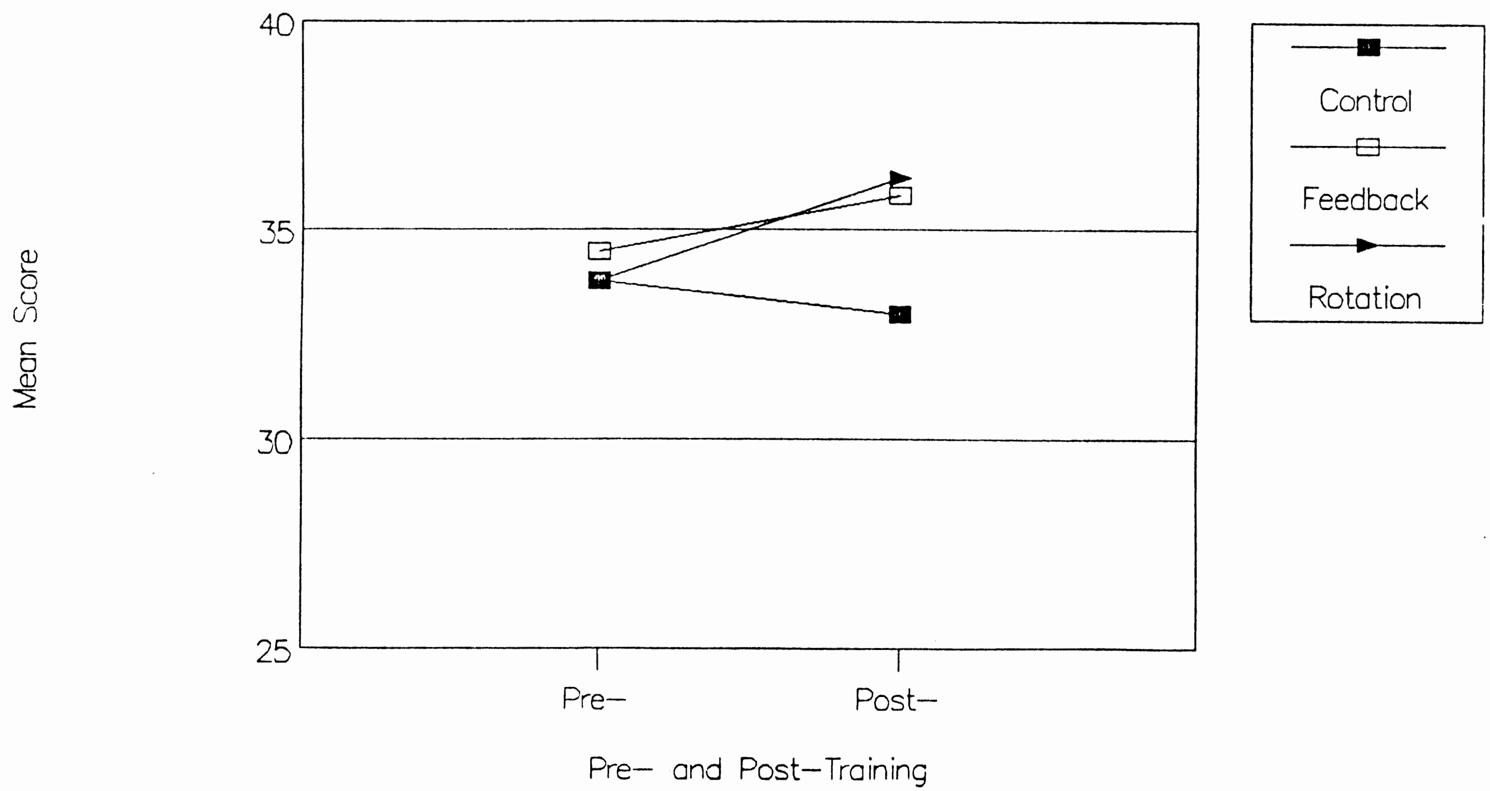


Figure 14. Pre- and Post-Training RTDOT Scores by Experimental Group

TABLE XXV

REPEATED MEASURES ANALYSIS OF VARIANCE ON
PRE- AND POST-TRAINING DPORT SCORES

Source	Sum of Squares	DF	Mean Square	F-Ratio	P
Between Subjects					
Sex	107.249	1	107.249	3.220	.076
Method	101.122	2	50.561	1.518	.224
Sex*Method	114.539	2	57.270	1.720	.184
Error	3330.245	100	33.302		
Within Subjects					
Trials	2661.310	1	2661.310	149.719	.001
Trials*Sex	13.580	1	13.580	.764	.384
Trials*Method	50.532	2	25.266	1.421	.246
Trials*Sex*Method	55.878	2	27.939	1.572	.213
Error	1777.539	100	17.775		

N = 106

Squared Multiple R = .05 on DPORT-C scores and .18 on DPORT-E scores.

training group significantly outperformed the control group [S (2, 100) = 3.40, p < .05] as did the mental rotation training group [S (2, 100) = 3.67, p < .05], and no difference between the two training groups was observed [S (2, 100) = .49, p >.05]. The trials by method interaction is shown in Figure 14.

Table XXV provides the summary ANOVA for DPORT scores, and as indicated, the only significant univariate F statistic was for the trials factor [F (1, 100) = 149.719, p < .01]. The multivariate statistics for the method by trials interaction were also significant [Rao's F statistic for Wilks' Lambda (4, 198) = 4.12, p < .01; Pillai's F approximation (4, 200) = 4.00, p < .01; Hotelling-Lawley F (4,196) = 4.25, p <.01].

Scheffe's S multiple comparison procedure was used to further assess the method by trials interaction. Three contrasts were of interest for each level of the trials factor: control and feedback training, control and mental rotation training, and feedback and mental rotation training. For the first level of the trials factor (i.e., pre-training DPORT scores) differences between treatment group means were insignificant. For post-training DPORT scores, the feedback training group significantly outperformed the control group [S (2, 100) = 3.20, p < .05] as did the mental rotation training group [S (2, 100) = 3.51, p < .05], and no difference between the two training groups was observed [S (2, 100) = .52, p >.05]. The trials by method interaction is shown in Figure 15.

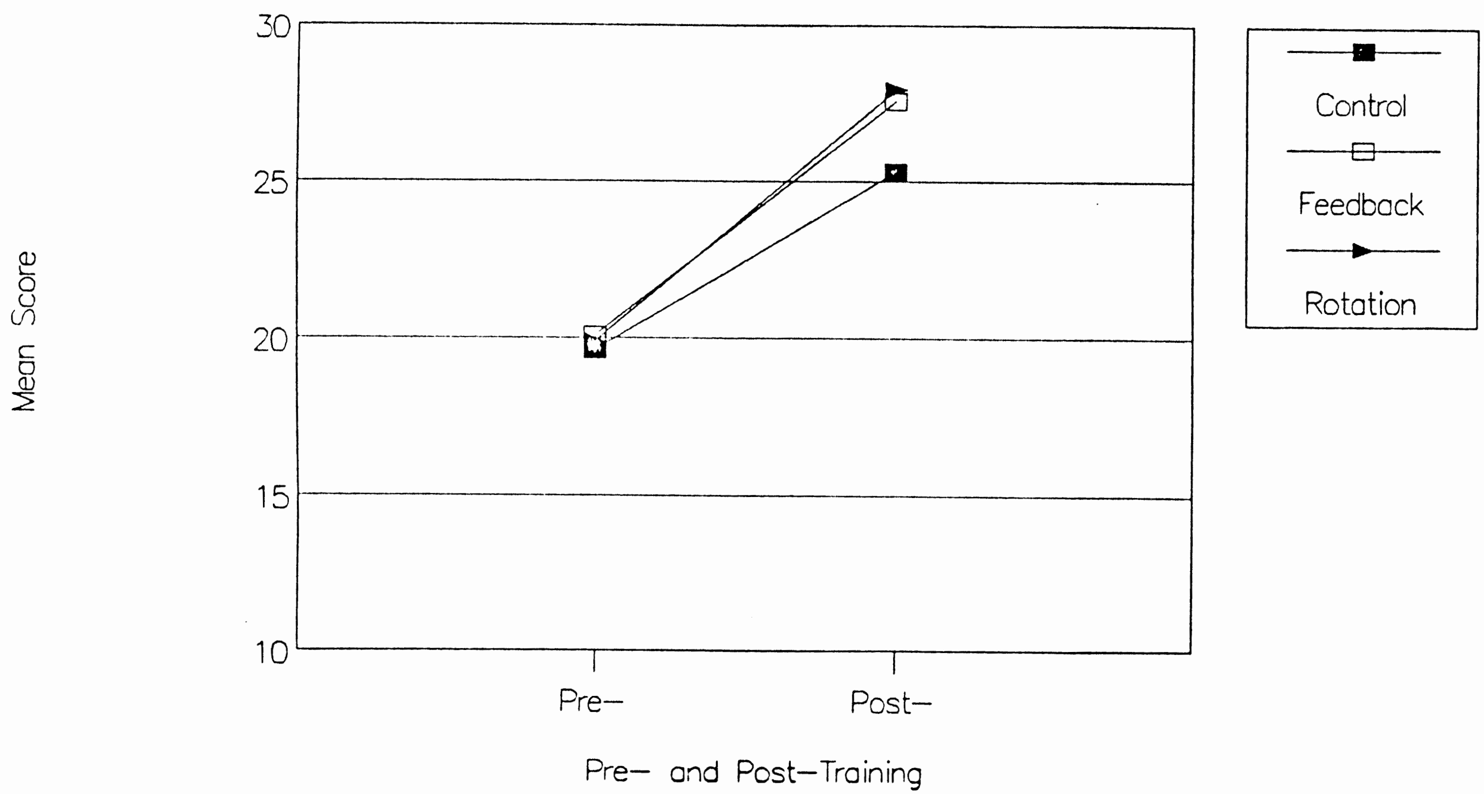


Figure 15. Pre- and Post-Training DPORT Scores by Experimental Group

CHAPTER V

CONCLUSIONS

This chapter summarizes the results and provides commentary. Some remarks regarding the dependent measures are provided first. Later sections consider the training results against the hypotheses posed in Chapter II.

