AD-A075 972 STANFORD UNIV CALIF SCHOOL OF EDUCATION F/G 5/10

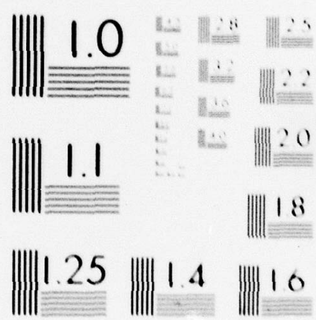
SPATIAL ABILITY: A REVIEW AND REANALYSIS OF THE CORRELATIONAL L--ETC(U)

OCT 79 D F LOHMAN. N00014-75-C-0882

UNCLASSIFIED TR-8 NL

1 OF 3

AD
A075972



MICROCOPY RESOLUTION TEST CHART

LEVEL II

AD A075972

SPATIAL ABILITY: A REVIEW AND REANALYSIS OF THE CORRELATIONAL LITERATURE

DAVID F. LOHMAN

TECHNICAL REPORT NO. 8
APTITUDE RESEARCH PROJECT
SCHOOL OF EDUCATION
STANFORD UNIVERSITY

Sponsored by

Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

and

Advanced Research Projects Agency

under

Contract No. N00014-75-C-0882

OCTOBER 1979

(12)

SPATIAL ABILITY: A REVIEW AND REANALYSIS OF THE CORRELATIONAL LITERATURE

DAVID F. LOHMAN

TECHNICAL REPORT NO. 8
APTITUDE RESEARCH PROJECT
SCHOOL OF EDUCATION
STANFORD UNIVERSITY

Sponsored by

Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

and
Advanced Research Projects Agency

under

Contract No. N00014-75-C-0882

OCTOBER 1979

14 TR-8

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 8	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 9
4. TITLE (and Subtitle) 6 SPATIAL ABILITY: A Review and Reanalysis of the Correlational Literature.		5. TYPE OF REPORT & PERIOD COVERED Technical Report #8
		6. PERFORMING ORG. REPORT NUMBER 8
7. AUTHOR(s) 10 David F. Lohman		8. CONTRACT OR GRANT NUMBER(s) 15 N00014-75-C-0882
9. PERFORMING ORGANIZATION NAME AND ADDRESS School of Education Stanford University Stanford, California 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 154-376
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Program Psychological Sciences Division, ONR, 458		12. REPORT DATE 11 October 1979
		13. NUMBER OF PAGES 204
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 12 226		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

UNLIMITED

18. SUPPLEMENTARY NOTES

This research was jointly sponsored by the Office of Naval Research and the Defense Advanced Research Projects Agency.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Spatial ability, solution strategies, speed, level, hierarchical models, multidimensional scaling, factor analysis, spatial factors, complexity dimension, spatial tests.

(of this report)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report is divided into four sections. The first and longest part reviews and reanalyzes the major American factor analytic studies of spatial ability in terms of a hierarchical model of ability organization. The most important questions are "What are the major dimensions of individual differences in spatial ability?" and "What are the implications of this research for a process understanding of spatial ability?" (over)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

391 909

JOB

(cont)

The second section ~~of this report~~ examines the effects of alternative solution strategies used by subjects on spatial tests. Some of the major confusions in the factor analytic studies are shown to result from individuals solving spatial problems in different ways. In addition to reviewing the literature on this topic, some new data are presented and discussed.

The third section reviews the relationship between speed, level, and complexity in test performance. The speed-level dimension is shown to be crucially important for all factor analytic work, particularly for the distinction between broad, general factors and narrow specifics. Implications of this research for chronometric studies of spatial ability are also discussed. It is argued that individual differences in speed may be largely independent of individual differences in level, especially on spatial tasks.

Finally, the fourth section summarizes the conclusions and implications of the review for future research on spatial ability.

PREFACE

This review was initially conceived as a short summary. We were planning some experimental studies on individual differences in spatial thinking, and merely wanted to understand the endpoints of the factor analytic research on spatial ability. But in attempting to summarize this work, it quickly became apparant that there were as many endpoints as investigators. The only way to integrate the research was to reanalyze the studies from a common theoretical perspective. The most important questions this review attempts to answer are: "What are the major dimensions of individual differences in spatial ability?" and "What are the implications of this research for a process understanding of spatial ability?"

