THE SUBJECTIVE DIFFICULTY OF SPATIAL ABILITY TESTS

by

KLAUS GERHARD SCHROEDER

B.Math., University of Waterloo, 1974

B.A., University of Waterloo, 1977

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Department of Psychology

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September, 1981

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Psychology

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date October 16, 1981

Abstract

Tests of Spatial Orientation (Card Rotations, Cube Comparisons) and Visualization (Form Board, Paper Folding, Surface Development) were administered to 537 (266 men, 268 women) university students. Participants rated the perceived difficulty of each of the tests on a 9 point scale ranging from 1 = very easy to 9 = very difficult. They were asked to indicate which of six problem solving strategies they used to solve the items on any particular test. The strategy statements were designed to tap part or whole problem solving strategies. Part strategies involved concentrating on salient aspects of a stimulus while whole strategies involved concentrating on an entire stimulus.

Since men scored higher than women on all five tests analyses were performed separately for the sexes. For both men and women the first principal component accounted for more than 50% of the variance. Thus, previous findings of two spatial factors for men and one spatial factor for women were not supported.

Problem solving strategy did not relate to performance on the spatial tests nor to difficulty ratings. There were no consistent sex differences in strategy except that women indicated that they guessed more on all tests. The limitations of introspective reports were discussed.

For both men and women the perceived difficulty of a particular test correlated more highly with the total score on that test than with the total score on any other test. On the basis of this finding it was concluded that the difficulty index is a valuable one worthy of further study. The finding that men and women did not differ on mean difficulty rating on three of the tests, even though they differed significantly in performance on all tests, was interpreted to mean that each person subjectively rank orders the tests in terms of difficulty. It was hypothesized that the perceived difficulty of a test is, therefore, a function of the other tests included for study.

There was moderate support for the hypothesis that, as the difference in rated difficulty for pairs of tests increases, the correlation between the two decreases. This was the case for six of 10 comparisons for men and three of 10 comparisons for women. It was suggested that this hypothesis would receive stronger support if tests of more distinct abilities were included in the same study.

Table of Contents

List of Tables

Acknowledgements

I wish to thank my committee members: the late Dr. Edro Signori for chairing the committee and for freely imparting his methodological knowledge, Dr. Susan Butt for her assistance in preparation of the final draft as well as for her encouragement, and Dr. John Yuille for his advice concerning the problem solving aspects of the thesis.

I also thank Dr. Ralph Hakstian for providing statistical guidance, and Marsha Schroeder for the drawings as well as for inspiration and emotional support throughout the project.

Historical Perspective on Spatial Ability Research

In summarizing the results of factor analytic studies of intellectual abilities Wolfle (1940) pointed out that a space factor was the second most frequently indentified factor in the literature. Although there was relatively little overlap in the tests used in many of the investigations of spatial ability to this time, various researchers provided nearly identical definitions of the ability they presumed to be measuring. McFarlane (1925), for instance, found a group factor distinct from general intelligence which measured an ability in which "uniqueness lies in the fact that those persons possessing it in a high degree analyse and judge better about concrete spatial relations" (p. 54). Koussy (1935) found evidence of a factor which he saw as the "ability to obtain and the facility to utilize visual spatial imagery" (p. 84) and Thurstone (1938) characterized his first primary mental ability as "facility in spatial and visual imagery" (p. 80).

While British researchers, for the most part, investigated the nature of this one general spatial factor for some time to come (Smith, 1964), numerous American investigators became interested in subfactors of the identified spatial ability. Notable are the findings of Guilford and Lacey (1947) from the U.S. Army Air Forces Aviation Psychology Research Program which mostly used armed services personnel. Their research led them to divide spatial ability into a spatial visualization and a spatial orientation factor. Spatial visualization was seen as the "ability to imagine the rotation of depicted objects, the folding and unfolding of flat patterns, the relative changes of position of objects in space, the motion of machinery. This visualization factor is strongest in tests that present a stimulus pictorially and in which some manipulation or transformation to another visual arrangement is involved" (McGee, 1979, p. 891). Spatial orientation was defined as the "ability to determine relationships between different spatially arranged stimuli and responses and the comprehension of the arrangement of elements within a visual stimulus pattern" (McGee, 1979, p. 891). It is not clear, however, on the basis of an analysis of the factor patterns listed by French (1951), how distinct these two factors might be. Many of the tests purported to load on one factor also load quite highly on the other, and tests later developed to measure these factors more parsimoniously (Guilford & Zimmerman, 1953) were often found to correlate substantially with one another (e.g. r = .67; Borich & Bauman, 1972).

As noted by McGee (1979), Guilford and Lacey (1947) were by no means the only researchers to split the spatial factor. Thurstone's (1950) work, for example, had a great impact on future investigations of spatial ability subfactors. Thurstone (1950) found three separate spatial factors, two of which (S1 and S3) are similar to the Guilford and Lacey (1947) spatial orientation factor, and one (S2) which resembles the spatial visualization factor. Citing Thurstone (1950), Barratt (1955) notes:

> "The first space factor (S1) represents the ability to recognize the identity of an object when it is seen from different angles;...the second space factor (S2) represents the ability to imagine the movement or internal displacement among the parts of a configuration that one is thinking about;...the third space factor (S3) represents the ability to think about those spatial relations in which the body orientation of the observer is an essential part of the problem" (p. 279).

In a review of spatial literature, however, Michael, Guilford, Fruchter, and Zimmerman (1957) concluded that S1 and S3 could be combined into one spatial relations and orientation factor due to the finding that tests of one of these factors often loaded on the other.

Following Thurstone (1950), French (1951) also identified three spatial factors after integrating the results of over sixty factor analyses, many of which were performed by or under the direction of Thurstone and Guilford. Space (S) was seen as the ability to perceive spatial patterns accurately and to compare them with each other; Spatial Orientation (SO) as the ability to remain unconfused by the varying orientations in which a pattern may be presented; and Visualization (Vi) as the ability to comprehend imaginary movements in 3-dimensional space or the ability to manipulate objects in imagination. The distinction between S and SO is somewhat unclear, however, and it seems that both factors might be included in the Guilford and Lacey (1947) definition of spatial orientation. The main distinction between the two seems merely to be that for S tests one has to compare two stimuli while for SO tests one works with the same stimulus. Further, in the analyses listed by French (1951) there is a considerable degree of overlap as to the tests with high loadings on these two factors. Given these findings and the fact that Michael et al. (1957) viewed their Spatial Relations and Orientation factor as a combination of the French (1951) S and SO factors, it is understandable that tests of only two spatial factors appeared in the 1963 version of the Kit of Reference Tests for Cognitive Factors (French, Ekstrom, & Price, 1963).

This trend continued with the 1976 version of the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, & Harman, 1976) when tests for two spatial factors, Spatial Orientation (S) and Visualization (Vz) were included. Thus it seems that researchers have converged in their belief that the earlier identified general spatial factor (e.g. Koussy, 1935; McFarlane, 1925; Thurstone, 1938) can be divided into two subfactors (McGee, 1979).

Purpose of the Study

Although there seems to be fairly widespread support for the view that S and Vz are distinct factors, there has, nevertheless, been considerable difficulty in clarifying the distinctions between the two. It is primarily this issue to which the present study is addressed. As Smith (1964) pointed out, even though the differences between the two have received detailed analysis in terms of such concepts as item complexity, task difficulty (e.g. Michael et al., 1957), and problem solving strategy (e.g. Barratt, 1953), none of the differences explains the distinctions between all different pairs of tests. Thus, for example, even though most Vz tests are assumed to be more difficult and comprised of more complex items than most S tests, counterexamples abound. Generally, there is widespread concensus that Vz tests are more difficult (e.g. Ekstrom et al., 1976; Werdelin & Stjernberg, 1969).

It has not been shown, however, that this phenomenon alone accounts for the two distinct factors. This is probably due to the inherent difficulty of obtaining a difficulty index for a speeded test. Further, not all tests are likely to be of the same relative difficulty for all persons and in order to show that the correlations between tests are affected by the difficulty of the tests involved one must compare correlations for various groups of people. This has not yet been researched and will be one of the concerns of the present study.

Another recurring finding has been that analysis of spatial test scores from men indicates more than one spatial factor while analysis of spatial test scores from women results in only one general spatial factor (Barratt, 1955; Michael, Zimmerman, & Guilford, 1950; Very, 1976). Men tend to score higher on various types of spatial tests (Barratt, 1955; Maccoby & Jacklin, 1974). One possible explanation for greater differentiation of abilities in men is afforded by research which indicates greater differentiation of abilities in groups which excel in these abilities (see Anastasi (1970) for a review). It is not clear, however, how the difficulty of a test relates to ability level. Thus, it is possible that the reported sex differences in structure are inextricably tied to difficulty. This will also be investigated.

One other area relating to the distinction between S and Vz which has received much attention is part versus whole problem solving strategy. It is generally believed that Vz tests require a person to keep an entire stimulus in mind (whole) while performance on S tests is enhanced by paying attention to the details (part) of the stimulus figures (Barratt, 1953, 1955; Ekstrom et al., 1957; Michael et al., 1950). The question of whether the problem solving strategy affects the correlations between tests will be investigated by comparing people who use the same strategies to those who use different strategies on the various tests used in this study.

The main focus of this study, then, lies not only in attempting to confirm previous findings related to sex differences and the part versus whole approach, but also in investigating hypotheses for distinct groups of people. An index of difficulty for speeded tests will be proposed and its psychometric properties investigated.

For the purposes of this study, the definitions of Spatial Orientation (S) and Visualization (Vz) are those provided by Ekstrom et al. (1976): S is the ability to "perceive spatial patterns or to maintain orientation with respect to objects in space" (p. 149) and Vz is the ability to "manipulate or transform the image of spatial patterns into other arrangements" (p. 173). The Card Rotations Test (S1) and the Cube Comparisons Test (S2) were used as measures of S and the Form Board Test (Vz1), the Paper Folding Test (Vz2), and the Surface Development Test (Vz3) were used as measures of Vz (see Appendix A for a description of these tests).

Review of the Literature

This section is divided into four main parts. The first deals with findings related to the psychometric distinctiveness of S (Spatial Orientation) and Vz (Visualization). The second, concerns sex differences and ability level. The third section deals with part versus whole problem solving strategies, while the fourth deals with the effects of test difficulty on the correlations between tests.

The Distinction between Spatial Subfactors

As stated previously, S tests sometimes load significantly on Vz factors and vica versa. Further, a factor is sometimes called spatial visualization even though no Vz-type tests are found to load on it. Examples of this can be found quite readily in French (1951). A case in point is an analysis (referred to as AFA) in which the visualization factor was loaded by Mechanical Principles along with mathematics and reading tests. In another analysis (AFE) both a spatial relations and a visualization factor were identified. While the visualization factor was loaded by Vz-type tests (as here defined) and the spatial relations factor was not, no S-type tests were included in the battery used. The second highest loading on the spatial relations factor was for a test in which the subject was required to adjust stick-and-rudder controls in response to a pattern of three lights. It is difficult to see how this would be tied to performance on the Cards Test (S1), say, where the subject merely indicates whether one pictured card is a reflection of another pictured card.

Other analyses shed more light on the distinction between S and Vz. In an analysis of a subset of Thurstone's (1938) data Fruchter (1948) found two factors which resemble S and Vz although there was some factorial complexity. The tests used were Cubes (S2), Flags (a test quite similar to the Cards Test - S1), Form Board (Vz1), and Paper Folding (Vz2). Although all tests loaded over .60 on the factor they were hypothesized to measure, Form Board loaded .50 on the S factor. Thus, while the distinction between S and Vz may not be perfectly clear, it does seem that, in this sample at least, an argument for two abilities could be made.