Dependent Measures

Data presented in the Results chapter (see tables XXII, XXIII, XXIV, and XXV) suggest that Mental Rotations, the Rotating Three-Dimensional Objects Test, and the Depth Plane Object Rotation Test were adequate psychometrically. Item-total score correlations, internal consistency reliability estimates, and test intercorrelations were acceptable.

As noted in Chapter IV, however, further examination of test score means and standard deviations indicated that there was a strong ceiling effect on post-training DPORT scores. Post-training performance on the DPORT was so high that the variance on this test was severely attenuated; everyone performed so well that there was little room for individual or group differences to be observed. I suspect that this resulted from the speed component present in the DPORT. Recall that pre-training percent completion rates suggested that both the RTDOT (99 percent completion) and the DPORT (85 percent completion) rely more heavily on power than on speed. The DPORT, however, did receive substantially lower percent completion rates pre-training than the RTDOT, and its item difficulties did suggest that a speed component was present. Post-training nearly all subjects completed this test (99 percent completion). This is a relatively common finding with speeded tests. That is, subjects, having become more test-wise, work faster the second time around and complete more items.

The consequences of the attenuated variance and the ceiling on DPORT scores are two-fold. First, all correlations between pre- and post-training performance are attenuated, and, thus, there is no way to discern whether the low correlations reflect a change of rank ordering of individuals on mental rotation ability or are simply an artifact of low variance. Second, because the post-training test score variance is attenuated, there is little "room," or variance, in which experimental effects can be observed, even if such effects are present; had the DPORT had a higher ceiling, a sex main effect and trial by sex interaction effect might have been observed. Without adequate variance on the post-training scores, it is not possible to draw strong conclusions about experimental treatment effects.

Evaluation of Hypotheses

Hypotheses about sex differences in mental rotation test scores, training effects, differential experience, and transfer of training were presented in Chapter II. Here, each hypothesis is considered against the empirical evidence reported in Chapter IV.

Hypothesis 1: Sex Differences

Before training, males were expected to outperform females on all three dependent measures, and data provided in Chapter IV indicate that males did outscore females. The magnitude of the sex difference on pre-training tests (as measured by the standardized difference between male and female mean scores) was .53 for RTDOT (Form A) and .32 on the DPORT (Form C). The sex difference in Mental Rotation Test scores was relatively large--over one-half standard deviation (.61)--and is consistent with the large difference found in previous studies using this test (Drauden, 1980; Freedman & Rovengno, 1981; McGee, 1978; Vandenberg & Kuse, 1978). As described in Chapter IV, pre-training sex differences were significant for Mental Rotations [F (1, 100) = 10.92, $p < .01$] and the RTDOT [F (1, 100) = 7.16, $p < .05$], and the pretest sex difference on the DPORT did not reach statistical significance.

Given that the RTDOT and DPORT were designed such that the items would be similar in appearance across the two tests and that both tests involve rotation in the depth plane, the smaller sex difference on the DPORT is not likely to be due to the mental processing requirements of the test (e.g., visualization and rotation of the objects pictured). The major differences between the two tests lie in the instructions and the presence of a speed component for the DPORT. Because previous research suggests that females rotate objects at a slower rate than do males (e.g., Kail, et al., 1979, 1984), test speededness might lead us to expect larger, not smaller, sex differences, if items are similar in type and difficulty. Although the instructions for the two tests differ, there is no clear rationale for how the difference in instructions might affect the magnitude of sex differences.

Hypotheses 2 and 3: General and Specific Training Effects

Recall that two hypotheses about general and specific training effects were presented in Chapter II. At a general level, trained groups were expected to show greater improvement than untrained groups, post-training, and with regard to specific training, training designed to provide direct experience rotating three-dimensional objects (i.e., mental rotation training) was expected to yield greater performance improvement than feedback training.

The significant interaction between training methods and trials (pre- and post-training) on Rotating Three Dimensional Objects (RTDOT) and Depth Plane Object Rotation Test (DPORT) scores suggests that training did enhance scores on these measures over the improvement of the control group. More specifically, Scheffe's S multiple comparison procedure indicated that the feedback training group and the mental rotation training group outperformed the control group, post-training, on the DPORT and the RTDOT. These results are, thus, consistent with previous work finding improvement of trained groups relative to the control (e.g., Brinkmann, 1966, Embretson, 1988; Lohman, 1988; Stringer, 1975).

There were no significant differences between the performance of the two trained groups (i.e., feedback and mental rotation) on any of the dependent measures. These data, therefore, do not support the hypothesis that mental rotation training would, by providing experience rotating tangible objects, result in greater improvement than feedback training. Moreover, the data suggest that increased familiarity with the testing situation and with test items and instructions are as influential in test performance improvement as is a training method designed to "teach" a construct.

It could also be argued that two one-hour training intervention periods are insufficient to permit observation of effects due to learning about the construct over and above those due to practice on test items. That is, in a lengthier training intervention, effects due to familiarity with items, or uninstructed practice on a test, might reach an asymptote, and construct oriented training might provide greater improvement, in the long run, by facilitating performance beyond the asymptote of the feedback only intervention.

Hypothesis 4: Differential Learning

The differential experience hypothesis suggests that males are functioning closer to the asymptote of their spatial ability than are females, and consequentially, that females should respond more favorably to practice than males. Scores of females should, therefore, show greater performance improvement pre- to post-training than those of males. A significant interaction between sex and trials (i.e., pre- to post-training) indicates that females respond more favorably to males across testing sessions, regardless of experimental group (i.e., control, feedback training, or mental rotation training) and, thus, supports the differential experience hypothesis. A significant interaction between sex, experimental group, and trials should emerge if the the training procedures play an important role in facilitating differential improvement by females.

The data presented in Chapter IV provide some support for the hypothesis of differential improvement of females compared to males. Particularly, the significant sex by trials interactions for both Mental Rotations and RTDOT scores suggests that females showed greater improvement than males across the two testing sessions on these dependent measures, regardless of the training method employed. Moreover, in both instances, the initial sex difference was significant (pre-test) and the post-test sex difference was not. These data support previous findings of differential improvement by females (e.g., Blatter, 1983; Connor et al, 1977, 1978). Lack of a significant sex difference pre-training in DPORT scores and lack of variance in post-training DPORT scores precluded the same observation on this measure.

The sex by method by trials interaction did not reach significance on any of the dependent measures. It, therefore, does not appear that females responded to training with greater performance improvement than males. In sum, the training procedures had little or no impact on differential improvement. Thus, the training procedures were not important in facilitating differential improvement; this finding is similar to that reported by Drauden (1980).

Hypothesis 5: Transfer of Training

Recall that the dependent measures were constructed with different degrees of training-relatedness in mind. That is, the Depth Plane Object Rotation Test (DPORT) was linked very closely with the training measures. The Rotating Three Dimensional Objects Test (RTDOT) was not a part of the training, but its items are similar in appearance to those on the DPORT. The Mental Rotations Test was similar to the other two dependent measures in that it is a three-dimensional rotation task; its items are quite different, in physical appearance, from those on the other tests, and it was not used in training procedures.

Although training effects were observed for the DPORT and the RTDOT (i.e., there was a significant trials by method interaction), no method effects were evident on Mental Rotations test scores. These data suggest that training on the DPORT did generalize to performance on the RTDOT and did not generalize to performance on Mental Rotations. These findings are consistent with those reported by Levine et al. (1979) who found significant improvement of trained groups on criterion measures highly related to training, but not on those less directly related to the intervention.