It is appropriate that this review be issued as an ONR Technical Report, as many of the studies reanalyzed herein were sponsored by the Office of Naval Research (e.g., Thurstone, 1951; Hoffman, Guilford, Hoepfner, & Doherty, 1968). While future research on aptitude will be quite unlike the studies reviewed in this report, it is important to understand the contributions as well as the limitations of this literature. We must begin again, but not from the beginning. The correlational studies provide a rough map of the terrain and a fertile ground for new hypotheses. But many of the problems that undermine the correlational literature are problems the new research on aptitude must also confront. Only by understanding the contributions and limitations of this older research can we avoid repeating the same mistakes or know if our new aptitudes are anything like the old ones.

This review is part of an ongoing research project aimed at understanding the nature and importance of individual differences in aptitude for learning. Requests for information regarding this project and for copies of this or other technical reports should be addressed to:

Professor Richard E. Snow, Principal Investigator
Aptitude Research Project
School of Education
Stanford University
Stanford, California 94305

TABLE OF CONTENTS

PREFACE ii

LIST OF TABLES v

LIST OF FIGURES vii

Introduction 1

REVIEW AND REANALYSIS OF CORRELATIONAL STUDIES OF SPATIAL ABILITY 4

The Hierarchical Perspective 4

Hierarchical Models 4
Hierarchical Factor Methods 5

Early Work 7

Thurstone's PMA Study 8

Thurstone's Analysis 8
The Zimmerman Reanalysis 10
A Reanalysis of the Spatial Tests 14
The Holzinger-Harman Reanalysis 17
Hypotheses about the SR-Vz Distinction 20
The Wrigley, Saunders, and Neuhaus Reanalysis 23
The Spearman Reanalysis 26
A Final Comparison 26

The Holzinger-Swineford Studies 29

Holzinger and Swineford (1939) 30
Swineford and Holzinger (1942) 33

The AAF Work 37

Reanalysis of the Perceptual Battery II Spatial Tests 38
The Sheppard Field Battery Analysis 44
Reanalyses of the Sheppard Field Battery Spatial Tests 48
Hypotheses about the SO Factor 53
Comparison of AAF and Thurstone Space Factors 56

Thurstone's Later Work 60

The Thurstone Perceptual Battery 60
The Thurstone Mechanical Aptitude Battery 67

Guilford's Postwar Work 84

Michael, Zimmerman, and Guilford (1950) 84
Michael, Zimmerman, and Guilford (1951) 89

Spatial Abilities in the SI Model 97

The Figural Cognition Battery 97
Evaluation of the SI Model. 116

The Upper Levels of the Hierarchy 120

Hierarchical Versus Gf-Gc Theory 120
The Horn and Cattell (1966) Study 121

Conclusions 126

Major Spatial Factors 127
Types of Spatial Transformations 128
Underlying Dimensions 128
Spearman Revisited 133
The Value of a Common Perspective 135

SOLUTION STRATEGIES ON SPATIAL TESTS 136

Early Work 136
Barratt 136
The Meyers Studies 138
The French Study 140
Yalow and Webb 142
Other Studies 143
Some New Data 144
Conclusions 149

SPEED AND LEVEL 151

Speed and Level Defined 152
Experimental Methods 152
Estimating Test Speededness 153
Part-Whole Correlation Studies. 154
Correlating Correctness with Time to Finish 155
Items Attempted as Speed 163
Time for Correct Responses as Speed 172
New Directions. 173

The Egan Study 173
Factors Affecting the Relationship between Speed and Level . . . 178
Latency and Correctness as Dependent Variables 180
Estimating the Correlation between Speed and Level 183
Implications for Speed Factors 185
Implications for Research on Aptitude 185