The results of Michael et al. (1950), with a sample of 500 male students, provide an even clearer distinction between S and Vz. The S tests were Cubes (S2), Flags, and the Guilford-Zimmerman Spatial Orientation Test (G-Z SO) and the Vz tests were Form Board (Vz1), Paper Folding (Vz2) and the Guilford-Zimmerman Spatial Visualization Test (G-Z SV). The results were quite unequivocal in that two of the S and two of the Vz tests loaded highly on their respective factors and not over .25 on the other factor. The G-Z tests were more complex and even though both loaded more highly on the hypothesized factors than other tests of those factors, they also loaded over .40 on the factors they were not meant to tap. Even so, this study, perhaps more than any other ever reported, shows S and Vz to be distinct spatial factors. To be noted is that neither of the G-Z tests were included in the 1976 edition of the Kit of Factor Referenced Cognitive Tests, although G-Z SO did appear as an S test in the 1963 Kit.

Other studies have not always factorially differentiated tests purported to define or load separate factors. Fleishman and Dusek (1971), for instance, administered 21 tests from the 1963 Kit to 90 army enlisted men and found that all spatial tests used (Cubes - S1, Paper Folding - Vz2, Surface Development - Vz3) loaded on the same factor. It is possible that this finding is due to the fact that only one S test was chosen for inclusion in the study. Using a greater variety of spatial tests with a sample of 177 eleventh and twelfth grade male students, French (1965) also found S and Vz tests to load on the same factor. The six spatial tests included in his battery were the three Vz tests from the 1963 Kit as well as Cards (S1), G-Z SO, and a concealed figures test. The latter test has at times been found to load with other S-type tests (French, 1951).

Very (1967) employed a number of spatial tests in a study of mathematical ability involving 193 male and 162 female college students. The tests used were mostly of the S variety (e.g. Cards - S1, Cubes - S2, G-Z SO) but also included Very's test of spatial relationships. This latter test included items requiring the subject to determine, for example, what would happen to the area of a rectangle if the length of each side were doubled. As concerns the spatial factors, the findings were quite sex specific. For women, all of the spatial tests, along with a number of other tests, combined to load on what was interpreted as a spatial factor. For men three spatial factors were identified, two of which seem to be relatively pure spatial factors. The first factor was defined by Cards (S1), Cubes (S2), and the spatial relations test while the second factor had its highest loading for G-Z SO but was also loaded by Cards, spatial relations, and a test involving the ability to determine the rule which relates two groups of letters. Thus, although this study does not provide a very stringent test of the hypothesis that S and Vz are distinct, due mainly to a lack of traditional Vz tests, it is interesting nonetheless. If one considers the notable loading of G-Z SO on a Vz factor mentioned previously (Michael et al., 1950) along with the finding that the same test split away from the S tests for the men, but not the women, it leads one to consider the possibility that Vz and S are more distinct for men than women.

Barratt's (1955) study provides some clarification of this matter because it included tests of both Vz and S and because factor analyses were performed for the sexes separately (103 women, 96 men). Although there were ten spatial tests and eight personality measures, discussion will be limited to tests which can be considered S or Vz or both. The S tests were Cards (S1), G-Z SO, Flags, and Thurstone's Figures test, a test which often loaded with Cards and Flags in the analyses presented by French

(1951). The Vz tests were the Minnesota Paper Form Board (similar to the Form Board - Vz1), G-Z SV, and the Spatial Relations subtest of the Differential Aptitude Test. This latter test is quite similar to the Surface Development Test (Vz3) in that both involve the folding of a 2-dimensional stimulus into a 3-dimensional object.

The results of the analyses were quite clear. For men, Flags, Figures, and Cards (S1), all S tests, defined one factor while all other tests, including G-Z SO, loaded on another factor. The only test with any notable complexity was the G-Z SV Test but its loading on what could be called the S factor was much lower (.43) than its loading on what could be seen as the Vz factor (.67). For women, on the other hand, all tests mentioned above loaded on one factor. This led Barratt (1955) to conclude that the "space tests in the female matrix did not show as clear a differentiation....as did the male analysis" (p. 286).

Sex Differences and Ability Level

The findings cited above lead one to ask whether S and Vz are separate abilities for men only. It should be remembered, for example, that much of the work reviewed by French (1951), including the work of Guilford and Lacey (1947), involved only male subjects. Further, the study which reported the clearest distinction between S and Vz (Michael et al., 1950) employed a male sample. Also, it is not clear whether Fruchter's (1948) results could be due to the sample employed since Thurstone (1938), whose data base Fruchter analyzed, did not describe his sample in terms of sex and pooled the sexes in determining correlations between tests. It may be conjectured, however, when one considers the time period, that men were overrepresented in Thurstone's sample.

If it is indeed the case that S and Vz are distinct for men only and that they collapse into one general spatial factor for women, the question arises as to why this should be the case. Possibly bearing on this issue is the finding that a sex difference in spatial ability has been one of the most consistent results in individual differences research (Garai & Scheinfeld, 1968; Harris, 1978; Maccoby & Jacklin, 1974). Results such as those reported by Barratt (1955) showing that men scored significantly higher than women on eight of ten spatial tests are not atypical in this regard. Although this sex difference is generally believed to obtain on both S and Vz tests (McGee, 1979) the literature related to this idea has never been systematically reviewed. In noting Werdelin's (1961) findings, Sherman (1967) wrote that "sex differences do not occur on tests in which visual organization is easily comprehended and which are to be manipulated as given, but on those in which the organization and reorganization aspects are exaggerated" (p. 293). This suggests, then, that the sex difference is limited to tests of Vz. Even if this is the case, it still does not explain why men exhibit two spatial factors and women only one.

Perhaps related to this is the evidence reviewed by Anastasi (1970) indicating greater differentiation of abilities in groups which excel in these abilities. Examples include more verbal factors for girls than boys and lower intercorrelations among primary mental abilities for higher scoring groups. Thus, it is possible that the sex difference in spatial ability is related to why S and Vz are more distinct for men than women.

Part Versus Whole Problem Solving Strategies

In relation to the greater differentiation of abilities for high scoring groups it might well be asked if groups who exhibit a relatively general factor, as opposed to more numerous specific factors, also tend to bring a more limited number of problem solving strategies to bear on the questions asked of them. In other words, is it possible that people who are not particularly adept at solving spatial items use the same strategies for all tests, while those who are more capable attempt different strategies for different tests? Further, does the type of problem solving strategy used influence the correlations between pairs of tests?

Most of the information related to these questions comes from investigations relying on introspective reports as well as numerous hypothetical statements.

French (1965) showed that the loading of a test on a factor can be substantially different for people who report employing different problem solving strategies. To begin, French divided his sample of 177 into two groups based on the following criteria: "Geometrical terms used in solving Cubes items;...Few visualization indications made in solving Cubes items;...When asked reports mentally rotating the cube on two separate axes" (p. 24). Subjects who were marked plus on two or more of these were deemed to have solved the Cubes (S2) items in an analytic manner and were compared to all other subjects. The most interesting findings were that the loading of Cubes on the Space-Visualization factor dropped from .52 for the non-analyzers to .07 for the analyzers and that, for the analyzers, Cubes loaded on an induction factor. Also, the correlations between various abilities were found to differ for people who used a system to solve problems (analyzers) and those who did not.

> "Subjects who attempt to solve spatial or other problems by ordinary common sense, perhaps by simple inspection until a solution seems to offer itself, succeed pretty much according to their general intelligence as measured by verbal and mathematical tests. For those subjects, therefore, the correlation between Space-Visualization and the verbal or mathematics factors is high. On the other hand, the subjects who use a system for solving problems have succeeded in developing some

> specialization of their abilities, and so the correlations drop down" (French, 1965; p. 22).

Although this study included both S- and Vz-type tests only one spatial factor was found. Thus, it was not possible to determine whether the highest loading for the Cubes Test would move from an S factor to a Vz factor, depending on the problem solving strategy used. Michael et al. (1950) have hypothesized that this might indeed be the case. It was their belief that people who mentally pick up and rotate a cube use a visualization process, whereas people who attempt to interrelate the positions of the sides or who picture themselves as walking around the cubes, use a spatial orientation strategy. Similarly, they believed that the Flags test (usually thought of as an S test) should load on the Vz factor if the flag is mentally picked up and turned, and that it should load on the S factor if it is solved with only the edges of the flag in mind. In summation, if a spatial item is solved by keeping the entire stimulus in mind (the 'whole' problem solving strategy) the test should load on a Vz factor. If an item is solved by concentrating only on various salient aspects of the stimulus (e.g. an edge, point, or side), perhaps one after the other (the 'part' problem solving strategy), the test should load on an S factor. Unfortunately, Michael et al. (1950) did not test this hypothesis.

In relation to the part versus whole strategy, Barratt (1953) found that people who used a part strategy (e.g. rotating only part of a figure) scored higher than people who rotated the whole figure. Also, people who attempted to relate the figures to something familiar (e.g. a letter of the alphabet, a concrete object) scored lower than people who employed a part strategy. Further, four types of problem solving strategies for the Space subtest of the Differential Aptitude Test were delineated. Some people folded the stimulus and noted relationships of parts; some unfolded the response figure and compared this unfolding to the pattern; others looked merely for cues such as angle intersections; and some people could not solve the problems and relied on guessing. Barratt did not report which problem solving strategy led to the highest scores. Also, correlations between the various tests used were not reported separately for groups who used different strategies. Thus, it is not known if different strategies might result in a different factor structure.

In summary, it seems that different strategies are employed by different people in solving the same items. For the most part, it is not known which strategies result in the highest scores. Also, we do not know if the correlation between two tests is higher when the same strategy is used for both than when a different strategy is used for each.

Test Difficulty and the Correlation between Tests

Although many investigators concerned with spatial ability have suggested that S tests are generally less difficult than Vz tests (Ekstrom et al., 1976; French, 1951; Michael et al., 1957), few empirical studies related to this point have been attempted. This is probably a result of the difficulty of obtaining a difficulty index for speeded tests. Thus, before discussing findings related to spatial ability, tests of which are, for the most part, speeded, studies involving more objective difficulty indices will be reviewed.

That test and item difficulty, at least in the absolute sense, is indeed a potent variable as concerns factor structure and the correlations between tests of varying difficulty has been amply documented. For example, Guilford (1941) factor analyzed the ten subscales of the Seashore test of pitch discrimination. Subscale scores were obtained for the 300 college students by tabulating which of the ten items at each of ten frequency differences they detected correctly. Frequency differences ranged from .5 to 30 cycles per second and subjects either passed or failed each subtest based on Guilford's attempt to obtain a non-skewed distribution for each of the subtests. Factor analysis based on tetrachoric correlations revealed three factors; Factor I had loadings which were inversely proportional to the difficulty of the items involved in the subtests; Factor

II had significant loadings for the more difficult subtests and Factor III had loadings for the moderately difficult subtests. Guilford (1941) summed up his study by noting that "we may accept the hypothesis that...(the factors)...represent three distinct human abilities involved in the comparison and judgment of tones as to pitch" (p. 74).

Ferguson (1941) obtained a result quite similar to Guilford's when he split a 100 word vocabulary test into six subtests of varying difficulty. This led Ferguson to suggest that "factors deduced from test batteries which are homogeneous with respect to difficulty, although heterogeneous with respect to content, would lend themselves more readily to psychologically meaningful interpretation than factors deduced from test batteries which are heterogeneous with respect to both difficulty and content" (p. 329). Thus, the implication is that different factors arising out of similar content are due to differential difficulty.

A study performed by Sargent (1940) in which subjects were asked to solve easy, difficult, and moderately difficult anagrams also relates to the above findings. In correlating mean solution times for the different groups of anagrams, it was found that correlations for tests one level removed in difficulty (i.e. easy-moderate, moderate-difficult) were substantially higher than correlations between tests two levels removed in difficulty (i.e. easy-difficult). Thus, as the difference in difficulty between the tests increased the correlation between the two decreased.

Given that difficulty can have an effect on correlations, it is still not at all clear that difficulty, per se, causes the correlations. Some researchers have postulated that as one type of item (e.g. an anagram) is made more difficult, its solution requires a different type of problem solving strategy or ability. Sargent (1940), for instance, had trained subjects verbalize their thoughts while solving the anagrams and found that while the easier anagrams were solved as a whole, the more difficult ones were solved by a part method which involved breaking the anagrams into syllables or familiar combinations. Along similar lines, Guilford (1941) hypothesized that the three factors discovered with the pitch discrimination test reflected auditory, attentional, and verbal components although he had no way of knowing which component was represented by which factor.