The lack of a significant improvement by trained groups over the control for the Mental Rotations test implies that either the training intervention was not strong enough to permit generalization to this test or that the effects of training on one test are so test-specific (having to do with instructions and item types) that it is unlikely generalization would occur. Because RTDOT instructions and items are different from those on the DPORT, it is unlikely that simple familiarity with instructions on the DPORT was related to improved RTDOT performance by the trained groups. It, therefore, appears that the training intervention was probably not intense or long enough to facilitate performance improvement on Mental Rotations.

Summary

Key conclusions are summarized below:

- The data support the idea that training enhances performance on mental rotation tests.
- The hypothesis that mental rotation training would result in greater performance improvement than feedback training was not supported. A lengthier training intervention may provide a better test of this hypothesis.
- The data suggest that there is some generalization of performance improvement on one test to performance on another test. A lengthier or more intense training intervention may have resulted in generalization of performance improvement to the Mental Rotations test as well as the RTDOT.
- Although there is some support for the idea that females would show greater improvement than males across trials, there is no evidence that females respond to training with greater performance improvement than do males.

Toward the Future

There are still many unanswered questions with regard to the trainability of mental rotation skills and the observed sex differences in mental rotation performance. Are training effects long-lived or do they deteriorate with time? Does a training intervention result in changes in the rank order of individuals on the trait or does it simply increase the magnitude of all scores? Are decreases in the magnitude of sex differences in spatial ability long lasting; are they test specific? Future work in this area should employ difficult tests that are not highly speeded to avoid problems associated with a ceiling effect and should provide a strong training intervention over time to examine the asymptote of trainability of spatial ability.

BIBLIOGRAPHY

Alderton, D. L., & Pellegrino, J. W. (1985). Sex differences in spatial ability: Componential analyses of process differences. Paper presented at the annual meeting of the American Psychological Association, Los Angeles, CA.

Alexander, W. P. (1935). Intelligence, concrete and abstract. British Journal of Psychology Monograph Supplement, 29.

Allen, M. J. (1974). Sex differences in spatial problem-solving styles. Perceptual and Motor Skills, 39, 843-846.

Allen, M. J., & Hogeland, R. (1978). Spatial problem-solving strategies as functions of sex. Perceptual and Motor Skills, 47, 348-350.

American College Testing (ACT) Program. (1977). Handbook for the ACT Career Planning Program. Iowa City, IA.

Anastasi, A. (1958). Differential psychology: Individual and group differences in behavior (3rd ed.). New York: MacMillan.

Begley, S. (April 11, 1988). Closing the gender gap: Equality of test scores. Newsweek, 73.

Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. Science, 222, 1029-1031.

Bennett, G. K., & Cruikshank, R. M. (1942). Sex differences in the understanding of mechanical problems. Journal of Applied Psychology, 26, 121-127.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1974). Manual for the differential aptitude tests: Forms S and T (5th ed.). New York: The Psychological Corporation.

Blade, M. F., & Watson, W. S. (1955). Increase in spatial visualization test scores during engineering study. Psychological Monographs: General and Applied, 69(12), 1-13.

Blatter, P. (1983). Training in spatial ability: A test of Sherman's hypothesis. Perceptual and Motor Skills, 57, 987-992.

Bock, R. D., & Kolakowski, D. (1973). Further evidence of a sex-linked major-gene influence on human spatial visualizing ability. American Journal of Human Genetics, 25, 1-14.

Boles, D. B. (1980). X-linkage of spatial ability: A critical review. Child Development, 51, 625-635.

Bouchard, T. J., & McGee, M. G. (1977). Sex differences in human spatial ability: Not an X-linked recessive gene effect. Social Biology, 24, 332-335.

Box, G. E. P. (1954).Some theorems on quadratic forms applied in analysis of variance problems. The Annals of Mathematical Statistics, 25, 290-302, 484-498.

Brinkmann, E. H. (1966). Programmed instruction as a technique for improving spatial visualization. Journal of Applied Psychology, 50, 179-184.

Broverman, D. M., Klaiber, E. L., Kobayashi, Y., & Vogel, W. (1968). Roles of activation and inhibition in sex differences in cognitive abilities. Psychological Review, 75, 23-50.

Brown, F. R. (1954). The effect of an experimental course in geometry on ability to visualize in three dimensions. Unpublished doctoral dissertation, Columbia University.

Brown, W., & Stephenson, W. A. (1933). A test of the theory of two factors. British Journal of Psychology, 23, 352-370.

Bryden, M. P. (1979). Evidence for sex-related differences in cerebral organization. In M. A. Wittig & A. C. Petersen (Eds.), Sex-related differences in cognitive functioning. New York: Academic Press.

Buffery, A. W. H., & Gray, J. A. (1972). Sex differences in the development of spatial and linguistic skills. In C. Ounstead and D. C. Taylor (Eds.), Gender differences: Their ontogeny and significance. London: Churchill.

Burnett, S. A. (1986). Sex differences in spatial ability: Are they trivial? American Psychologist, 41, 1012-1014.

Burnett, S. A., Lane, D. M., & Dratt, L. M. (1979). Spatial visualization and sex differences in quantitative ability. Intelligence, 3, 345-354.

Burnett, S. A., Lane, D. M., & Dratt, L. M. (1982). Spatial ability and handedness. Intelligence, 6, 57-68.

Campbell, J. P. (1976). Psychometric theory. In M.D. Dunnette (Ed.), Handbook of Industrial and Organizational Psychology, 185-222.

Caplan, P. J., MacPherson, G. M., & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? A multilevel critique with new data. American Psychologist, 40, 786-799.

Carter, P., Pazak, B., & Kail, R. (1983). Algorithms for processing spatial information. Journal of Experimental Child Psychology, 36, 284-304.

Churchill, B. D., Curtis, J. M., Coombs, C. H., & Hassell, T. W. (1942). Effect of engineer school training on the Surface Development Test. Educational Psychological Measurement, 2, 279-280.

Cohen, D. (1976). Usefulness of the group-comparison method to demonstrate sex differences in spatial orientation and spatial visualization in older men and women. Perceptual and Motor Skills, 43, 388-390.

Cole, N. S. (1981). Bias in testing. American Psychologist, 36, 1067-1077.

Connor, J. M., Schackman, M., & Serbin, L. A. (1978). Sex-related differences in response to practice on a visual-spatial test and generalization to a related test. Child Development, 49, 24-29.

Connor, J. M., Serbin, L. A., & Schackman, M. (1977). Sex differences in children's response to training on a visual-spatial test. Developmental Psychology, 13, 293-294.

Cooper, L. A., & Shepard, R. W. (1975). Mental transformations in the identification of left and right hands. Journal of Experimental Psychology: Human Perception and Performance, 104, 48-56.

Cox, J. W. (1928). Mechanical aptitude. London: Meuthuen.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52(4), 281-302.

Cureton, E. E. (1971). The stability coefficient. Educational and Psychological Measurement, 31, 45-55.

DeFries, J. C., Ashton, G. C., Johnson, R. C., Kuse, A. R., McClearn, G. E., Mi, M. P., Rashad, M. N., Vandenberg, S. G., & Wilson, J. R. (1976). Parent-offspring resemblance for specific cognitive abilities in two ethnic groups, Nature, 261, 131-133.

Denmark, F. L., & Francois, F. (1987). Research context of studies for sex differences in mathematics performance. Psychology and Educational Policy, 517, 61-68.

Donlon, T. F. (1987). Issues in the construction and use of tests in studies of sex differences in mathematics. Psychology and Educational Policy, 517, 87-98.

Drauden, G. (1980). Training effects on sex differences in spatial abilities. Unpublished doctoral dissertation, University of Minnesota.

Dunnette, M. D., Corpe, V. A., & Toquam, J. L. (1987). Cognitive paper-and-pencil measures: Field test. In N. G. Peterson (Ed.), Development and field test of the trial battery for Project A (Technical Report No. 739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Earle, F. M. (1929). Tests of mechanical ability (p. 321). NIIP.

Ehrlichman, H. I. (1972). Hemispheric functioning and individual differences in cognitive abilities (Doctoral dissertation, New School for Social Research). Dissertation Abstracts International, 33, 2319B.

Ekstrom, R. B. (1973). Cognitive factors: Some recent literature (Technical Report No. 2, ONR Contract N00014-71-C-0017, NR 150-329.). Princeton, NJ: Educational Testing Service.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. Multivariate Behavioral Research Monographs, 79-2, P. 1-84.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). Manual for kit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service.

Eliot, J. (1980). Classification of figural spatial tests. Perceptual and Motor Skills, 51, 847-851.

Eliot, J. (1986). Comment on Caplan, MacPherson, and Tobin. American Psychologist, 41, 1011.

Embretson, S. E. (1987). Improving the measurement of spatial aptitude by dynamic testing. Intelligence, 11, 333-358.

Eysenk, H. J. (1939). Review of "primary mental abilities" by L. L. Thurstone. British Journal of Psychology, 9, 270-275.