Summary and Conclusions 186

GENERAL CONCLUSIONS 188

Spatial Ability 188
Solution Strategies 190
Speed-Level 192
General Comments 193

REFERENCES 195

Footnotes 204

LIST OF TABLES

Table

1. Tests Loading on the Space Factor 9
2. The Spatial Relations Factor 11
3. The Visualization Factor 12
4. Two Factor Solution for the 14 PMA Space Tests 16
5. General and Space Factor Loadings of the 14 PMA Space Tests 18
6. Residual PMA Correlations After Removing General and Group Spatial Factors 21
7. Factor Analysis of Holzinger and Harman (1938) Residual Correlations 22
8. Summary of General Factor Loadings, Bipolar, SR-Vz Factor Loadings, and Speededness Estimates for SR and Vz Defining Tests . . 25
9. Variance in the 14 Space Tests Explained by Each Factor 27
10. Contributions of General, Space, and All Factors to the Total Variance 28
11. General and Space Factor Loadings for the Six Spatial Tests 32
12. General and Space Factor Loadings and Residual Correlations for the Six Spatial Tests 35
13. Bipolar Factor for Residual Space Test Correlations 36
14. Tests Loading .35 or More on Spatial Factors Identified in the AAF Studies 39
15. Unrotated and Varimax Rotated Factor Matrices for a Reanalysis of the Spatial Tests in Perceptual Battery II 42
16. Tests Defining Selected Factors in the Shepard Field Battery Analysis 46
17. Shepard Field Battery Reference Cluster Intercorrelations, and Correlations Between Selected Tests and Reference Clusters 50
18. Average Cluster Correlations for the Shepard Field Battery Reanalysis 52
19. Factor Analyses of Shepard Field Battery Reference Cluster Correlations 54
20. Speededness Estimates for Tests in the Shepard Field Battery Reference Clusters 57
21. Raw and Residual Correlations Between Shepard Field Battery Reference Clusters and AAF Spatial Relations Tests 59
22. Tests Loading on Selected Factors 61
23. Simplified Oblique Factor Pattern Matrix for the Mechanical Aptitude Battery 68

LIST OF TABLES (continued)

Table

24. First Order Factor Correlations 69

25. Orthogonal, Hierarchical Factor Structure Matrix for the Thurstone (1951) Mechanical Aptitude Battery 81

26. Correlations and Factor Loadings of the Six Spatial Tests 86

27. Factor Matrices for the Reanalysis of Michael, Guilford, & Zimmerman (1950) 88

28. Factor Matrices for Males for Reanalysis of Michael, Zimmerman & Guilford (1951) 94

29. Factor Matrices for Females for Reanalysis of Michael, Zimmerman & Guilford (1951) 95

30. Figural Slice of the Structure of the Intellect Model 98

31. Principal Component Factor Structure Matrix after Varimax Rotation for the Figural Cognition Battery 100

32. Varimax Rotated 22 Factor Maximum Likelihood Solution for the 44 Variable Matrix 107

33. Primary Factors, Abbreviations, and Tests 123

34. Means, Standard Deviations, and Intercorrelations of Marking Indices and Scores on the Three Spatial Tests 146

35. Correlations between Selected Reference Variables and Paper Folding Marking Indices (N=123), and between Paper Folding Total Correct and Selected Reference Variables for Less (N=62) and More (N=61) Marking Groups 147

36. Median Raw and Disattentuated Correlations Between Speed and Level 157

37. Zero-order Correlations, Beta Coefficients, and Multiple Correlations in the Prediction of Time Limit Scores (T) from Speed (S) and Level (L) Scores 160

38. Rotated Principal Components for the Davidson & Carroll (1945) Data 161

39. Average Speed, Level, and Reference Test Correlations 165

40. Average Spatial Test Correlations 167

41. Average Within and Between Session Correlations for Latency and Correctness for Three Experiment 1 Spatial Tests 176

42. Average Correlations for Latency and Correctness on Three Experiment 2 Spatial Tests 176

LIST OF FIGURES

Figure

1. Spatial tests in the SR-Vz factor space 13

2. Rotated factor loadings for the 14 spatial tests 15

3. Speededness versus bipolar factor loading for the 14 spatial tests 24

4. Rotated factor loadings for the reanalysis of Perceptual Battery II 43

5. Two dimensional scaling of spatial tests 66

6. Clustering of the disattentuated correlations by the diameter method 72

7. Multidimensional scaling of the mechanical aptitude battery 74

8. Diameter method hierarchical clusters superimposed on the scaling of the Mechanical Aptitude Battery 75

9. Clustering of the first and second order clusters by the diameter method 79

10. Two-dimensional scaling of the ten first order clusters 80

11. Clustering by (a) diameter method, and (b) rotated factor loadings superimposed on two dimensional scalings of the male data 91