In the area of spatial ability Zimmerman (1954) hypothesized that "merely by varying item difficulty and complexity a single kind of test could be made to emphasize each of four factors in succession from Perceptual Speed through Space and Visualization to Reasoning" (p. 398).

Werdelin and Stjernberg (1969) attempted to test this hypothesis by administering S and Vz tests as well as perceptual speed, numerical, and reasoning tests to a sample of eighth grade boys and girls. Although separate S and Vz factors were not found for this sample, a result not inconsistent with other findings employing relatively young samples (e.g. French, 1951), the results indicated, for some spatial tests, a "clear tendency for increased correlations with the P (perceptual speed) and N (numerical) factors and somewhat lower correlations with the S (space) factor the more the test is practiced" (p. 190). Thus, although this finding does not directly address the distinction between S and Vz, it does show that as a test becomes easier (i.e. is practiced) it may measure a different ability. Werdelin and Stjernberg (1969) also pointed out, however, that the method by which the items are solved may change with increased practice. Thus, for the test which showed the most pronounced practice effect, a test quite similar to Cards (S1), they noted that "the problem...might after the practice period at least part of the time be solved without the subjects having to turn the various figures around in their minds" (p. 191).

The findings related to spatial ability, in contrast to the studies by Guilford (1941) and Ferguson (1941), do not involve an operational definition of difficulty. For instance, Zimmerman (1954) assumes that keeping track of three turns of a plane is more difficult than keeping track of merely one turn. Werdelin and Stjernberg (1969) assume that as a test is practiced it becomes easier. While this kind of logic is hardly debatable, consideration of difficulty in this manner does not allow one to assess its effects empirically. For example, most people would agree that a test involving three turns is more difficult than one involving one turn when the same stimuli are used in both (as is the case with Zimmerman, 1954). There would probably not, however, be the same degree of agreement if people were asked to compare the difficulty of a test involving three turns of one stimulus to the difficulty of a test involving the folding and unfolding of a piece of paper. Thus, what is needed is a measurement of difficulty.

To this end, it is proposed that the difficulty a person has in completing a test can be assessed by asking how difficult he or she found the test. While this is a relatively subjective index, and not necessarily related to difficulty indices based on pass/fail considerations (see Nunnally, 1978), it is believed to be a valuable one. This is especially so because it allows for careful consideration of individual differences. Not only can one determine how difficult a person found a test, but how difficult a person found one test relative to others. Further, it would be possible, empirically, rather than anecdotally, to investigate the relationship between difficulty and the topics reviewed above (i.e. sex differences, problem solving strategies, structure of spatial ability). For instance, one could readily determine whether people who use certain problem solving strategies find a test more or less difficult than people who do not use these strategies. Also, the effect that test difficulty has on the correlation between tests is open to investigation on the basis of subjective test difficulty indices for speeded tests.

Hypotheses

Hypothesis 1

The difficulty rating assigned to a test will correlate more negatively with the total score on that test than with the total score on any other test.

Rationale: This is a test of the adequacy of the difficulty rating scale. If this is not generally the case then it would be difficult to justify interperson comparisons based on the difficulty indices. Further, groups created on the basis of the difficulty indices could not, meaningfully, be compared on any performance or ability variable.

Hypothesis 2

Men will score significantly higher on Vz, but not on S tests, than women.

Rationale: This is a test of Werdelin's (1961) hypothesis. Further, it is an attempt to support the often cited sex difference in spatial ability.

Hypothesis 3

Men will rate Vz tests, but not S tests, as significantly less difficult than will women.

Rationale: If rated difficulty correlates with ability or performance level, the difficulty findings should parallel those of ability (Hypothesis 2).

Hypothesis 4

People using a part approach on one test and a whole approach on another test will exhibit a lower correlation between these tests than people using the same approach on both tests.

Rationale: This is an indirect test of the finding that people who use different approaches for different tasks have "succeeded in developing some specialization of their abilities" (French, 1965; p. 22). Also, the hypothesis that any particular test can change from being an S test to being a Vz test depending on the strategy employed (Michael et al., 1950) can be tested in this way.

Hypothesis 5

The sexes will differ in the strategies they employ to solve the items on various tests.

Rationale: This hypothesis relates to the statement made by Werdelin and Stjernberg (1969) that the way in which items are solved may change with practice. If the 'with practice' aspect of this postulate is taken to imply that the items become easier and if women do, in fact, find certain tests more difficult than men this hypothesis would indirectly test the assumption put forth by Werdelin and Stjernberg.

Hypothesis 6

For S tests, people using a part approach will score significantly higher than people using a whole approach. For Vz tests the reverse will be the case.

Rationale: The S tests do not require a restructuring of the stimulus. It would be more efficient, as found by Barratt (1953), to compare the salient aspects of the figures rather than the figures as a whole. The Vz tests, on the other hand, do require a restructuring of the stimulus in order to compare it to the figure(s) from which the answer must be chosen. Thus, it would probably be more efficient to keep the entire restructured stimulus in mind when searching for an answer than to make the many comparisons necessary for correct solution.

Hypothesis 7

For S tests, people using a part approach will find the tests less difficult than people using a whole approach. For Vz tests the reverse will be the case.

Rationale: If rated difficulty correlates with ability, the difficulty findings should parallel those of ability (Hypothesis 6).

Hypothesis 8

S tests will be rated as less difficult than Vz tests.

Rationale: This hypothesis tests statements made to this effect by Zimmerman (1954) and by Ekstrom et al. (1976) in the Manual of the Kit of Factor-Referenced Cognitive Tests. It is not clear that this hypothesis has ever been tested either with objective or subjective (as used here) difficulty ratings.

Hypothesis 9

Principal component analysis of the five spatial test scores will yield a first principal component of smaller variance for men than for women.

Rationale: This is an attempted confirmation of Very (1967) and Barratt (1955) that S and Vz are more distinct for men than women.

Hypothesis 10

For pairs of tests, as the difference in rated difficulty increases the correlation between the two will decrease.

Rationale: The evidence discussed previously indicates that the correlations between tests decrease as the difference in objective difficulty increases (Ferguson, 1941; Guilford, 1941; Sargent, 1940). This hypothesis tests these findings at a more individual level in that it takes into account how difficult each test was for each individual rather than at a group level.

Method

Questionnaires

The Card Rotations Test (S1) and the Cube Comparisons Test (S2) were used as measures of Spatial Orientation (S), and the Form Board Test (Vz1), the Paper Folding Test (Vz2), and the Surface Development Test (Vz3) were used as measures of Visualization (Vz). These tests were taken from the 1976 version of the Kit of Factor-Referenced Cognitive Tests (Ekstrom et al., 1976). Part one of each test was used. Tests Vz3 and Vz1 were completed third and fifth, respectively. The order of the other three tests (S1, S2, and Vz2) was randomized.

For the difficulty rating a 9 point scale labelled at three points (1 = very easy, 5 = moderately easy/difficult, 9 = very difficult) was used (see Appendix B). There was one rating scale for each test and the test to be rated on each scale was made salient by preceding the scale with stimulus items from the appropriate test.

Problem solving strategy was assessed with six items for each test (see Appendix C). To each statement participants were to answer either 'TRUE' or 'FALSE'. Statements for each test were again preceded by sample items from that test (see Appendix D for an explanation of the statements used).

Procedure

After a brief introduction to the purposes of the study, participants were asked to complete a title page by indicating age, sex, year of university, university faculty and major. They were also asked to provide their student identification numbers on this page if they wanted anonymous feedback.

After participants indicated that they had completed the title page they were asked if there were any questions. After answering these, the following procedure was employed for each of the tests. Participants were asked to turn the page and read the instructions to the next test (see Ekstrom et al., 1976 for the instructions). After 1.5 minutes they were asked to indicate if they had not finished the instructions by raising their hands. If a majority of people had finished, an additional 15 seconds were allowed. Otherwise, an additional 30 seconds were given. At the end of this time, participants were told to 'start now'. At the end of the time allowed for completion of each test (see Appendix A), participants were told to 'stop'.

After the last test (Vz1) participants were asked to rate the difficulty of each test (see Appendix B) and to indicate how they solved the items on each of the tests (see Appendix C). This part of the study was not timed. Upon completion participants were asked to take their booklets to the front of the room.

At a later date feedback was provided indicating an individual's performance as well as means for each of the tests.

Participants

Participants were recruited in introductory psychology classes during the fall term of the 1980/81 academic year. Although 572 people returned booklets, 35 of these were dropped from all analyses because one or more of the tests were not completed. People were deleted due to the possibility that they worked on a test at an inappropriate time (i.e. when they should have been completing another test). Thus, all analyses are based on a maximum of 537 people.

The mean age of the 537 participants (266 men, 268 women, 3 sex not specified) was 18.78 (minimum 17, maximum 40). Of the 484 people who reported a faculty, 51.7% were registered in Arts, 21.9% in Science, 12.2% in Engineering, and 14.2% in various other faculties. Of the 492 people who reported year of registration, 73.2% were in first, 23.8% in second, 2.6% in third, and 0.4% in fourth year.

Results

The means and standard deviations for the difficulty ratings and all test scores (corrected for guessing) are presented by sex and for the total group in Table 1.

Due to the large number of people involved in most analyses to be discussed below, trends (i.e. $.05 < p < .10$) were treated as nonsignificant. Further, when post hoc comparisons were performed Scheffe's method (Hays, 1973), with the significance level set at .05, was used exclusively. Moreover, all 2x2 Chi-square values reported are based on Yates' correction regardless of expected cell frequency (Ferguson, 1976).

In consideration of the widely cited sex differences in spatial ability all analyses were performed for the sexes separately as well as for the total group. In the case of analysis of variance (ANOVA) sex is always a factor.

Hypothesis 1

The difficulty rating assigned to a test will correlate more negatively with the total score on that test than with the total score on any other test.

The correlations related to this hypothesis are presented in Table 2. For women, men, and the total group, the correlations were as predicted. For men the correlations between difficulty rating and test score ranged from -.24 (Cubes) to -.59 (Surface Development) while for women they ranged from -.16 (Form Board) to -.64 (Surface Development).

Thus, although there are substantial differences in the magnitude of the correlations, perceived difficulty does appear to be related to performance. Further, the perceived difficulty of any particular test is more related to performance on that test than to performance on any other test.

Hypothesis 2

Men will score significantly higher on Vz, but not on S tests, than women.

In order to stringently test this hypothesis a Hotelling's T^2 analysis was performed with the five spatial test scores serving as the dependent variables. Men and women were found to differ significantly (F (5,528) = 26.42, $p < .01$) and post hoc multiple comparisons revealed that men scored significantly higher than women on all tests (see Table 1 for means and standard deviations).

It seems, therefore, that the sex difference in spatial ability is not limited only to tests of Vz, but manifests itself on S tests as well.

Hypothesis 3

Men will rate Vz tests, but not S tests, as significantly less difficult than will women.

For this analysis the five difficulty ratings were treated as repeated measures and a Sex x Test Difficulty repeated measures ANOVA was performed (see Table 3). This revealed significant effects for Sex (F (1,529) = 62.07, $p < .001$), Test Difficulty (F (4,2116) = 327.17, $p < .001$), and a Sex x Test Difficulty interaction (F (4,2116) = 24.50, $p < .01$). Post hoc comparisons indicated that women rated both the Cubes and Surface

Development tests as more difficult than did men. There were no significant differences on the other tests (see Table 1 for means and standard deviations).

Considering that there were sex differences on only two of the tests (one S and one Vz) support for this hypothesis does not seem adequate. The sex differences in ability (i.e. significant differences on all tests) do not parallel the findings concerning difficulty.

Within sex comparisons based on this analysis will be discussed under Hypothesis 8.

Hypothesis 4

People using a part approach on one test and a whole approach on another test will exhibit a lower correlation between these tests than people using the same approach on both tests.

The number of people who answered 'true' to the various statements regarding problem solving strategy are given in Table 4. Of interest to this hypothesis are the first four statements for each test.