Faubian, R. W., Cleveland, E. A., & Hassell, T. W. (1942). The influence of training on mechanical aptitude test scores. Educational Psychological Measurement, 2, 91-94.

Feingold, A. (1988). Cognitive gender differences are disappearing. American Psychologist, 43, 95-103.

Ferguson, G. A. (1954). On learning and human ability. Canadian Journal of Psychology, 8, 95-112.

Ferguson, G. A. (1956). On transfer and the abilities of man. Canadian Journal of Psychology, 10, 121-131.

Fox, L. H. (1987). Sex differences among the mathematically gifted. Psychology and Educational Policy, 517, 99-112.

Freedman, R. J., & Rovengno, L. (1981). Ocular dominance, cognitive strategy, and sex differences in spatial ability. Perceptual and Motor Skills, 52, 651-654.

French, J. W. (1951). The description of aptitude and achievement tests in terms of rotated factors. Psychometric Monographs, 5. Chicago: University of Chicago Press.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). Kit of reference tests for cognitive factors. Princeton, NJ: Educational Testing Service.

Ghiselli, E. E. (1966). The validity of occupational aptitude tests. New York: Wiley.

Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. Personnel Psychology, 26, 461-477.

Goldstein, A. G., & Chance, J. E. (1965). Effects of practice on sex-related differences in performance on Embedded Figures. Psychonomic Science, 3, 361-362.

Guilford, J. P. (1967). The nature of human intelligence. New York: McGraw-Hill.

Guilford, J. P., & Lacy, J. I. (1947). Printed Classification Tests, A.A.F., Aviation Psychological Progress Research Report, 5. Washington, DC: U.S. Government Printing Office.

Guilford, J. P., & Zimmerman, W. S. (1947). Some A.A.F. findings concerning aptitude factors. Occupations, 26, 154-159.

Gustaffson, J. E. (1984). A unifying model for the structure of intellectual abilities. Intelligence, 8, 179-203.

Halpern, D. F. (1986). A different answer to the question, "Do sex-related differences in spatial abilities exist?". American Psychologist, 41, 1014-1015.

Halpern, D. F. (1987). Sex differences in cognitive abilities. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hamley, H. R. (1935). The testing of intelligence. London: Evans.

Harris, L. J. (1978). Sex differences in spatial ability: Possible environmental, genetic, and neurological factors. In M. Kinsbourne (Ed.), Asymmetrical function of the brain. New York: Cambridge University Press.

Harris, L. J. (1981). Sex-related variations in spatial skills. In Spatial representation and behavior across the lifespan. Academic Press Inc.

Hartlage, L. C. (1970). Sex-linked inheritance of spatial ability. Perceptual and Motor Skills, 31, 610.

Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. Journal of Educational Statistics, 7, 119-137.

Hills, J. R. (1957). Factor analyzed abilities and success in college mathematics. Educational Psychological Measurement, 17, 615-622.

Hisock, M. (1986). On sex differences in spatial abilities. American Psychologist, 41, 1011-1012.

Hobson, J. R. (1947). Sex differences in primary mental abilities. Journal of Educational Research, 41, 126-132.

Hoffman, K. I., Guilford, J. P., Hoepfner, R., & Doherty, W. J. (1968). A factor analysis of the figural-cognition and figural-evaluation abilities. University of Southern California Psychological Laboratory (Rep. No. 40), Los Angeles.

Hogrebe, M. C. (1987). Gender differences in mathematics. American Psychologist, 42, 265-266.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligence. Journal of Educational Psychology, 57, 253-270.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). Meta analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage.

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using w^2 and d. American Psychologist, 36, 892-901.

Johnson, E. S. (1984). Sex differences in problem solving. Journal of Educational Psychology, 76, 1359-1371.

Johnson, S. (1976). _Effect of practice and training in spatial skills on sex related differences in performance on Embedded Figures_. Unpublished master's thesis, George Mason University, Fairfax, VA.

Johnson, S., Flinn, J. M., & Tyer, Z. E. (1979). Effects of practice and training in spatial skills on Embedded Figures scores of males and females. _Perceptual and Motor Skills_, _48_, 975-984.

Jones, B., & Anuza, T. (1982). Effects of sex, handedness, stimulus, and visual field on Mental Rotation. _Cortex_, _18_, 501-514.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. _Psychological Review_, _92_, 137-172.

Kail, R. V., Carter, P., & Pellegrino, J. W. (1979). The locus of sex differences in spatial ability. _Perceptual and Psychophysics_, _26_, 182-186.

Kail, R. V., Stevenson, M. R., & Black, K. N. (1984). Absence of a sex difference in algorithms for spatial problem solving. _Intelligence_, _8_, 37-46.

Kaplan, R. M., & Saccuzzo, D. P. (1982). _Psychological Testing: Principles, Applications, and Issues_. Monterey, CA: Brooks/Cole Publishing.

Kelley, T. L. (1928). _Crossroads in the mind_. Stanford, CA: Stanford University Press.

Kepner, M. D., & Neimark, E. D. (1984). Test-retest reliability and differential patterns of score change on the group Embedded Figures Test. _Journal of Personality and Social Psychology_, _46_, 1405-1413.

Koussey, A. A. H., El. (1935). The visual perception of space. _British Journal of Psychology Monograph Supplement_, _20_.

Landsdell, H. (1962). A sex difference in effect of temporal lobe neurosurgery on design preference. _Nature_, _194_, 852-854.

Levine, J. M., Brahlek, R. E., Eisner, E. J., & Fleishman, E. A. (1979). _Trainability of abilities: Training and transfer of abilities related to electronic fault-finding_ (ARRO Technical Report, 3010 R79-2). Washington, DC: Advanced Research Resources Organization.

Levine, J. M., Schulman, D., Brahlek, R. E., & Fleishman, E. A. (1980). Trainability of abilities: Training and transfer of spatial visualization (ARRO Technical Report, 3010 TR3). Washington, DC: Advanced Research Resources Organization.

Levy, J. (1969). Possible basis for the evolution of lateral specialization of the human brain. Nature, 224, 614-615.

Levy, J. (1976). Cerebral lateralization and spatial ability. Behavior Genetics, 6, 171-189.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child Development, 56, 1479-1498.

Lohman, D. F. (1979). Spatial ability: A review and reanalysis of the correlational literature (Aptitude Research Project Technical Report No. 8). Stanford, CA: Stanford University.

Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.) Advances in the psychology of human intelligence (Vol. 4, pp. 181-248). Hillsdale, NJ: Lawrence Erlbaum Associates.

Maccoby, E. E., & Jacklin, C. N. (1974). The psychology of sex differences. Stanford, CA: Stanford University Press.

McFarlane, M. (1925). A study of practical ability. British Journal of Psychology Monograph Supplement, 8.

McGee, M. G. (1976). Laterality, hand preference, and human spatial ability. Perceptual and Motor Skills, 42, 781-782.

McGee, M. G. (1977). A family study of human spatial ability. (Doctoral dissertation, University of Minnesota). Dissertation Abstracts International, 37, 6396.

McGee, M. G. (1978a). Effects of two problem solving strategies on Mental Rotation Test scores. Journal of Psychology, 100, 83-85.

McGee, M. G. (1978b). Effects of training and practice on sex differences in Mental Rotation Test scores. Journal of Psychology, 100, 87-90.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. Psychological Bulletin, 86, 889-918.

McGlone, J. (1980). Sex differences in human brain asymmetry: A critical survey. The Behavioral and Brain Sciences, 3, 215-263.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1987, April). Project A validity results: The relationship between predictor and criterion domains. Paper presented at the mid-year conference of the Society for Industrial and Organizational Psychology in Atlanta, GA.

McKenna, F. P. (1984). Measures of field dependence: Cognitive style or cognitive ability? Journal of Personality and Social Psychology, 47, 593-603.

Michael, W. B., Guilford, J. P., Fruchter, B., & Zimmerman, W. S. (1957). The description of spatial-visualization abilities. Educational and Psychological Measurement, 17, 185-199.

Mumaw, R. J., Pellegrino, J. W., & Kail, R. V. (1984). Different slopes for different folks: Process analysis of spatial aptitude. Memory and Cognition, 12, 515-521.

Murphy, L. W. (1936). The relation between mechanical ability tests and verbal and non-verbal intelligence tests. Journal of Psychology, 2, 353-366.

Newcombe, N. (1982). Sex-related differences in spatial ability. In M. Potegal (Ed.), Spatial abilties: Development and physiological foundations. New York: Academic Press.

Newcombe, N., Bandura, M., & Taylor, D. G. (1983). Sex differences in spatial ability and spatial activities. Sex Roles, 9, 377-386.

Park, J., Johnson, R. C., DeFries, J. C., McClearn, G. E., Mi, M. P., Rashad, M. N., Vandenberg, S. G., & Wilson, J. R. (1978). Parent-offspring resemblance for specific cognitive abilities in Korea. Behavior Genetics, 8, 43-53.