12. Clustering by (a) diameter method, and (b) rotated factor loadings superimposed on two dimensional scalings of the female data 92

13. Two dimensional scaling of the 44 variable matrix 103

14. Diameter method hierarchical cluster analysis of 44 variables from the Figural Cognition Battery 104

15. Principal components superimposed on the scaling of Figure 13, excluding divergent production tests 105

16. Diameter method hierarchical cluster analysis of 60 variables from the Figural Cognition Battery 110

17. Two dimensional nonmetric scaling of disattenuated correlations from the 60 variable Figural Cognition Battery 111

18. Three dimensional nonmetric scaling of disattentuated correlations from the 60 variable Figural Cognition Battery 112

19. Median general factor loadings of the 23 SI abilities identified by Hoffman, et al (1968) 118

20. Two dimensional scalings of the Horn & Cattell (1966) data, (A) clustered by Horn & Cattell factor analyses, and (B) cluster in two or three different analyses 124

21. Proposed two dimensional scaling of spatial ability factors 130

22. Relating spatial tests and factors to process, speed, and complexity 132

LIST OF FIGURES (continued)

23. Average method clusters superimposed on two dimensional scaling of the Davidson & Carroll (1945) data 162

24. Plot of spatial tests for (A) the principal components reanalysis and (B) two spatial factors in Lord's (1956) solution 168

25. Two dimension scalings of (A) rights and (B) wrongs with average method clusters superimposed 171

26. Task complexity by ability. Meaningful latency data in solid boxes 182

27. Hypothetical plot of response latency versus item complexity for one subject 182

28. Hypothetical regressions of response latency on item complexity . . .184

Introduction

Research on aptitude for learning has entered a new era. Instructional studies have established that individual differences among learners often interact with instructional treatment variables (Cronbach and Snow, 1977; Snow, 1977). Much of this work has also underscored the need for deeper, more process-oriented understanding of the psychological nature of aptitudes. Cognitive psychologists have begun the experimental analysis of individual differences in information processing, and there is now reason to hope that coordination of these lines of work will lead to process theories of aptitude for learning from instruction (Snow, 1978).

One kind of aptitude of particular interest in both instructional and laboratory research has been spatial ability. That the difference between spatial and verbal aptitudes would interact with instructional treatments emphasizing one or the other form of representation has been a popular ATI hypothesis. But results have been conflicting and unsatisfactory largely, it seems, because our understanding of spatial tests is inadequate. Further, it is not clear just where and how spatial abilities fit into current structural models of ability organization or how they differ from verbal abilities in process terms (Snow, 1978). Recent experimental research, however, has begun to demonstrate that spatial processing appears to be fundamentally different from verbal-symbolic-sequential processing (Cooper and Shepard, 1976). Newer research that seeks a process understanding of individual differences in spatial ability would benefit from a clearer understanding of the end points of the psychometric tradition, specifically the number, nature, and apparent psychological differences between the various spatial tests and their factors. There is thus good reason to re-examine past research on individual differences in spatial ability with the new concepts and data techniques now available. This report, then, reviews and reanalyzes past findings to clarify the nature and measurement of spatial ability.

The report is divided into four sections. The first and longest part reinterprets the major American factor analytic studies on spatial ability in terms of a hierarchical model of ability organization. British factorists have, for the most part, interpreted their work from a hierarchical perspective, so no reinterpretation of that work is necessary (see Smith, 1964, for a comprehensive review). There are other reasons, however, for bypassing most of the British work. A major goal of this review is to examine the nature of the minor space factors, to determine how many there are and where they fit into the hierarchical model, and, if possible, to shed some light on the psychological processes which may underlie their differences. British work has paid scant attention to the subdivisions of the broad group space factor, and so is only marginally related to this concern.