For any particular test, people who answered 'true' to one or more of the part strategies (as indicated in Appendix D) and answered 'false' to all of the whole strategies are defined as the 'part' subgroup. The 'whole' subgroup consists of people who answered 'true' to one or more of the whole strategies and 'false' to all of the part strategies. A third subgroup, to be considered later, consists of people who answered 'true' to at least one part strategy as well as at least one whole strategy. This group is referred to as the 'part/whole' subgroup. Thus, for the Paper Folding Test, for instance, a person who answered 'true' to strategies one and two and 'false' to three and four would fall into the part/whole subgroup; a person who answered 'true' to strategy one and 'false' to the rest would be in the whole subgroup; a person who answered 'true' to strategy two and 'false' to the rest would be in the part subgroup; and people who answered 'false' to all four strategies would not be considered in any analysis involving problem solving strategy. For pairs of tests, then, people who fell into the same subgroup on both tests (i.e. into the part subgroup on both tests or into the whole subgroup on both tests) were deemed to have used the 'same' problem solving approach to both tests. Those who fell into different subgroups for the two tests (i.e. the part subgroup for one test and the whole subgroup for the other) were deemed to have employed 'different' problem solving approaches to the two tests.

The number of people who fell into the 'same' and 'different' groups for each pair of tests as well as the correlations between the tests for the two groups are given in Table 5.

For the total sample, as well as for the sexes separately, the correlations between the 'same' and 'different' groups were compared by first using Fisher's _r_ to _Z_ transformation (Hays, 1973). Of the 30 comparisons, 13 negative and 17 positive z values were obtained. None of these surpassed the critical value for a one-tailed test (1.65).

Given these findings, it may be concluded that the distinction between part and whole problem solving (as here measured) has little to do with the correlations between tests.

Hypothesis 5

The sexes will differ in the strategies they employ to solve the items on various tests.

To assess this hypothesis 2x2 Chi-square (male/female versus true/false) analyses were performed for each of the 30 strategy statements (see Table 4 for the number of people in each group).

On all tests women indicated that they guessed more ($p < .05$ in all cases). Further, women answered 'true' to the following statements more often than men: I mentally folded the paper, punched a hole, and then mentally unfolded it and compared it to the possible answers (from Paper Folding; $\chi^2(1) = 4.12$, $p < .05$); I didn't use any particular strategy (from Surface Development; $\chi^2(1) = 23.20$, $p < .001$); I noted some distinct aspect of the figure (e.g. bottom part or top left corner) and solved the figures with this feature in mind (from Cards; $\chi^2(1) = 6.81$, $p < .01$); I used a pencil to draw the large figure out of the smaller ones (from Form Board; $\chi^2(1) = 4.80$, $p < .05$).

At a more global level, Chi-square tests were performed to determine if, on each test, sex was independent of overall strategy employed (i.e. part, whole, part/whole). The only test to reach significance was for the Cards test ($\chi^2(2) = 8.19$, $p < .05$). This was highlighted by the fact that a greater percentage of men (36.9 compared to 25.6% of the women) fell into the part subgroup, while a greater percentage of women (64.9 compared to 53.6% of the men) fell into the part/whole subgroup (see Tables 6 to 10 for the number of men and women in each group for each test).

Generally, then, this hypothesis may be rejected. Only three of the twenty statements which were designed to reflect part or whole problem solving strategy discriminated men from women. Also, on only one test did the sexes differ in overall strategy employed and this finding may well have been influenced by the small number of people in the part subgroup.

Hypothesis 6

For S tests, people using a part approach will score significantly higher than people using a whole approach. For Vz tests the reverse will be the case.

For each test a Sex x Strategy (part, whole, part/whole) ANOVA was performed (see Tables 6 to 10 for the source tables and means). In all cases, men scored significantly higher than women (see Hypothesis 2 for a more appropriate statistical test of these findings). Of major concern for this hypothesis are the Strategy and Sex x Strategy findings.

For the Cards Test a significant effect for Strategy was found (F (2,519) = 4.33, $p < .05$), although post hoc comparisons (Scheffe) indicated no significant differences between any two of the subgroups. On the other S test (Cubes) there was a significant Sex x Strategy interaction (F (2,518) = 4.07, $p < .05$). Although post hoc comparisons revealed no within sex differences, the male part/whole subgroup scored higher than the female whole and part/whole subgroups, and the male part subgroup did not score higher than any female subgroup. In considering these results it is important to note the relatively few people in the part subgroups (19 men, 21 women).

For the Vz tests, there was a significant Strategy effect for Paper Folding (F (2,521) = 3.53, $p < .05$) and significant Sex x Strategy (F (2,512) = 5.14, $p < .01$) and Strategy (F (2,512) = 5.52, $p < .01$) effects for Surface Development. For the Form Board Test the only significant effect was Sex.

Post hoc comparisons for the Vz test findings indicated that, for Paper Folding, the part subgroup scored lower than the part/whole subgroup. To be noted is that the part subgroup consisted of only 25 people. The Sex x Strategy interaction for Surface Development was highlighted by the finding that the male whole subgroup scored lower than the male part subgroup. There were no other within sex differences, and all male subgroups scored higher than all female subgroups.

Given the significant sex differences in performance (see Hypothesis 2) the analyses of major concern to this hypothesis are the Strategy and Sex x Strategy post hoc comparisons. The only within sex comparison to achieve significance (Surface Development) does not support the hypothesis since the male whole subgroup scored lower than the male part subgroup. Further, the only post hoc comparison related to Strategy which was significant (Paper Folding) did not indicate that the part and whole subgroups differed. Thus, it seems that performance is not tied to problem solving strategy in a consistent manner and the hypothesis may be largely rejected.

Hypothesis 7

For S tests, people using a part approach will find the tests less difficult than people using a whole approach. For Vz tests the reverse will be the case.

For each test a Sex x Strategy (part, whole, part/whole) ANOVA was performed with difficulty rating serving as the dependent variable (see Tables 11 to 15 for the source tables and means). Except for the Form Board Test women rated all tests as more difficult than did men. Since sex differences related to difficulty were discussed previously (Hypothesis 3) based on a more appropriate statistical procedure, they will not be discussed further here.

Two other effects also reached significance. On the Paper Folding Test post hoc analyses of the Strategy effect ($\underline{F}$ (2,526) = 5.47, $\underline{p} < .01$) revealed that the part subgroup rated the test as more difficult than each of the other two groups. Although there was a significant Sex x Strategy interaction for the Surface Development Test ($\underline{F}$ (2,512) = 4.69, $\underline{p} < .01$), post hoc comparisons did not reveal any within sex differences. All male subgroups rated the test as less difficult than all female subgroups.

Since only one finding was in the hypothesized direction (the Strategy effect for Paper Folding) this hypothesis may be largely rejected.

Hypothesis 8

S tests will be rated as less difficult than Vz tests.

This hypothesis was assessed by performing within sex post hoc comparisons on the various mean difficulty ratings (see Hypothesis 3 for a discussion of the repeated measures design employed). For both men and women the Paper Folding Test and the Cubes Test were the only two tests which did not differ significantly in mean difficulty rating. Women rated the tests, from easiest to most difficult, in the following order: Cards (S1), Paper Folding (Vz2), Cubes (S2), Surface Development (Vz3), Form Board (Vz1). Except that the order of Surface Development and Form Board was reversed, men ordered the tests in the same way.

Support for this hypothesis is tenuous. Although two Vz tests were rated most difficult (Surface Development and Form Board) the other Vz test (Paper Folding) did not differ in rated difficulty from the more difficult S test (Cubes).

Hypothesis 9

Principal component analysis of the five spatial test scores will yield a first principal component of smaller variance for men than for women.

The correlation matrices which were factored are presented in Table 16. To be noted is that, with few exceptions, the correlations among tests of the same type (i.e. within S and within Vz test correlations) are higher than the correlations between S and Vz tests. Even though this is the case, however, the eigenvalue corresponding to the second principal component was never greater than unity. The first principal component accounted for 52.5, 52.1, and 56.8% of the variance for men, women, and the total group, respectively. The corresponding percentages for the second principal component were 17.1, 16.3, and 15.1. Further, visual inspection of the matrices presented in Table 16 reveals few differences in magnitude in the correlations for men and women.

Regardless of the findings presented above and ignoring the possibility of overfactoring when the second component accounts for so little variance (see Harman, 1976), two principal components were extracted and rotated to simple structure using a Varimax rotation. The obtained pattern matrices are presented in Table 17. Interpretation of these matrices would indicate that the Vz tests define Factor I and that the S tests define Factor II.

The analyses for two factors are included for illustrative purposes only. This is also true of the analyses for the total group because the test for inequality of the variance/covariance matrices (see Morrison, 1976) indicated that the data for men and women should not, ideally, have been pooled ($\chi^2(15) = 32.71$, $\underline{p} < .01$). The likelihood that the two-factor solutions represent an overfactoring of the data is suggested by the application of the three most commonly used tests for the number of factors to retain (see Hakstian, 1973). The 'Kaiser-Guttman rule' (eigenvalues greater than unity) suggests one factor for both men and women. The 'maximum likelihood-ratio test' suggests two factors for men ($\chi^2(1) = .59$, $\underline{p} = .44$) and one factor for women ($\chi^2(5) = 10.08$, $\underline{p} = .07$). Finally, the 'Scree test' (finding the break in the curve of plotted eigenvalues) indicates one factor for both men and women.

Thus, three of three tests for women and two of three tests for men suggest one spatial factor. Due to the author's belief that the data have probably been overfactored the two-factor findings will not be pursued further.

Hypothesis 10

For pairs of tests, as the difference in rated difficulty increases the correlation between the two will decrease.

For any two given tests, the absolute value of the difference in rated difficulty was calculated for each person. On the basis of the distribution of this value a median split procedure was employed to yield two groups; a group with a 'low' difference in difficulty (the low difference group), and a group with a 'high' difference in difficulty (the high difference group). This strategy was followed for men and women separately, and also for the group as a whole. The values used to define the two groups were the same for the male, female, and total groups with two exceptions (see Table 18 for the values used).

Following the creation of the low and high difference groups the correlation between the two tests involved was computed for each group. For clarification of the procedure employed, two tests will be chosen (Paper Folding and Surface Development).

For each person in the total group the difficulty rating of Paper Folding was subtracted from the difficulty rating of Surface Development. The absolute value of this difference was taken and a frequency distribution of, in this case, 537 values was obtained. Inspection of this distribution revealed that 268 people had absolute values of two or less while 269 people had absolute values of three or more. Since it was desired to create two groups of approximately equal size, those people with values of two or less defined the low difference group, and people with values greater than two comprised the high difference group. Following this, the correlation between Paper Folding and Surface Development was computed for both the low and high difference groups. This procedure was followed for each pair of tests (i.e. 10 pairs for five tests).

The correlations for the high and low difference groups were compared by first using Fisher's r to Z transformation. These correlations, along with the number of people entering into their computation, are presented in Table 18.

For the total sample five of the 10 comparisons were significant ($p < .05$, one-tailed) in that the low difference group exhibited a higher correlation between the two tests than the high difference group. For men, six of the 10 comparisons were significant and for women three of the comparisons were significant.

To address the possibility that these results may, in some way, be artifactual due to a truncation of range or lessened variability, the variances of the low and high difference groups were compared on the tests involved. On those comparisons for which a significant difference in correlations was found, the following differences in variability were noted; on the Paper Folding/Cards comparison the low difficulty group was more variable on Paper Folding (F (252,284) = 1.34, $p < .01$); on the Paper Folding/Surface Development comparison the low difficulty female group was more variable on Surface Development (F (115,163) = 1.36, $p < .05$); on the Surface Development/Cubes comparison the low difficulty male group was more variable on Cubes (F (160,104) = 1.36, $p < .05$).

Thus, there seems to be some support for the hypothesis that perceived differences in difficulty are related to the correlations between tests. The effect, however, seems to be much stronger for men (six of 10) than for women (three of 10).