Paterson, D. G., Elliott, R. M., Anderson, L. D., Toops, H. A., & Heidreder, E. (1930). Minnesota mechanical ability tests. Minneapolis: University of Minnesota Press.

Pawlik, K. (1966). Concepts in human cognition and aptitudes. In R. B. Cattell (Ed.), Handbook of multivariate experimental psychology. Chicago: Rand McNally.

Peterson, A. C., & Crockett, L. J. (1987). Biological correlates of spatial ability and mathematical performance. Psychology and Educational Policy, 517, 69-86.

Peterson, N. G. (1987). Development and field test of the trial battery for Project A (Technical Report No. 739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Peterson, N. G., & Bownas, D. A. (1982). Skill, task structure, and performance acquisition. In E. A. Fleishman and M. D. Dunnette, (Eds.), Human performance and productivity: Human capability assessment. Hillsdale, NJ: Lawrence Erlbaum Associates.

Porteus, S. D. (1965). Porteus maze tests: Fifty years' application. Palo Alto, CA: Pacific Books.

Ranucci, E. R. (1952). The effect of the study of solid geometry on certain aspects of space perception abilities. Unpublished doctoral dissertation, Columbia University.

Regian, J. W., & Pellegrino, J. W. (1984). Practice and transfer effects in two-dimensional mental rotation. Unpublished manuscript, University of California, Santa Barbara, CA.

Rossi, J. S. (1983). Ratios exaggerate gender differences in mathematical ability. American Psychologists, 38, 348.

Samuel, W. (1983). Sex differences in spatial ability reflected in performance in IQ subtests by black or white examiners. Personality and Individual Differences, 4, 219-221.

Sanders, B., Cohen, M. R., & Soares, M. P. (1986). The sex difference in spatial ability: A rejoiner. American Psychologist, 41, 1015-1016.

Sanders, B., Soares, M. P., & D'Aquila, J. M. (1982). The sex difference on one test of spatial visualization: A non-trivial difference. Child Development, 53, 1106-1110.

Scarr, S. (1988). Race and gender as psychological variables: Social and ethical issues. American Psychologist, 43, 56-59.

Sevy, B. A. (1983). Sex-related differences in spatial ability: The effects of practice. Doctoral Dissertation, University of Minnesota.

Shepard, R. N., & Feng, C. A. (1972). A chronometric study of mental paper folding. Cognitive Psychology, 3, 228-243.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. Science, 171, 701-703.

Sherman, J. A. (1967). Problem of sex differences in space perception and aspects of intellectual functioning. Psychological Review, 74, 290-299.

Sherman, J. (1980). Mathematics, spatial visualization, and related factors: Changes in girls and boys, grades 8-11. Journal of Educational Psychology, 72, 476-482.

Shute, V. J., Pellegrino, J. W., Hubert, L., & Reynolds, R. W. (1983). The relationship between androgen levels and human spatial abilities. Bulletin of the Psychonomic Society, 21, 465-468.

Smith, I. M. (1948). Measuring spatial abilities in school pupils. Occupational Psychology, 22, 150-159.

Smith, I. M. (1964). Spatial ability: Its educational and social significance. London: University of London.

Spearman, C. (1904). The proof and measurement of the association between two things. American Journal of Psychology, 15, 72-101.

Spearman, C. (1927). The abilities of man. London: Macmillan.

Spearman, C. (1939). Thurstone's work reworked. Journal of Educational Psychology, 30, 1-16.

Stafford, R. E. (1961). Sex differences in spatial visualization as evidence of sex-linked inheritance. Perceptual and Motor Skills, 13, 428.

Stanley, J. C., & Benbow, C. P. (1982). Huge sex ratios at the upper end. American Psychologist, 37, 972.

Stringer, P. (1975). Drawing training and spatial ability. Ergonomics, 18, 101-108.

Sullivan, A. M. (1964). The relationship between intelligence and transfer. Unpublished doctoral dissertation, McGill University, Montreal.

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, 122-130.

Teegarden, L. (1942). Manipulative performance of young adult applicants at a public employment office: Part II. Journal of Applied Psychology, 26, 754-769.

Thomas, H. (1983). Familial correlation analyses, sex differences, and the X-linked gene hypothesis. Psychological Bulletin, 93, 427-440.

Thurstone, L. L. (1938). Primary mental abilities. Chicago: University of Chicago Press.

Thurstone, L. L. (1950). Some primary abilities in visual thinking. University of Chicago Psychometric Laboratory (Rep. No. 59): Chicago.

Thurstone, L. L., & Thurstone, T. G. (1941)., The Primary Mental Abilities Tests. Chicago: Science Research Associates.

Tobin, P. (1982). The effects of practice and training on sex differences in performance on a spatial task. Unpublished master's thesis, University of Toronto, Toronto, Canada.

Toquam, J. L., Dunnette, M. D., Corpe, V. A., Houston, J. S., Peterson, N. G., Russell, T. L., & Hanson, M. A. (1987). Cognitive paper-and-pencil measures: Pilot testing. In N. G. Peterson (Ed.), Development and field test of the trial battery for Project A (Technical Report No. 739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Tyler, L. (1965). The psychology of human differences (3rd edition). New York: Appleton-Century-Crofts.

U.S. Civil Service Commission, U.S. Equal Employment Opportunity Commission, U.S. Department of Justice, and U.S. Department of Labor. (1978). Uniform guidelines on employee selection procedures. Federal Register, 43(166), 38295-38309.

Vandenberg, S. G. (1975). Sources of variance in performance on spatial tests. In J. Eliot and N. J. Salkind (Eds.), Children's spatial development. Springfield, IL: Charles C. Thomas.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental Rotations, a group test of three-dimensional spatial visualization. Perceptual and Motor Skills, 47, 599-604.

Van Voorhis, W. R. (1941). The improvement of space perception ability by training. Unpublished doctoral dissertation, Pennsylvania State University.

Vernon, P. E. (1950). The structure of human abilities. London: Meuthuen.

Vernon, P. E. (1960). Intelligence and attainment tests. London: University of London Press.

Vernon, P. E. (1965). Ability factors and environmental influences. American Psychologist, 20, 723-733.

Vernon, P. E. (1969). Intelligence and cultural environment. London: Meuthuen.

Waber, P. P., Carlson, D., & Mann, M. (1982). Developmental and differential aspects of mental rotation in early adolescence. Child Development, 53, 1614-1621.

Williams, T. (1975). Family resemblance in abilities: The Weschler scales. Behavior Genetics, 5, 405-409.

Wilson, J. R., & Vandenberg, S. G. (1978). Sex differences in cognition: Evidence from the Hawaii family study. In T. E. McGill, D. A. Dewsbury, and B. J. Sachs (Eds.), Sex and behavior. Plenum Publishing Corporation.

Wilson, J. R., DeFries, J. C., McClearn, G. E., Vandenberg, S. G., Johnson, R. C., & Rashad, M. N. (1975). Cognitive abilities: Use of family data as a control to access sex and age differences in two ethnic groups. International Journal of Aging and Human Development, 6, 261-276.

Winer, B. J. (1971). Statistical Principles in Experimental Design. New York: McGraw-Hill.

Wise, L. L., Campbell, J., & Peterson, N. G. (1987, April). Identifying optimal predictor composites and testing for generalizability across jobs and performance constructs. Paper presented at the mid-year conference of the Society for Industrial and Organizational Psychology in Atlanta, GA.

Witkin, H. A. (1950). Individual differences in ease of perception of Embedded Figures. Journal of Personality, 19, 1-15.

Witkin, H. A., Dyk, R. A., Faterson, G. E., Goodenough, D. R., & Karp, S. A. (1962). Psychological differentiation. New York: Wiley.

Witkin, H. A., Goodenough, D. R., & Karp, S. A. (1967). Stability of cognitive style from childhood to young adulthood. Journal of Personality and Social Psychology, 7, 291-300.

Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, K., Meissner, P., & Wapner, S. (1954). Personality through perception. New York: Harper.

Wittig, M. A., & Petersen, A. C. (Eds). (1979). Sex-related differences in cognitive functioning. New York: Academic Press.

Yen, W. M. (1975). Sex-linked major-gene influence in selected types of spatial performance. Behavior Genetics, 5, 281-298.

Zajanc, R. B. (1986). The decline and rise of scholastic aptitude scores. American Psychologist, 41, 862-867.

Zimmerman, W. S. (1953). A revised orthogonal rotation solution for Thurstone's original primary mental abilities test battery. Psychometrika, 18, 77-93.

Zimmerman, W. S. (1954a). The influence of item complexity upon the factor composition of a spatial visualization test. Educational and Psychological Measurement, 14, 106-119.

Zimmerman, W. S. (1954b). Hypotheses concerning the nature of spatial factors. Educational and Psychological Measurement, 14, 396-400.