On the other hand, American investigators, using multiple factor methods and following primary factor theories, have identified a number of different space factors. Thurstone (1951) claimed to have identified three, plus several others such as Closure speed (Cs), Flexibility of Closure (Cf), Perceptual Speed (Ps), and Kinasthetic (K) that correlated with the three space factors in varying degrees. Guilford and Lacey (1947) reported four orthogonal space factors: Visualization (Vz), Spatial Relations (SR or S1), Space 2 (S2), and Space 3 (S3). But there are substantial differences between these factors and those identified by Thurstone. French, Ekstrom, and Price (1963) listed three space factors: Visualization (Vz), Spatial Orientation (SO), and Spatial Scanning (Ss). The Vz factor was essentially the same as that identified by Guilford. The SO factor was a combination of Guilford's SR factor and Thurstone's S1, while Ss was the same factor Guilford, Fruchter, and Zimmerman (1952) called Planning Speed. Finally, Cattell (1971) placed Vz in the second stratum of the heirarchy under the label Gv (Horn and Cattell, 1966), and later, pv (Cattell, 1971). Gv was defined as a second order factor combining the first order primaries for Cf, Cs, S, DFT, and Vz. Further, the primaries that composed Gv were initially placed under Fluid Intelligence (Gf), with Cf and Vz loading strongly. Cattell recognized that complex spatial tests of the Vz and Cf sort measure Gf in part, but forced them under Gv nonetheless (see also Horn, 1976).

In short, there is much confusion in the American work on spatial ability. Are Cf and Vz really different abilities? How do the Thurstone factors map onto the Guilford factors? What elaborations are required by Guilford's (1967) later work with the Structure of the Intellect model, which posits thirty separate abilities within the figural content slice of that model? Finally, where do the replicable factors fit within a hierarchical model? Are Horn and Cattell correct when they assert that the various spatial primaries form a second order factor that is largely independent of Gf and Gc?

Such questions simply cannot be answered by a typical "litany of the saints" review of literature. The labels investigators have attached to their factors are often more misleading than helpful. Identical tests appear with different names in different studies, and tests with the same name are sometimes quite different. More difficult to detect are the subtle changes in test format and administration that alter the factorial composition of a test. Changing the dependent measure from solution time to number correct also changes the factor structure of a test. As will become evident, these "minor" changes in test format, administration procedures, and dependent variable can be as important as differences in the subject populations and range of tests entered into the analysis. Most important, however, are the ubiquitous differences in factor extraction and rotation criteria used by different investigators, and even by the same investigator over time.

The potentially most significant contribution of this review is the effort to reanalyze and reinterpret the major American factor analytic studies on spatial ability from a heirarchical perspective. While some may quibble with the utility of a hierarchical model it should be evident that reanalyzing a host of conflicting studies from some common theoretical perspective is the only way to reach meaningful integration.

It is impossible to review every factor analytic study that identified a space factor, as most well designed test batteries include at least a few spatial tests. Rather, this review concentrates on those studies that were designed to clarify the nature of spatial ability (e.g., Michael, Guilford, and Zimmerman, 1950), contained a particularly interesting combination of spatial tests (e.g., Thurstone, 1938), or supported important new models of ability organization (e.g., Horn and Cattell, 1966; Hoffman, Guilford, Hoepfner, and Doherty, 1968). Those seeking a broader review of the educational, practical, and personality correlates of spatial ability are referred to Smith (1964).

The second part of this review examines the effects of alternative solution strategies used by subjects on spatial tests. Some of the major confusions in the factor analytic studies are shown to result from individuals solving spatial problems in different ways. In addition to reviewing the literature on this topic, some new data are presented and discussed.

The third section reviews the relationship between speed, power, and complexity in test performance. The speed-power dimension is shown to be crucially important for all factor analytic work, particularly for the distinction between broad, general factors and narrow specifics. A method for examining the relationships between speed, power, and complexity is presented. It is argued that this method has important implications both for differential psychology and for cognitive psychology, and for attempts to coordinate the two.

Finally, the fourth section summarizes the conclusions and implications of the previous sections.

REVIEW AND REANALYSIS OF CORRELATIONAL STUDIES OF SPATIAL ABILITY

The Hierarchical Perspective

Hierarchical Models

British psychologists have long advocated hierarchical models of ability organization. Spearman's early two factor theory implied a crude hierarchy with "g" sitting atop a host of uncorrelated specific factors. When group factors were identified they were inserted between "g" and the specifics. Perhaps the best example of this sort of hierarchy can be found in the later work of Spearman's protégé Holzinger, using Holzinger's bi-factor method of factor analysis (e.g., Holzinger and Harman, 1938).