Discussion

In general, the major purpose of this study was to clarify the distinction between two subfactors of spatial ability in terms of problem solving strategy and perceived difficulty of the tests involved. Secondarily, spatial tests were the vehicle used to test some theoretical notions that differences in perceived difficulty between two tests would affect the correlation between these same tests.

It was found that two subfactors (i.e. Spatial Orientation and Visualization) did not manifest themselves in this study. This was the case for both men and women:

a result which runs contrary to the findings of Barratt (1955) and Very (1967) that men would exhibit more than one spatial factor while women would exhibit only one.

Referenced previously was the idea that, depending on the problem solving strategy employed, a test could move from one factor to the other. Since two factors were not found, this hypothesis could not be tested. It was possible, however, to test whether or not a group of people who used the same strategy on each of two tests would exhibit a higher correlation between these two tests than a group which used different strategies on the two tests. Of interest in this regard was the finding that people who use different strategies for different tasks have "succeeded in developing some specialization of their abilities" (French, 1965; p. 22). This, again, was not borne out by the data. Of the 30 comparisons (10 for each of men, women, and the total group) not one was significant.

Further, the idea that use of part problem solving strategies on S tests and whole problem solving strategies on Vz tests would enhance performance (Barratt, 1953; Michael et al., 1950) received no support. Even though the distinction between S and Vz is lost for the purposes of this study, the formation of subgroups on the basis of problem solving strategy (i.e. part, whole, part/whole) was not helpful in determining who would score in a predicted direction on any tests.

Add to this the findings which largely rejected the hypotheses that perceived difficulty would be tied to problem solving strategy and that the sexes would differ in the strategies employed, one is left wondering what problem solving strategy has to do with any of the variables mentioned thus far. Of particular importance in this regard is the fact that so many investigators have documented a relationship between problem solving strategy and one or more of these variables.

One possibility which must be addressed in an attempt to explain these nonsignificant results is that the participants did not respond seriously to this aspect of the study. There are a number of findings, however, which mitigate against such a conclusion. For instance, on all tests women indicated that they guessed more. When this is considered in conjunction with the finding that men scored higher than women on all tests, it is not surprising. Also, although the relationship is not perfect, both men and women indicated that they guessed more on the tests they perceived to be more difficult. Further, there was an indication that people responded 'TRUE' to 'I didn't use any particular strategy' more often for the more difficult tests. This is also not surprising if one assumes that a person who has difficulty in solving a particular item will bring a variety of problem solving strategies to bear on that item. Thus, the evidence is not consistent with the idea that respondents did not take these items seriously.

One problem, however, is that most participants tended to answer in the same direction on a number of statements designed to tap part versus whole problem solving strategy. On two of these (one from Paper Folding, one from Cubes) over 80% of the respondents answered 'TRUE' while on four (one each from Cards and Surface Development, two from Cubes) over 80% answered 'FALSE'. With the exception of Surface Development, these response patterns led to very few people being included in the part subgroups (i.e. never more than 10%) as compared to the whole and part/whole subgroups. This, then, resulted in a loss of power to detect differences among groups (Hays, 1973).

One other major concern with the assessment of problem solving strategy is the introspective nature of the task. For instance, response to the strategy statements may require a person to verbalize some higher order cognitive processes and it has been suggested that people have little or no introspective access to these processes (Nisbett & Wilson, 1977). Thus, it has been proposed that when people actually do attempt to report on cognitive processes they may, in fact, base their reports on the extent to which a stimulus (i.e. a strategy statement) is a plausible explanation for some behaviour (i.e. solving a spatial item). An example of this may be afforded by the strategy statements for the Cubes test.

Although a significant portion of French's (1965) sample agreed to having used either geometrical terms, or rotating the cube on separate axes, or both, only 6.8% of the present sample indicated that they 'had an X and Y axis in mind when solving the items'. On the other hand, 89.5% indicated agreement with Strategy 1 (i.e. comparing symbols).

It has also been suggested that introspective reports are, to some extent, based on social desirability (Di Vesta, Ingersoll, & Sunshine, 1971). Thus, for instance, if people believe that solving items in one fashion is more valued than solving them in some other fashion, their responses might not be indicative of the strategies they actually employed.

Thus, the proposals of both Nisbett and Wilson (1977) and Di Vesta et al. (1971) may relate to why so many people responded in the same way on many of the problem solving statements included in this study. Since there is no way to test either of these assertions, however, it must be concluded that the failure of problem solving strategy to relate to other variables in this study cannot be readily explained. Therefore, in what follows, problem solving strategy will not be discussed further. It is suggested, however, that if one wishes to pursue this matter it would be preferable not to supply participants with problem solving strategies in order to, if nothing else, avoid the problem of experimenter induced social desirability of response (Barratt, 1953).

As concerns sex differences, the similarities were more striking than the differences. Although men scored higher than women on all tests (not just Vz as predicted), large differences in the correlations between tests did not manifest themselves. For instance, the largest discrepancy in the test score matrices was .14. Similarly, the largest sex difference for correlations between test difficulty and test score was .18. Also, with the exception of one reversal on the two most difficult tests, the sexes rank ordered the tests on the basis of difficulty in the same way. Finally, even though men scored higher than women on all tests, women and men did not differ significantly on the difficulty rating of three of the tests.

The last point, then, leads one to a discussion of the difficulty index. Why, it may be asked, does difficulty rating correlate substantially with test performance (for the total group as well as for the sexes separately) when it does not necessarily differentiate the sexes? Although the answer is by no means obvious, one explanation may be that any one individual considers the difficulty of the tests as a group before rating any one test individually. This might result in what could be seen as a rank ordering of the tests. If, for instance, a person considers that the tests as a group are moderately easy/difficult, each test may be compared to this difficulty index before being rated. The actual scale point of the difficulty rating may not be of the utmost importance, but rather the numerical rating of one test compared to another. This, then, leads one to conclude that the difficulty of a test (as measured in this study) is a relative matter and will be influenced by other tests included for study. Thus, for example, the average difficulty rating of the Cards tests might change from 4.0 in a study comprised solely of spatial tests to 7.0 in a study comprised of trivial arithmetic and verbal tests.

If this is the case then it is an empirical matter to determine whether difficulty rating correlates with performance when tests of different abilities are included in the same study.

Also in relation to perceived difficulty, the factor analytic results of this study do not support the findings of Guilford (1941) that the first factor had loadings which were inversely proportional to difficulty indices. A possible explanation is that, in this study, contrary to the studies of Guilford (1941) and Ferguson (1941), different people perceived different tests to be more or less difficult. For instance, 45% of the men perceived Cubes to be more difficult than Paper Folding while 38% perceived Paper Folding to be more difficult than Cubes. The results were similar for women. Thus, there seem to be large individual differences in how difficult one test is perceived to be in relation to another. Further support for this contention is supplied by an analysis of the correlations between the difficulty ratings of the various tests. The largest correlations in these matrices were, for men and women, respectively, .30 and .22. Thus, while it may be a simple matter to say that it is more difficult to detect frequency differences of .5 cycles than 30 cycles, it is quite another matter to say that S tests, for instance, are less difficult than Vz tests (Ekstrom et al., 1976; Michael et al., 1950). It is probably for this reason that the correlation matrices among difficulty ratings were not similar to the correlation matrices among test scores.

That difficulty is, indeed, an important variable is attested to by the findings that the correlation between two tests can be affected by how discrepant the difficulty ratings of these two tests are. The findings related to these correlational differences only used information about the difficulty of one test relative to the other. That is, no consideration was given to how difficult the two tests actually were.

To be noted, again, is that this study involved tests of only one general ability. Thus, it can only be suggested that this same research strategy be applied to studies involving two, and possibly more, distinct abilities. It may well be the case that the individual differences in perceived <u>relative</u> difficulty would, in such a study, be lessened. It would then be a simple matter to compare the correlations among difficulty ratings to the test score correlation matrices. It is hypothesized that, with tests from different domains, these matrices would be much more similar than the ones in the present study. This is, of course, based on the finding that as the difference in perceived difficulty of two tests increases, the correlation between the two may decrease. It is possible that a study of this nature would explain, or clarify, why a difficult vocabulary test can load on a spatial factor (French, 1951). Perhaps the structure of intellect is inextricably tied to the structure of test difficulty.

It may be concluded that subjective difficulty ratings are valuable in that they provide an operational definition for the difficulty of speeded tests. This is evidenced by the finding that test difficulty correlates, at times substantially, with total test score. It would be interesting to know if this finding generalizes to tests of other ability domains (e.g. reasoning, numerical) and if it makes any difference whether all tests measure the same ability.

The findings related to problem solving are unfortunate; not because of their insignificance, but because the reason for their insignificance is not clear. Although it seems that difficulty is not related to how spatial items are solved, it may be that the measures used in this study were not sensitive enough to point out any relationship which may exist.

Finally, the relationship between sex differences in performance and sex differences in difficulty are enlightening. Although men scored higher on all tests, sex differences in difficulty ratings were found on only two of the tests. It would be interesting to know the pattern of difficulty ratings from a study which included tests of two abilities; one in which women excel (e.g. verbal) and one in which men excel (e.g. spatial).

The use of subjective test difficulty ratings as difficulty indices for speeded tests is strongly recommended. The ratings are both easy to obtain and seem to be valid measures. Since they have not, to the author's knowledge, been used previously their potential is unknown.

References

Anastasi, A. On the formation of psychological traits. _American Psychologist_, 1970, _25_, 899-910.

Barratt, E.S. An analysis of verbal reports of solving spatial problems as an aid in defining spatial factors. _Journal of Psychology_, 1953, _36_, 17-25.

Barratt, E.S. The space-visualization factors related to temperament factors. _Journal of Psychology_, 1955, _39_, 279-287.

Borich, G.D., & Bauman, P.M. Convergent and discriminant validation of the French and Guilford-Zimmerman spatial orientation and spatial visualization factors. _Educational and Psychological Measurement_, 1972, _32_, 1029-1033.

Di Vesta, F.J., Ingersoll, G., & Sunshine, P. A factor analysis of imagery tests. _Journal of Verbal Learning and Verbal Behavior_, 1971, _10_, 471-479.

Ekstrom, R.B., French, J.W., & Harman, H.H. _Manual for Kit of Factor Referenced Cognitive Tests_. Princeton: Educational Testing Service, 1976.

Ferguson, G.A. The factorial interpretation of test difficulty. _Psychometrika_, 1941, _6_, 323-329.

Ferguson, G.A. Statistical Analysis in Psychology and Education (4th ed.). New York: McGraw-Hill, 1976.

Fleishman, J.J., & Dusek, E.R. Reliability and learning factors associated with cognitive tests. Psychological Reports, 1971, 29, 523-530.

French, J.W. The description of aptitude and achievement tests in terms of rotated factors. Psychometric Monographs (No. 5). Chicago: University of Chicago Press, 1951.

French, J.W. The relationship of problem-solving styles to the factor composition of tests. Educational and Psychological Measurement, 1965, 25, 9-28.

French, J.W., Ekstrom, R.B., & Price, L.A. Kit of Reference Tests for Cognitive Factors. Princeton: Educational Testing Service, 1963.

Fruchter, B. The nature of verbal fluency. Educational and Psychological Measurement, 1948, 8, 33-47.

Garai, J.E., & Scheinfeld, A. Sex differences in mental and behavioral traits. Genetic Psychology Monographs, 1968, 77, 169-299.

Guilford, J.P. The difficulty of a test and its factor composition. Psychometrika, 1941, 6, 67-77.

Guilford, J.P., & Lacey, J.I. Printed Classification Tests, A.A.F. (Aviation Psychological Progress Research Report, No. 5). Washington: U.S. Government Printing Office, 1947.

Guilford, J.P., & Zimmerman, W.S. Guilford-Zimmerman Aptitude Survey. Orange, California: Sheridan Psychological Services, 1953.

Harman, H.H. Modern Factor Analysis (3rd ed.). Chicago: University of Chicago Press, 1976.

Hakstian, A.R., & Muller, V.J. Some notes on the number of factors problem. Multivariate Behavioral Research, 1973, 8, 461-475.

Harris, L.J. 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, 1978.

Hays, W.L. Statistics for the Social Sciences (2nd ed.). New York: Holt, Rinehart and Winston, 1973.