APPENDIX A

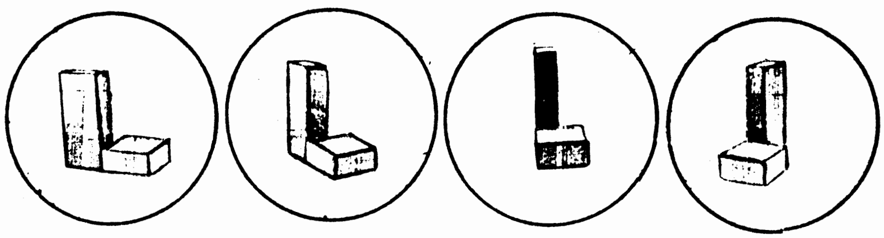
INSTRUCTIONS FOR THE ROTATING THREE-DIMENSIONAL OBJECTS TEST AND THE DEPTH PLANE OBJECT ROTATION TEST

Rotating Three-Dimensional Objects Test
(RTDOT) Form A

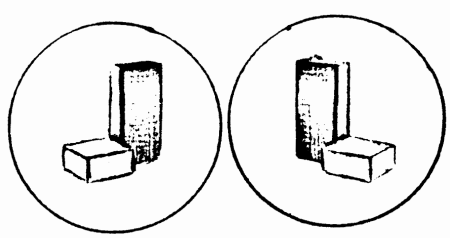
This is a test of your ability to imagine what a three-dimensional object will look like if it is rotated in the depth plane.

For each problem, there will be pictures of two three-dimensional objects. You are to decide whether the pictures are actually two views of the same object or whether the objects are different from each other. That is, the objects are different if there is no way to rotate one of the objects so that it looks exactly like the other view. If one of the objects can be rotated to look exactly like the other, the objects are the same.

For example, look at the series of pictures below. They show different views of one object as it is being rotated in the depth plane. Can you see that the pictures are all actually different views of the same object? Is the object rotating clockwise or counter-clockwise?

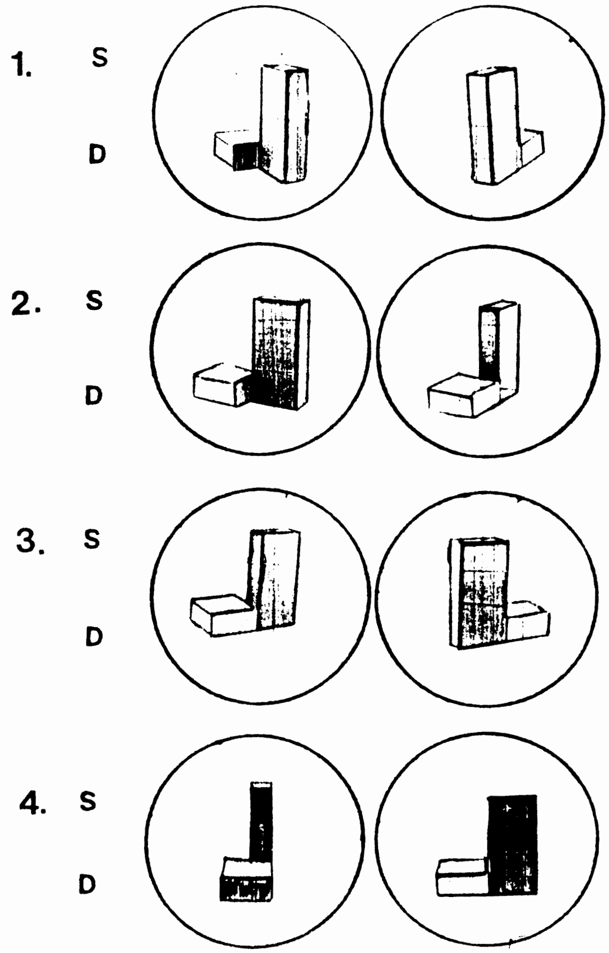


In the four pictures above, the object is rotating clockwise. Now, look at the two objects below. They are three-dimensional reflections or mirror-images of each other. Try rotating the object on the right. No matter how it is rotated it will not look exactly like the object on the left. The two objects are, therefore, different.



The items on this test are composed of actual pictures of objects. Sometimes, shadows are present. Please remember that the objects are always either exactly the same or mirror images.

Now, work the sample problems on the next page. Circle the **S** on the answer sheet, if the objects are the same. Circle the **D** if the objects are different because they are mirror-images of each other. Do not write on the test booklet. Please write all you answers on the answer sheet.



Imagine that the object on the left in item number 1 has been rotated clockwise about 60 degrees. It will look exactly like the object on the right in item number 1. Thus, the objects in problem number 1 are the <u>same</u>, and **S** is the correct answer. The answers to the remaining problems are: #2 is **D**; #3 is **D**; and #4 is **S**.

Did you get the correct answers? If not, feel free to ask for assistance.

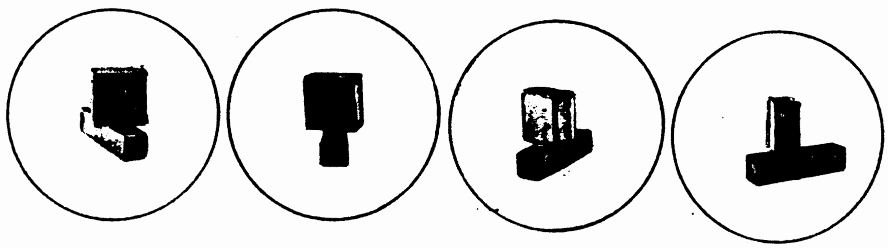
There are 40 items on this test, and you will have seven minutes. Please work rapidly and accurately. Remember to write all your answers on the answer sheet. Do not write on the test booklet.

DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

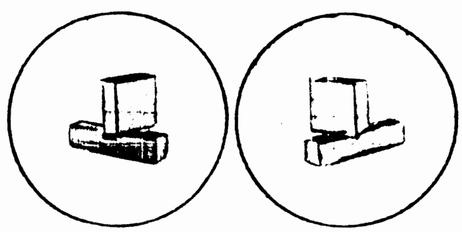
Depth Plane Object Rotation Test
(DPORT) Form E

This is a test of your ability to imagine what a three-dimensional object will look like if it is rotated in the depth plane.

Look at the series of pictures below. They show different views of one object as it is being rotated in the depth plane. Can you see that the pictures are all actually different views of the same object? Is the object rotating clockwise or counter-clockwise?

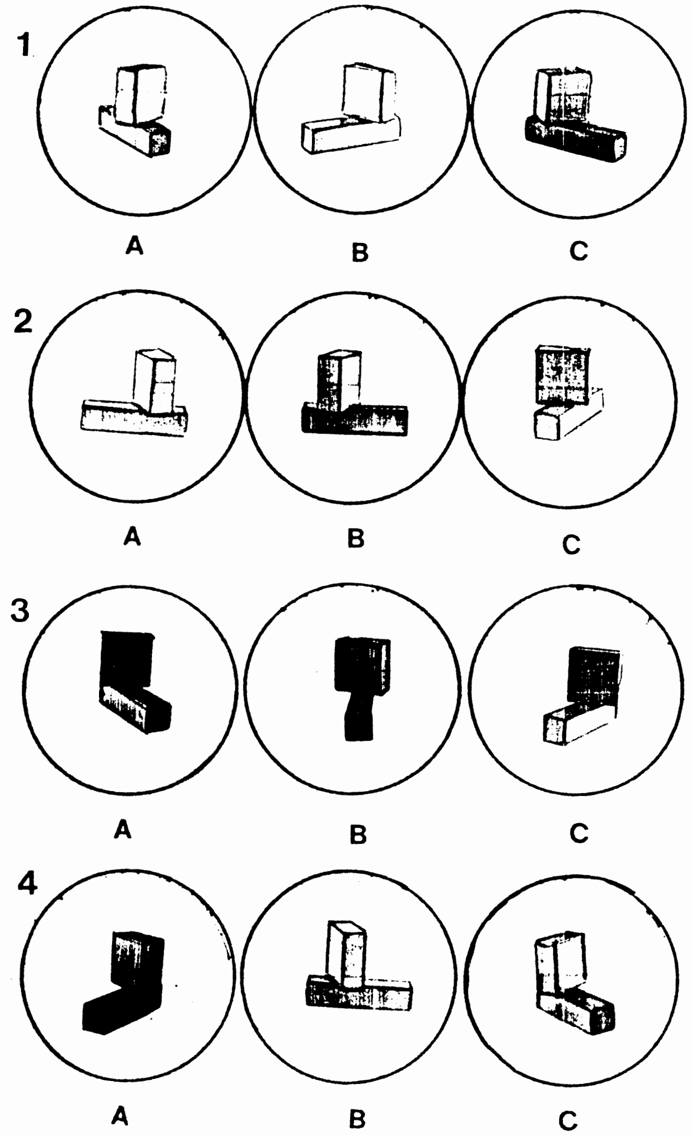


The object is rotating clockwise. Now, look at the two objects below. They are three-dimensional reflections or mirror-images of each other. Try rotating the object on the right. No matter how it is rotated it will not look exactly like the object on the left.



For each problem in this test there will be three pictures of three-dimensional objects. Two of the pictures are actually two views of the same object taken from different angles. One of the pictures is different from the other two; it shows an object which is a mirror-image or reflection of the others. There is no way to rotate this different object so that it looks like the other two. The two objects that can be rotated to look like each other are the same. You are to identify the picture of the object with is different from the other two.

Now, work the sample problems on the next page. On the answer sheet, circle the letter: **A**, **B**, or **C** that corresponds to the **object that is different from the other two**.



Look at problem #1. If object B is rotated clockwise about 60 degrees, it will look just like object A. Object C can not be rotated to look like A or B. C is, therefore, the correct answer to problem #1 because object C is different from A and B. The answers to the remaining problems are: #2 is B, #3 is C; and #4 is A.

Did you get the correct answers? If not, feel free to ask for assistance.

There are 30 items on this test, and you will have six minutes. Please work rapidly and accurately. Remember to write your answers on the answer sheet. Do not write on the test booklet.

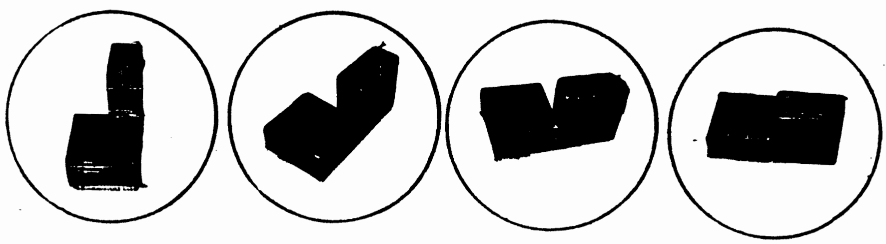
DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

Rotating Three-Dimensional Objects Test
(RTDOT) Form B

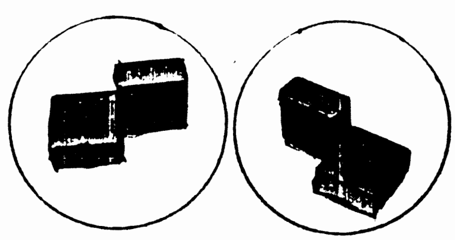
This is a test of your ability to imagine what a three-dimensional object will look like if it is rotated in the depth plane.

For each problem, there will be pictures of two three-dimensional objects. You are to decide whether the pictures are actually two views of the same object or whether the objects are different from each other. That is, the objects are different if there is no way to rotate one of the objects so that it looks exactly like the other view. If one of the objects can be rotated to look exactly like the other, the objects are the same.

For example, look at the series of pictures below. They show different views of one object as it is being rotated in the depth plane. Can you see that the pictures are all actually different views of the <u>same</u> object? Is the object rotating clockwise or counter-clockwise?

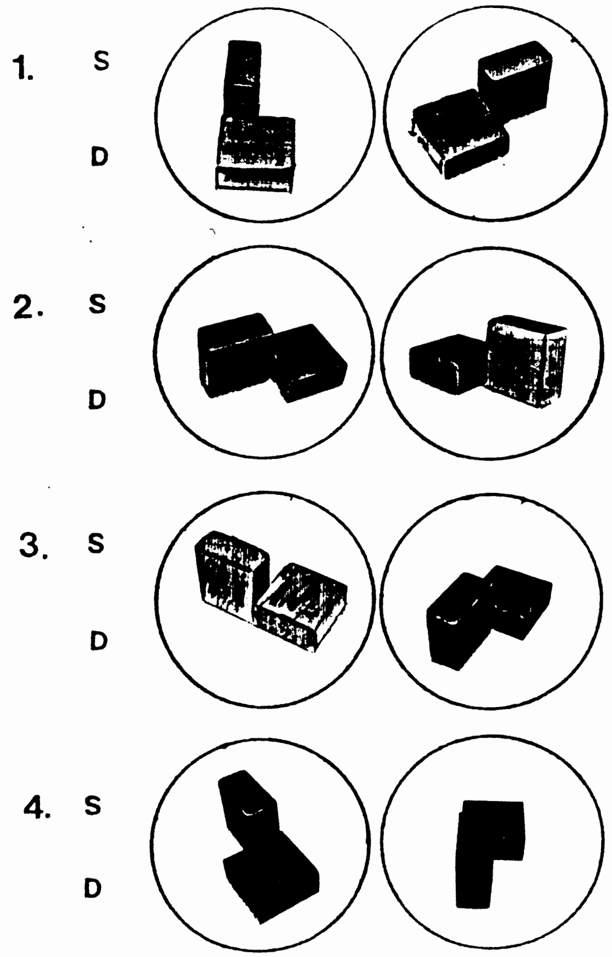


In the four pictures above, the object is rotating clockwise. Now, look at the two objects below. They are three-dimensional reflections or mirror-images of each other. Try rotating the object on the right. No matter how it is rotated it will not look exactly like the object on the left. The two objects are, therefore, <u>different</u>.



The items on this test are composed of actual pictures of objects. Sometimes, shadows are present. Please remember that the objects are always either exactly the same or mirror images.

Now, work the sample problems on the next page. Circle the **S** on the answer sheet, if the objects are the <u>same</u>. Circle the **D** if the objects are different because they are mirror-images of each other. Do not write on the test booklet. Please write all you answers on the answer sheet.



Imagine that the object on the left in item number 1 has been rotated clockwise about 30 degrees. It will look exactly like the object on the right in item number 1. Thus, the objects in problem number 1 are the <u>same</u>, and **S** is the correct answer. The answers to the remaining problems are: #2 is **D**; #3 is **D**; and #4 is **S**.

Did you get the correct answers? If not, feel free to ask for assistance.

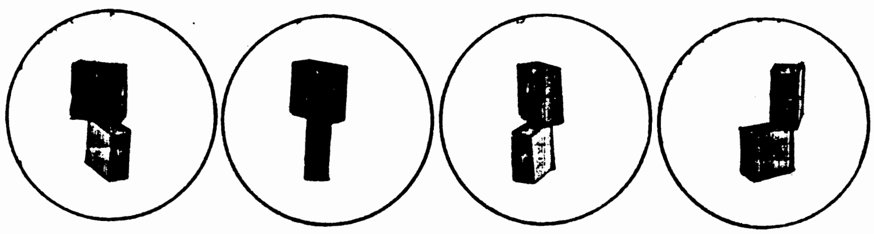
There are 40 items on this test, and you will have seven minutes. Please work rapidly and accurately. Remember to write all your answers on the answer sheet. Do not write on the test booklet.

DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

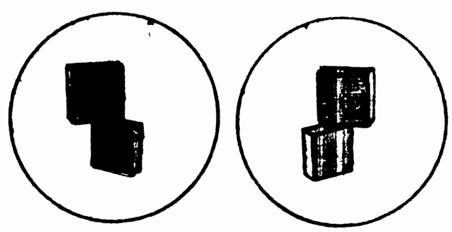
Depth Plane Object Rotation Test
(DPORT) Form C

This is a test of your ability to imagine what a three-dimensional object will look like if it is rotated in the depth plane.

Look at the series of pictures below. They show different views of one object as it is being rotated in the depth plane. Can you see that the pictures are all actually different views of the same object? Is the object rotating clockwise or counter-clockwise?