Koussy, A.A.H. The visual perception of space. British Journal of Psychology (Monograph Supplement), 1935, No. 20.

Maccoby, E.E., & Jacklin, C.N. The Psychology of Sex Differences. Stanford, California: Stanford University Press, 1974.

McFarlane, M. A study of practical ability. British Journal of Psychology (Monograph Supplement), 1925, No. 8.

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

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

Michael, W.B., Zimmerman, W.S., & Guilford, J.P. An investigation of two hypotheses regarding the nature of spatial-relations and visualization. Educational and Psychological Measurement, 1950, 10, 187-213.

Morrison, D.F. Multivariate Statistical Methods (2nd ed.). New York: McGraw-Hill, 1976.

Nisbett, R.E., & Wilson, T.D. Telling more than we can know: Verbal reports on mental processes. Psychological Review, 1977, 84, 231-259.

Nunnally, J.C. Psychometric Theory (2nd ed.). New York: McGraw-Hill, 1978.

Sargent, S.S. Thinking processes at various levels of difficulty: A quantitative and qualitative study of individual differences. Archives of Psychology, 1940, No. 249.

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

Smith, I.M. Spatial Ability: Its Educational and Social Significance. London: University of London Press, 1964.

Thurstone, L.L. Primary Mental Abilities. Chicago: University of Chicago Press, 1938.

Thurstone, L.L. Some primary abilities in visual thinking. University of Chicago Psychometric Laboratory Report, 1950, No. 59.

Very, P.S. Differential factor structures in mathematical ability. Genetic Psychology Monographs, 1967, 75, 169-207.

Werdelin, I. Geometrical Ability and the Space Factors in Boys and Girls. Lund, Sweden: University of Lund Press, 1961.

Werdelin, I., & Stjernberg, G. On the nature of the perceptual speed factor. Scandanavian Journal of Psychology, 1969, 10, 185-192.

Wolfle, D. Factor analysis to 1940. Psychometric Monographs (No. 3). Chicago: University of Chicago Press, 1940.

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

Tables

Table 1

Means and Standard Deviations for Spatial Test Scores and Difficulty Ratings

Total Group

	Test Score	Difficulty Rating
Form Board		
Mean	54.59	6.49
Standard Deviation	19.56	1.71
Number of People	537	537
Paper Folding		
Mean	6.19	4.50
Standard Deviation	2.46	1.81
Number of People	537	537
Surface Development		
Mean	17.12	6.42
Standard Deviation	8.45	2.18
Number of People	537	537
Card Rotations		
Mean	50.89	3.38
Standard Deviation	16.79	1.86
Number of People	537	536
Cube Comparisons		
Mean	7.81	4.79
Standard Deviation	5.10	1.75
Number of People	537	534

Table 1 continued

Men

	Test Score	Difficulty Rating
Form Board		
Mean	60.46	6.52
Standard Deviation	19.37	1.68
Number of People	266	266
Paper Folding		
Mean	6.88	4.26
Standard Deviation	2.24	1.81
Number of People	266	266
Surface Development		
Mean	20.71	5.49
Standard Deviation	7.67	2.17
Number of People	266	266
Card Rotations		
Mean	55.50	3.17
Standard Deviation	15.48	1.84
Number of People	266	265
Cube Comparisons		
Mean	9.33	4.42
Standard Deviation	5.13	1.70
Number of People	266	264

Table 1 continued

Women

	Test Score	Difficulty Rating
Form Board		
Mean	48.60	6.44
Standard Deviation	17.88	1.75
Number of People	268	268
Paper Folding		
Mean	5.50	4.74
Standard Deviation	2.48	1.79
Number of People	268	268
Surface Development		
Mean	13.56	7.37
Standard Deviation	7.68	1.75
Number of People	268	268
Card Rotations		
Mean	46.19	3.60
Standard Deviation	16.73	1.86
Number of People	268	268
Cube Comparisons		
Mean	6.26	5.15
Standard Deviation	4.59	1.74
Number of People	268	268

Table 2

Correlations between Difficulty Ratings

and Spatial Test Scores[1]

Total Group

		Difficulty Rating				
		Vz1	Vz2	Vz3	S1	S2
	Vz1	-.23	-.22	-.37	-.09	-.07
	Vz2	.05	-.52	-.40	-.11	-.08
Test Score	Vz3	.15	-.28	-.68	-.11	-.09
	S1	.10	-.12	-.30	-.45	-.19
	S2	.13	-.08	-.30	-.15	-.28
Number[2]		537	537	537	536	534

Table 2 continued

Men

Difficulty Rating

		Vz1	Vz2	Vz3	S1	S2
	Vz1	-.34	-.23	-.28	-.12	-.03
	Vz2	-.01	-.53	-.35	-.09	.07
Test Score	Vz3	.15	-.24	-.59	-.10	.07
	S1	.05	-.13	-.17	-.40	-.13
	S2	.12	.00	-.20	-.16	-.24
Number		266	266	266	265	264

Table 2 continued

Women

		Difficulty Rating				
		Vz1	Vz2	Vz3	S1	S2
	Vz1	-.16	-.16	-.29	.02	.01
	Vz2	.09	-.49	-.31	-.07	-.11
Test Score	Vz3	.16	-.27	-.64	-.03	-.06
	S1	.12	-.06	-.25	-.49	-.16
	S2	.14	-.09	-.18	-.09	-.22
Number		268	268	268	268	268

[1] Vz1 = Form Board
Vz2 = Paper Folding
Vz3 = Surface Development
S1 = Card Rotations
S2 = Cube Comparisons

[2] The correlations in each column are all based on the same number of people.

Table 3

Sex by Test Repeated Measures Analysis of Variance on Difficulty Ratings

Source	df	MS	F
Sex (S)	1	305.02	62.07***
Error between	519	4.91	
Test (T)	4	938.32	327.17***
S x T	4	70.27	24.50***
Error within	2116	2.87	

Means and Number of People

	Men	Women
Form Board	6.54	6.44
Paper Folding	4.28	4.74
Surface Development	5.50	7.37
Card Rotations	3.19	3.60
Cube Comparisons	4.41	5.15
Number of People	263	268

*** $p < .001$

Table 4

Frequency of Answers to Problem Solving Statements

	Total Group		Men		Women	
	True	False	True	False	True	False
Form Board						
Strategy 1	300	229	138	122	161	105
Strategy 2	346	184	177	84	168	98
Strategy 3	226	301	99	161	127	138
Strategy 4	276	249	127	132	147	116
Strategy 5	104	425	38	223	64	201
Strategy 6	51	479	28	233	21	245
Paper Folding						
Strategy 1	434	103	206	60	227	41
Strategy 2	116	418	57	208	56	210
Strategy 3	123	411	65	201	56	209
Strategy 4	156	374	86	174	70	197
Strategy 5	87	443	27	234	59	207
Strategy 6	34	495	18	244	15	249
Surface Development						
Strategy 1	155	381	77	189	77	190
Strategy 2	339	197	169	97	167	100
Strategy 3	68	464	29	235	37	228
Strategy 4	307	224	155	110	149	114
Strategy 5	135	401	42	223	92	176
Strategy 6	46	490	14	251	32	236
Card Rotations						
Strategy 1	70	462	40	224	30	235
Strategy 2	364	169	166	99	195	70
Strategy 3	237	296	113	152	122	143
Strategy 4	344	184	171	94	170	90
Strategy 5	49	484	11	254	37	228
Strategy 6	17	517	7	258	10	256
Cube Comparisons						
Strategy 1	478	56	233	30	242	26
Strategy 2	205	326	107	155	96	170
Strategy 3	36	496	24	238	12	255
Strategy 4	28	504	13	249	15	252
Strategy 5	91	441	25	237	65	202
Strategy 6	35	496	14	248	21	245

Table 5

Correlation between Test Scores for People who use the 'Same' and 'Different' Problem Solving Strategies

Total Group

Tests[1]	Same Strategy		Different Strategy		z-value[2]
	n[3]	r[4]	n	r	
Vz1/Vz2	115	.46	45	.63	-1.34
Vz1/Vz3	86	.39	60	.56	-1.29
Vz1/S1	73	.42	49	.26	.96
Vz1/S2	121	.38	43	.42	-.26
Vz2/Vz3	133	.57	82	.68	-1.22
Vz2/S1	123	.42	44	.30	.76
Vz2/S2	227	.38	43	.30	.58
Vz3/S1	81	.50	62	.48	.15
Vz3/S2	134	.42	78	.38	.30
S1/S2	118	.43	59	.58	-1.23

Table 5 continued

Men

Tests	Same Strategy		Different Strategy		z-value
	n	r	n	r	
Vz1/Vz2	63	.52	21	.72	-1.22
Vz1/Vz3	43	.35	29	.30	.19
Vz1/S1	41	.41	23	.28	.51
Vz1/S2	57	.29	18	.37	-.29
Vz2/Vz3	70	.52	39	.62	-.76
Vz2/S1	70	.47	24	.32	.70
Vz2/S2	100	.26	22	.30	-.19
Vz3/S1	50	.43	29	.36	.34
Vz3/S2	61	.37	35	.34	.16
S1/S2	66	.46	27	.55	-.51

Table 5 continued

Women

Tests	Same Strategy		Different Strategy		z-value
	n	r	n	r	
Vz1/Vz2	52	.35	24	.40	-.19
Vz1/Vz3	43	.38	31	.51	.66
Vz1/S1	32	.32	26	.08	.90
Vz1/S2	64	.37	25	.46	-.43
Vz2/Vz3	63	.56	43	.29	1.62
Vz2/S1	53	.30	20	.21	.36
Vz2/S2	127	.43	20	.06	1.58
Vz3/S1	31	.44	33	.40	.19
Vz3/S2	73	.41	43	.22	1.07
S1/S2	52	.30	32	.49	-.95

[1] Vz1 = Form Board
Vz2 = Paper Folding
Vz3 = Surface Development
S1 = Card Rotations
S2 = Cube Comparisons

[2] z-value for the comparison of the correlations for the same and different strategy groups.

[3] The number of people.

[4] The correlation between the two tests involved.

Table 6

Sex by Strategy Analysis of Variance for

Total Score on the Card Rotations Test

Source	df	MS	F
Sex (Se)	1	11832.42	47.37***
Strategy (St)	2	1082.26	4.33
Se x St	2	188.30	.75
Error	519	249.77	

Means and Number of People

	Whole	Part	Part/Whole
Men	52.47 (97)	58.60 (25)	57.27 (141)
Women	43.96 (67)	43.92 (25)	47.99 (170)

* $p < .05$
** $p < .01$
*** $p < .001$

Table 7

Sex by Strategy Analysis of Variance for

Total Score on the Cube Comparisons Test

Source	df	MS	F
Sex (Se)	1	1179.08	50.57***
Strategy (St)	2	24.06	1.03
Se x St	2	94.92	4.07*
Error	518	23.32	

Means and Number of People

	Whole	Part	Part/Whole
Men	8.68 (151)	8.89 (19)	10.57 (88)
Women	6.56 (170)	5.86 (21)	5.80 (75)

* $p < .05$
** $p < .01$
*** $p < .001$

Table 8

Sex by Strategy Analysis of Variance for

Total Score on the Paper Folding Test

Source	df	MS	F
Sex (Se)	1	252.52	45.92***
Strategy (St)	2	19.39	3.53*
Se x St	2	3.58	.65
Error	521	5.49	

Means and Number of People

	Whole	Part	Part/Whole
Men	6.90 (164)	5.32 (14)	7.17 (85)
Women	5.54 (177)	4.89 (11)	5.59 (76)

* $p < .05$
** $p < .01$
*** $p < .001$

Table 9

Sex by Strategy Analysis of Variance for Total Score on the Surface Development Test

Source	df	MS	F
Sex (Se)	1	6553.63	117.22***
Strategy (St)	2	308.56	5.52**
Se x St	2	287.12	5.14**
Error	512	55.91	

Means and Number of People

	Whole	Part	Part/Whole
Men	18.54 (95)	24.48 (49)	21.28 (117)
Women	13.60 (84)	13.52 (49)	14.11 (124)

* $p < .05$
** $p < .01$
*** $p < .001$

Table 10

Sex by Strategy Analysis of Variance for

Total Score on the Form Board Test

Source	df	MS	F
Sex (Se)	1	18241.85	52.20***
Strategy (St)	2	91.21	.26
Se x St	2	86.54	.25
Error	513	349.44	

Means and Number of People

	Whole	Part	Part/Whole
Men	60.81 (85)	57.78 (18)	60.66 (151)
Women	47.31 (74)	48.04 (26)	49.26 (165)

* $p < .05$
** $p < .01$
*** $p < .001$

Table 11

Sex by Strategy Analysis of Variance for
Difficulty Rating of the Form Board Test

Source	df	MS	F
Sex (Se)	1	.50	.17
Strategy (St)	2	.94	.32
Se x St	2	.06	.02
Error	513	2.91	

Means and Number of People

	Whole	Part	Part/Whole
Men	6.52 (85)	6.33 (18)	6.56 (151)
Women	6.50 (74)	6.27 (26)	6.48 (165)

* $p < .05$
** $p < .01$
*** $p < .001$

Table 12

Sex by Strategy Analysis of Variance for Difficulty Rating of the Paper Folding Test

Source	df	MS	F
Sex (Se)	1	34.82	11.08***
Strategy (St)	2	17.19	5.47**
Se x St	2	4.73	1.50
Error	521	3.14	

Means and Number of People

	Whole	Part	Part/Whole
Men	4.12 (164)	5.86 (14)	4.22 (85)
Women	4.66 (177)	5.18 (11)	4.87 (76)

* $p < .05$
** $p < .01$
*** $p < .001$

Table 13

Sex by Strategy Analysis of Variance for

Difficulty Rating of the Surface Development Test

Source	df	MS	F
Sex (Se)	1	458.13	120.02***
Strategy (St)	2	1.93	.51
Se x St	2	17.91	4.69**
Error	512	3.82	

Means and Number of People

	Whole	Part	Part/Whole
Men	5.78 (95)	4.80 (49)	5.44 (117)
Women	7.13 (84)	7.65 (49)	7.31 (124)

* $p < .05$
** $p < .01$
*** $p < .001$

Table 14

Sex by Strategy Analysis of Variance for
Difficulty Rating of the Card Rotations Test

Source	*df*	*MS*	*F*
Sex (Se)	1	20.30	6.05*
Strategy (St)	2	.85	.25
Se x St	2	.97	.28
Error	518	3.36	

Means and Number of People

	Whole	Part	Part/Whole
Men	3.24 (97)	3.08 (25)	3.11 (140)
Women	3.55 (67)	3.84 (25)	3.49 (170)

* $p < .05$
** $p < .01$
*** $p < .001$

Table 15

Sex by Strategy Analysis of Variance for Difficulty Rating of the Cube Comparisons Test

Source	df	MS	F
Sex (Se)	1	78.05	27.14***
Strategy (St)	2	2.20	.77
Se x St	2	8.17	2.84
Error	516	2.88	

Means and Number of People

	Whole	Part	Part/Whole
Men	4.44 (149)	4.42 (19)	4.25 (88)
Women	4.94 (170)	5.29 (21)	5.53 (75)

* $p < .05$
** $p < .01$
*** $p < .001$

Table 16

Correlations between Spatial Test Scores[1]

Total Group (n = 537)

	Vz1	Vz2	Vz3	S1	S2
Vz1	1.00	.52	.52	.39	.36
Vz2		1.00	.59	.41	.36
Vz3			1.00	.48	.47
S1				1.00	.48
S2					1.00

Table 16 continued

Men (n = 266)

	Vz1	Vz2	Vz3	S1	S2
Vz1	1.00	.51	.39	.35	.33
Vz2		1.00	.53	.36	.26
Vz3			1.00	.40	.44
S1				1.00	.48
S2					1.00

Table 16 continued

Women (n = 268)

	Vz1	Vz2	Vz3	S1	S2
Vz1	1.00	.44	.53	.31	.26
Vz2		1.00	.55	.36	.34
Vz3			1.00	.43	.36
S1				1.00	.38
S2					1.00

[1] Vz1 = Form Board
Vz2 = Paper Folding
Vz3 = Surface Development
S1 = Card Rotations
S2 = Cube Comparisons

Table 17

Unrotated and Rotated Component Pattern Matrices

Total Group (n = 537)

	Unrotated		Rotated	
Tests[1]	Factor I	Factor II	Factor I	Factor II
Vz1	.74	-.38	.81	.19
Vz2	.77	-.38	.83	.20
Vz3	.82	-.12	.71	.44
S1	.72	.42	.29	.78
S2	.70	.53	.19	.86

Eigenvalues and Percentage of Variance

1.	2.84	56.8%
2.	.75	15.1%
3.	.52	10.4%
4.	.49	9.9%
5.	.39	7.9%

Table 17 continued

Men (n = 266)

Tests	Unrotated Factor I	Unrotated Factor II	Rotated Factor I	Rotated Factor II
Vz1	.71	-.36	.77	.19
Vz2	.74	-.48	.88	.13
Vz3	.77	-.07	.63	.46
S1	.71	.41	.26	.78
S2	.68	.56	.14	.87

Eigenvalues and Percentage of Variance

1.	2.63	52.5%
2.	.85	17.1%
3.	.61	12.2%
4.	.53	10.5%
5.	.39	7.7%

Table 17 continued

Women (n = 268)

Tests	Unrotated Factor I	Unrotated Factor II	Rotated Factor I	Rotated Factor II
Vz1	.71	-.46	.85	.06
Vz2	.76	-.20	.73	.30
Vz3	.82	-.20	.77	.34
S1	.68	.40	.29	.73
S2	.62	.60	.13	.86

Eigenvalues and Percentage of Variance

1.	2.60	52.1%
2.	.82	16.3%
3.	.62	12.4%
4.	.54	10.9%
5.	.42	8.4%

[1] Vz1 = Form Board
Vz2 = Paper Folding
Vz3 = Surface Development
S1 = Card Rotations
S2 = Cube Comparisons

Table 18

Correlation between Test Scores for People with High and Low Differences in Difficulty Ratings

Total Group

Tests[1]	Value[2]	Low Difference		High Difference		z-value[3]
		n[4]	r[5]	n	r	
Vz1/Vz2	2	290	.58	247	.48	1.55
Vz1/Vz3	1	236	.58	301	.48	1.60
Vz1/S1	3	299	.49	237	.39	1.45
Vz1/S2	2	311	.37	224	.42	-.67
Vz2/Vz3	2	268	.68	269	.53	2.82**
Vz2/S1	1	252	.53	284	.35	2.61**
Vz2/S2	1	269	.51	266	.22	3.89***
Vz3/S1	3	284	.66	252	.41	4.00***
Vz3/S2	2	287	.54	248	.41	1.97*
S1/S2	1	233	.48	301	.50	-.30

Table 18 continued

Men

Tests	Value	Low Difference		High Difference		z-value
		n	r	n	r	
Vz1/Vz2	2	139	.57	127	.50	.78
Vz1/Vz3	1	114	.48	152	.31	1.60
Vz1/S1	3	141	.49	124	.32	1.72*
Vz1/S2	2	144	.38	120	.38	-.03
Vz2/Vz3	2	152	.62	114	.44	1.99*
Vz2/S1	1	126	.51	139	.26	2.37**
Vz2/S2	1	122	.54	142	.03	4.56***
Vz3/S1	2	125	.58	140	.35	2.36**
Vz3/S2	2	160	.55	104	.31	2.32*
S1/S2	1	124	.51	139	.47	.39

Table 18 continued

Women

Tests	Value	Low Difference		High Difference		z-value
		n	r	n	r	
Vz1/Vz2	2	150	.52	118	.36	1.56
Vz1/Vz3	1	121	.60	147	.50	1.11
Vz1/S1	2	117	.43	151	.30	1.14
Vz1/S2	2	166	.25	102	.32	-.65
Vz2/Vz3	2	115	.67	153	.50	2.16*
Vz2/S1	1	124	.53	144	.27	2.53**
Vz2/S2	1	146	.41	122	.29	1.10
Vz3/S1	3	115	.64	153	.44	2.21*
Vz3/S2	2	124	.41	144	.35	.58
S1/S2	1	108	.31	160	.44	-1.19

[1] Vz1 = Form Board
Vz2 = Paper Folding
Vz3 = Surface Development
S1 = Card Rotations
S2 = Cube Comparisons

[2] This value was used to define the low and high difference groups. For any pair of tests, if the absolute value of the difference in difficulty rating of the two tests was less than or equal to this value a person would be in the low difference group.

[3] z-value for the comparison of the correlations for the low and high difference groups.
* $p < .05$, one tailed
** $p < .01$, one tailed
*** $p < .001$, one tailed

[4] The number of people.

[5] The correlation between the two tests involved.

Appendix A

Description of Spatial Tests

Card Rotations Test (S1)

"Suggested by Thurstone's Cards. Each item gives a drawing of a card cut into an irregular shape. To its right are....(eight)....other drawings of the same card, sometimes merely rotated and sometimes turned over to its other side. The subject indicates whether or not the card has been turned over" (Ekstrom et al., 1976, p. 150).

Length: 3 minutes for 10 items (80 possible answers).

Cube Comparisons Test (S2)

"Suggested by Thurstone's Cubes. Each item presents two drawings of a cube. Assuming no cube can have two faces alike, the subject is to indicate which items present drawings that can be of the same cube and which present drawings that cannot be of the same cube" (Ekstrom et al., 1976, p. 150).

Length: 3 minutes for 21 items.

Form Board Test (Vz1)

"Each item presents 5 shaded drawings of pieces, some or all of which can be put together to form a figure presented in outline form. The task is to indicate which of the pieces, when fitted together, would form the outline" (Ekstrom et al., 1976, p. 174).

Length: 8 minutes for 24 items (120 possible answers).

Paper Folding Test (Vz2)

"Suggested by Thurstone's Punched Holes. For each item successive drawings illustrate two or three folds made in a square sheet of paper. The final drawing of the folded paper shows where a hole is punched in it. The subject selects one of 5 drawings to show how the punched sheet would appear when fully reopened" (Ekstrom et al., 1976, p. 176).

Length: 3 minutes for 10 items.

Surface Development Test (Vz3)

"Suggested by Thurstone's test of the same name. In this test, drawings are presented of solid forms that could be made with paper or sheet metal. With each drawing there is a diagram showing how a piece of paper might be cut and folded so as to make the solid form. Dotted lines show where the paper is folded. One part of the diagram is marked to correspond to a marked surface in the drawing. The subject is to indicate which lettered edges in the drawing correspond to numbered edges or dotted lines in the diagram" (Ekstrom et al., 1976, p. 174).

Length: 5 minutes for 6 items (30 possible answers).

Appendix B

Difficulty Rating Questionnaire

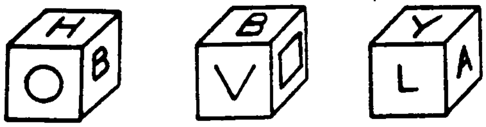
On the following page use the scale presented below to indicate how difficult you found each of the tests you just completed.

For instance, if you found a test to be very easy, circle 1. If you found it to be very difficult, circle 9. If you found it to be moderately easy/difficult, circle 5.

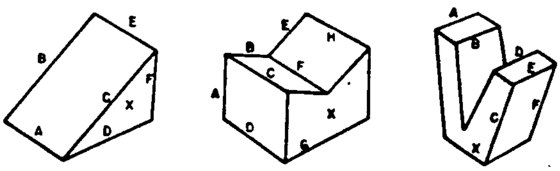
Circle the number on the scale which reflects how difficult you found the test.