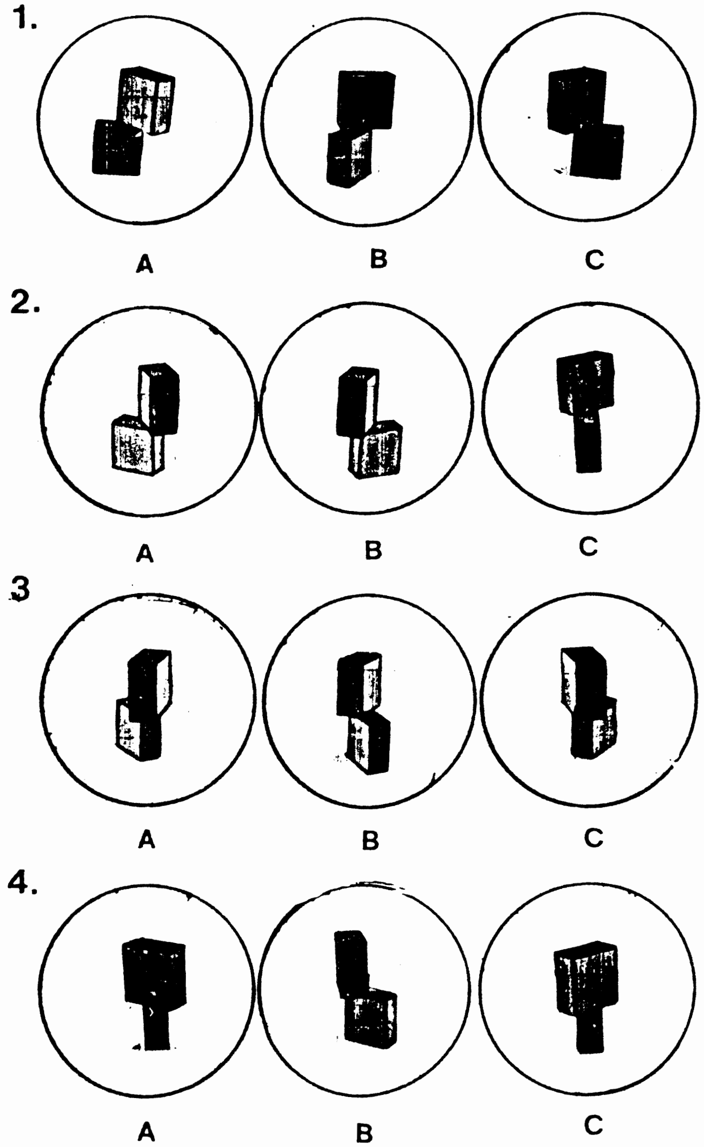


The object is rotating clockwise. Now, look at the two objects below. They are three-dimensional reflections or mirror-images of each other. Try rotating the object on the right. No matter how it is rotated it will not look exactly like the object on the left.



For each problem in this test there will be three pictures of three-dimensional objects. Two of the pictures are actually two views of the same object taken from different angles. One of the pictures is different from the other two; it shows an object which is a mirror-image or reflection of the others. There is no way to rotate this different object so that it looks like the other two. The two objects that can be rotated to look like each other are the same. You are to identify the picture of the object with is different from the other two.

Now, work the sample problems on the next page. On the answer sheet, circle the letter: **A**, **B**, or **C** that corresponds to the **object that is different from the other two.**



Look at problem #1. If object B is rotated clockwise about 60 degrees, it will look just like object A. Object C can not be rotated to look like A or B. C is, therefore, the correct answer to problem #1 because object C is different from A and B. The answers to the remaining problems are: #2 is B, #3 is C; and #4 is A.

Did you get the correct answers? If not, feel free to ask for assistance.

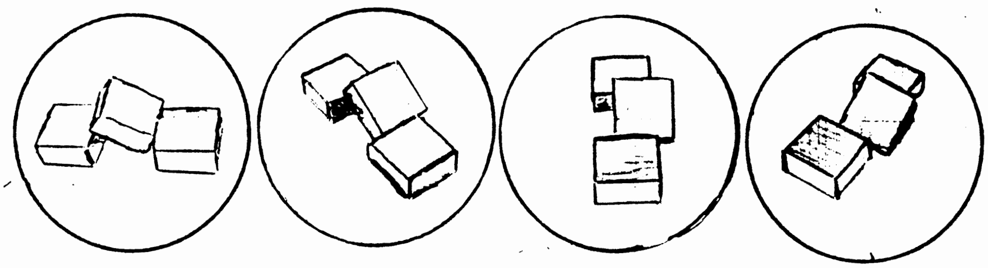
There are 30 items on this test, and you will have six minutes. Please work rapidly and accurately. Remember to write your answers on the answer sheet. Do not write on the test booklet.

DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

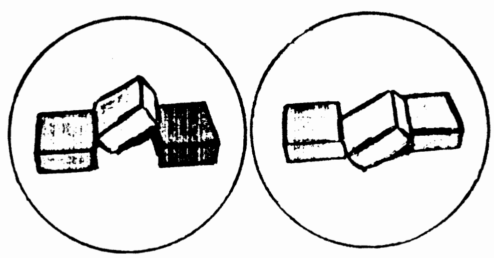
Depth Plane Object Rotation Test
(DPORT) Form D

This is a test of your ability to imagine what a three-dimensional object will look like if it is rotated in the depth plane.

Look at the series of pictures below. They show different views of one object as it is being rotated in the depth plane. Can you see that the pictures are all actually different views of the same object? Is the object rotating clockwise or counter-clockwise?

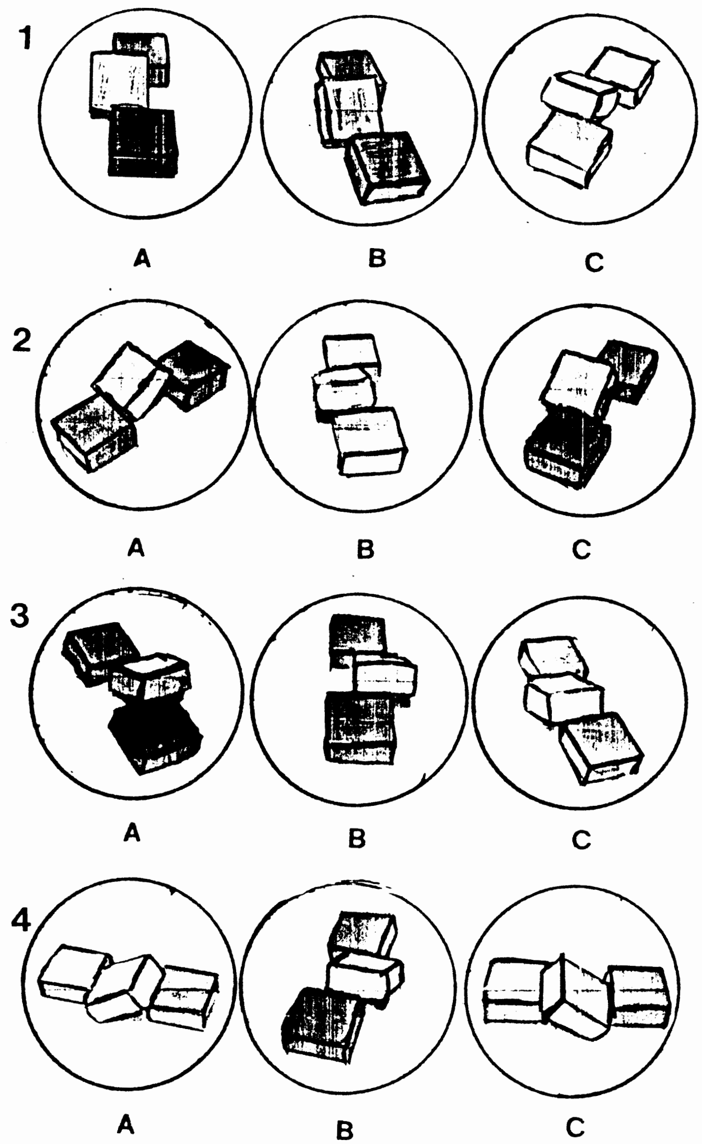


The object is rotating clockwise. Now, look at the two objects below. They are three-dimensional reflections or mirror-images of each other. Try rotating the object on the right. No matter how it is rotated it will not look exactly like the object on the left.



For each problem in this test there will be three pictures of three-dimensional objects. Two of the pictures are actually two views of the same object taken from different angles. One of the pictures is different from the other two; it shows an object which is a mirror-image or reflection of the others. There is no way to rotate this different object so that it looks like the other two. The two objects that can be rotated to look like each other are the same. You are to identify the picture of the object with is different from the other two.

Now, work the sample problems on the next page. On the answer sheet, circle the letter: **A**, **B**, or **C** that corresponds to the **object that is different from the other two**.



Look at problem #1. If object B is rotated clockwise about 30 degrees, it will look just like object A. Object C can not be rotated to look like A or B. C is, therefore, the correct answer to problem #1 because object C is different from A and B. The answers to the remaining problems are: #2 is B, #3 is C; and #4 is A.

Did you get the correct answers? If not, feel free to ask for assistance.

There are 30 items on this test, and you will have six minutes. Please work rapidly and accurately. Remember to write your answers on the answer sheet. Do not write on the test booklet.

DO NOT TURN THE PAGE UNTIL TOLD TO DO SO.

VITA

Teresa Lynne Russell

Candidate for the Degree of

Doctor of Philosophy

Thesis: SEX DIFFERENCES IN MENTAL ROTATION ABILITY AND THE EFFECT OF TWO TRAINING METHODS

Major Field: Psychology

Biographical:

Personal Data: Born in Woodward, Oklahoma, November 5, 1957, the daughter of Stanley V. and Leona M. Russell.

Education: Graduated from Alva High School, Alva, Oklahoma, in May, 1976; received Bachelor of Science degree in Psychology from Oklahoma State University in December, 1980; received Master of Science degree from Oklahoma State University in December, 1983; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in December, 1988.

Professional Experience: Teaching Assistant, Department of Psychology, Oklahoma State University, August 1981 to December 1983; Research Assistant, CONOCO, Inc., September to December of 1982; Research Assistant, Personnel Decisions Research Institute, January to December of 1984; Research Associate, Personnel Decisions Research Institute, April 1985 to present.