1	2	3	4	5	6	7	8	9
very easy				moderately easy/difficult				very difficult

1	2	3	4	5	6	7	8	9
very easy				moderately easy/difficult				very difficult

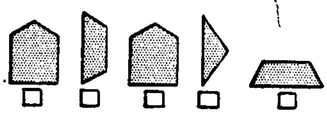


1	2	3	4	5	6	7	8	9
very easy				moderately easy/difficult				very difficult



1	2	3	4	5	6	7	8	9
very easy				moderately easy/difficult				very difficult

1	2	3	4	5	6	7	8	9
very easy				moderately easy/difficult				very difficult



1	2	3	4	5	6	7	8	9
very easy				moderately easy/difficult				very difficult

Appendix C

Strategy Questionnaire

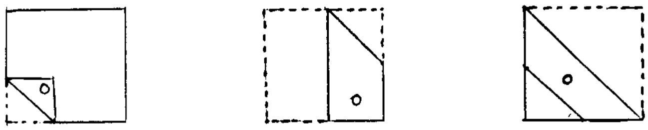
Listed on the following pages are a number of questions related to how you solved the items in the tests you just completed.

At the top of each page a number of the items from each test are reproduced. With these in mind think back to when you were working on the test in question.

Then, keep in mind how you <u>generally</u> solved the items and circle TRUE for those statements which describe how you solved the problems and FALSE for those statements which do not.

NOTE: Do not respond on the basis of how you solved one or two of the items.

RESPOND ON THE BASIS OF HOW YOU SOLVED THE ITEMS IN GENERAL.



I mentally folded the paper, punched a hole, and then mentally unfolded it and compared it to the possible answers. TRUE FALSE

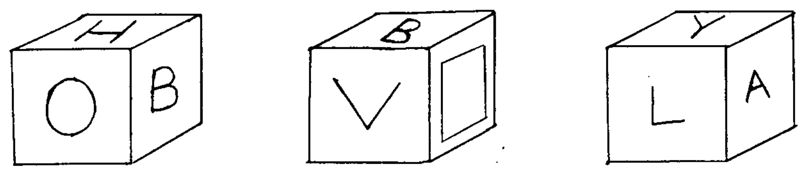
I noted where one or two of the holes would be and then tried to find an answer that had these one or two holes in the same place (i.e. I didn't actually fold and unfold the paper). TRUE FALSE

I tried to determine through how many thicknesses of paper the hole would have to go and then picked one of the answers that had that many holes (i.e. I didn't actually unfold the paper with the holes). TRUE FALSE

I worked backwards; I looked at the drawing with the hole and then unfolded it mentally (i.e. I didn't actually fold the paper up to begin with). TRUE FALSE

I guessed when I wasn't sure. TRUE FALSE

I didn't use any particular strategy. TRUE FALSE



I mentally picked up one of the cubes and rotated it to line it up with a certain symbol on the other one and then checked whether or not the other symbols were in the right place. TRUE FALSE

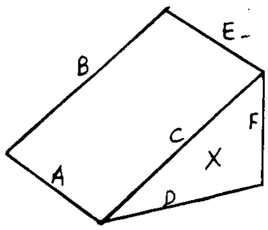
Without mentally picking up the cube, I determined if any two of the symbols on both cubes bore the same relation to one another. TRUE FALSE

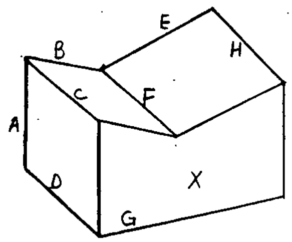
I had an X and Y axis in mind when solving the items. TRUE FALSE

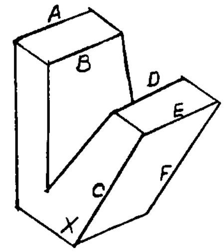
I pictured myself as walking around the cubes. TRUE FALSE

I guessed when I wasn't sure. TRUE FALSE

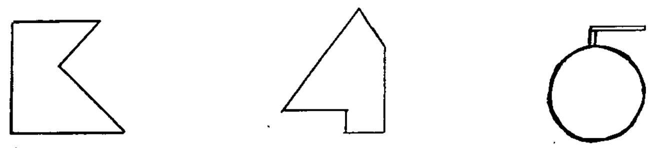
I didn't use any particular strategy. TRUE FALSE







I mentally folded up the figure and mentally placed it on top of the figure already folded up.	TRUE	FALSE
I mentally folded up the figure and then, without necessarily placing it on top of the figure already folded up, I tried to match the letters and numbers.	TRUE	FALSE
I tried to match the numbers and letters up without ever doing any mental folding.	TRUE	FALSE
I chose a number I wanted to find a letter for and then mentally folded only those parts of the figure that had to do with that number.	TRUE	FALSE
I guessed when I wasn't sure.	TRUE	FALSE
I didn't use any particular strategy.	TRUE	FALSE



I noted that some of the figures looked like letters of the alphabet or familiar objects and solved them with this in mind. TRUE FALSE

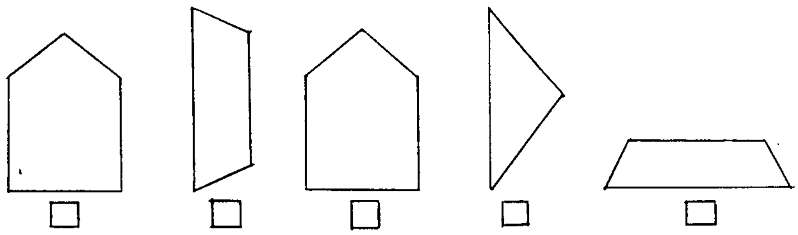
I noted some distinct aspect of the figure (eg. bottom part or top left corner) and solved the figures with this feature in mind. TRUE FALSE

I mentally rotated a figure and mentally placed it on top of the other figures to see if they were the same. TRUE FALSE

I mentally rotated one of the figures and then, without necessarily placing one figure on top of the other, compared the two. TRUE FALSE

I guessed when I wasn't sure. TRUE FALSE

I didn't use any particular strategy. TRUE FALSE



I mentally placed one item after the other into the figure until it was full.	TRUE	FALSE
I kept the large figure in mind and tried to arrange the smaller pieces to fit it.	TRUE	FALSE
I used a pencil to draw the large figure out of the smaller ones.	TRUE	FALSE
I used trial and error.	TRUE	FALSE
I guessed when I wasn't sure.	TRUE	FALSE
I didn't use any particular strategy.	TRUE	FALSE

Appendix D

Rationale for including Strategy Statements

For each test six strategy statements were included. Two of these (the last two) were the same for each test. 'I didn't use any particular strategy' was included to ensure that people who felt that none of the other strategies applied to them would, in fact, have a reasonable choice.

'I guessed when I wasn't sure' was included on the strength of Barratt's (1953) finding that in solving the items of the Space subtest of the Differential Aptitude Test, some people could not solve the items and relied on guessing.

The remaining problem solving statements are all different and are designed to tap either part or whole problem solving strategies. If an answer of 'TRUE' to a statement implied that the whole figure was kept in mind, the item was deemed to be a whole problem solving strategy. If an answer of 'TRUE' implied that only a certain aspect or part of a stimulus figure was kept in mind to attain solution of the problem, the item was considered to be a part problem solving strategy.

<u>Card Rotations Test (S1)</u>

Answering 'TRUE' to 'I noted that some of the figures looked like letters of the alphabet or familiar objects and solved them with this in mind' (Strategy 1) is a <u>whole</u> problem solving strategy. This statement was included because of Barratt's (1953) finding that people who, on the Figures test, attempted to relate the items to something familiar scored lower than people who used a part strategy.

Answering 'TRUE' to 'I noted some distinct aspect of the figure (e.g. bottom part or top left corner) and solved the figures with this feature in mind' (Strategy 2) is a part problem solving strategy, while answering 'TRUE' to 'I mentally rotated a figure and mentally placed it on top of the other figures to see if they were the same' (Strategy 3) is a whole problem solving strategy. Both items were suggested by Barratt's (1953) finding that people who rotated only part of a figure scored higher than people who rotated the whole figure.

Answering 'TRUE' to 'I mentally rotated one of the figures and then, without necessarily placing one figure on top of the other, compared the two' (Strategy 4) is a whole problem solving strategy. Although the statement was suggested by Barratt (1953), it was primarily included in order to determine how people solved the Cards items, given that they were using a whole approach.

Cube Comparisons Test (S2)

Answering 'TRUE' to 'I mentally picked up one of the cubes and rotated it to line it up with a certain symbol on the other one and then checked whether not the other symbols were in the right place' (Strategy 1) is a whole problem solving strategy. The rationale for including this statement is the same as that for including Strategy 4 for Cards.

Answering 'TRUE' to 'Without mentally picking up the cube, I determined if any two of the symbols on both cubes bore the same relation to one another' (Strategy 2) is a part problem solving strategy. This statement was suggested by Strategy 2 from Cards in that only part of the stimulus is considered.

Answering 'TRUE' to 'I had an X and Y axis in mind when solving the items' (Strategy 3) is a whole problem solving strategy. The statement was suggested by French (1965) who noted that some people reported solving the Cubes items by mentally rotating the cubes on two separate axes.

Answering 'TRUE' to 'I pictured myself as walking around the cubes' (Strategy 4) is a whole problem solving strategy. This statement is based on the idea that people who picture themselves as walking around the cubes use a spatial orientation strategy (Michael et al., 1950), while those who mentally pick up and rotate a cube use a visualization process (Strategy 1).

Form Board Test (Vz1)

Inspection of the stimuli for this test might suggest that, since every item consists of five distinct pieces, a part strategy would be the only possible one. However, consideration must also be given to the figure at the top of each column which was to be completed, or filled in, by some subset of the five distinct pieces. Thus, statements which indicated that this figure was kept in mind as a whole were termed whole problem solving strategies. If, however, the figure was not considered mentally (i.e. using the actual figure rather than a mental representation of it) it was assumed that a part strategy was used.

Due to the lack of published reports relating to problem solving strategy for this test, the above rationale for including the various statements applies to all statements.

Answers of 'TRUE' to 'I mentally placed one item after the other into the figure until it was full' (Strategy 1) or 'I kept the large figure in mind and tried to arrange the smaller pieces to fit it' (Strategy 2) were <u>whole</u> problem solving strategies.

Answers of 'TRUE' to 'I used a pencil to draw the large figure out of the smaller ones' (Strategy 3) or 'I used trial and error' (Strategy 4) were <u>part</u> problem solving strategies.

<u>Paper Folding Test (Vz2)</u>

Answers of 'TRUE' to 'I mentally folded the paper, punched a hole, and then mentally unfolded it and compared it to the possible answers' (Strategy 1) or 'I worked backwards; I looked at the drawing with the holes and then unfolded it mentally (i.e. I didn't actually fold the paper up to begin with)' (Strategy 4) were whole problem solving strategies.

Answers of 'TRUE' to 'I noted where one or two of the holes would be and then tried to find an answer that had these one or two holes in the same place (i.e. I didn't actually fold and unfold the paper)' (Strategy 2) or 'I tried to determine through how many thicknesses of paper the hole would have to go and then picked one of the answers that had that many holes (i.e. I didn't actually unfold the paper with the holes)' (Strategy 3) were part problem solving strategies.

All statements were suggested by the problem solving strategies listed by Barratt (1953) for the Space subtest of the Differential Aptitude Test. Strategies 2 and 3 were meant to tap whether people were merely looking for cues (i.e. position of holes, number of thicknesses); and Strategies 1 and 4 involved folding and/or unfolding of figures.

Surface Development Test (Vz3)

Answers of 'TRUE' to 'I mentally folded up the figure and mentally placed it on top of the figure already folded up' (Strategy 1) or 'I mentally folded up the figure and then, without necessarily placing it on top of the figure already folded up, I tried to match the letters and numbers' (Strategy 2) were whole problem solving strategies.

Answers of 'TRUE' to 'I tried to match the numbers and letters up without ever doing any mental folding' (Strategy 3) or 'I chose a number I wanted to find a letter for and then mentally folded only those parts of the figure that had to do with that number' (Strategy 4) were part problem solving strategies.

These statements were included on the strength of the Michael et al. (1950) hypothesis that if entire figures are kept in mind (Strategies 1 and 2) a test would load on a visualization factor, whereas if only salient aspects of the figures are considered (Strategies 3 and 4) the test would load a spatial orientation factor.