Hierarchical theories of ability organization have only recently gained credence in this country. Shortly after Thurstone introduced his centroid method in the Primary Mental Abilities study (Thurstone, 1938), multiple factor theory captured American theorists' attention. Its popularity has continued to the present; Guilford's facet model of abilities is the most recent attempt to keep all cognitive factors on equal footing (Guilford, 1967).

However, Thurstone himself initiated the first rapproachement between the two systems when he introduced the notion of oblique first-order factors. The matrix of these factor correlations could itself be factored to extract one or more second order factors. Continuing this process should eventually produce a factor akin to Spearman's "g."

Thurstone's idea was never really pursued because higher order factors were known to be unstable. Factorists were pressed to defend the psychological reality of first order factors; never mind factors of factors. Besides, multiple factor theory allowed aspiring students the hope of discovering new factors as important as those already in the catalog. Thus the number of "primary" factors climbed from Thurstone's seven to Guilford's 120.

The most compelling argument for a hierarchical factor theory is parsimony. Early defenders of the "separate but equal" theory had to remember only a handful of factors, and so hierarchical theory was not really simpler or more parsimonious. But French (1951) listed 59 factors in his monograph, and Guilford claimed to have identified 98 (Guilford and Hoepfner, 1971); parsimony is no longer irrelevant.

The more recent formulations of the hierarchical model place two or more broad group factors between "g" and the narrow group factors. One such model clusters verbal abilities and educational achievements together in a factor labeled v:ed, while spatial, practical, and mechanical abilities are clustered under a factor called k:m. This model was initially proposed by Burt (see Burt, 1949) and was later revised by Vernon (1950). A more elaborate version was suggested by Cronbach (1970). He split g into two broad group factors called Verbal Analytic and Figural Analytic. The v:ed factor was placed under the Verbal Analytic factor, while k:m was placed under the Figural Analytic factor. These factors were in turn subdivided and the process repeated until only test specific factors remained.

Another influential ability model was proposed by Cattell (1957, 1963) and later modified by Horn (Horn & Cattell, 1966; Horn & Bramble, 1967) and Cattell (1971). The earliest formulation distinguished Fluid Intelligence (Gf) and Crystallized Intelligence (Gc) as two correlated, second order factors derived from first order primaries enumerated by French (1951) and French, Ekstrom, and Price (1963).

Fluid ability was represented most strongly by tests highly correlated with Spearman's "g," such as Matrices, Classification, Cattell's "culture-fair" tests, and complex spatial tests such as Thurstone's Form Board. It was thought to represent the major measurable outcome of biological factors on intellectual development. Crystallized ability, on the other hand, was defined by the Verbal, Reasoning, and Number primaries. It was thought to represent the crystalization of fluid ability in specific achievement or skill areas, primarily through formal education and cultural experience.

More recent formulations of the model have relied heavily on a study by Horn and Cattell (1966) where three other second order factors were identified: General Visualization (Gv), General Speed (Gs), and General Fluency (Gr).

Neither the original Gf-Gc theory, nor its newer versions are truly hierarchical theories. Even though the second order factors are oblique, the theories deny that a third order factor is necessary. Cattell is particularly emphatic about this. On the other hand, Horn has referred to G as a combination of second order general factors, particularly Gf and Gc (Horn, 1976).

Hierarchical Factor Methods

While some American factorists now recognize the utility of Hierarchical models, many continue to analyze their data in traditional multiple factor ways. Even those who perform oblique rotations and extract higher order factors rarely transform the series of factor structure matrices into an orthogonal, hierarchical factor matrix. Appropriate procedures were developed some years ago by

Schmid and Leiman (1957) and Wherry (1959). In addition to reducing redundancy, a hierarchical transformation allows the investigator to examine the loadings of the tests, not just the loadings of the factors, on the higher order factors.

Several reanalyses are reported below in which oblique factors were extracted at several levels and the results transformed into an orthogonal, hierarchical factor structure matrix by the Wherry (1959) procedure. However, reanalyzing a large matrix in this way is time consuming and expensive, so the usual procedure was to refactor a submatrix of spatial test intercorrelations. The hierarchy was then constructed from the top down. The first unrotated centroid or principal factor extracted from such a matrix represents the group spatial factor plus all higher order factor loadings. The second unrotated factor is usually bipolar and represents the next bifurcation into minor group spatial factors. Thus, if an investigator claimed to have isolated three spatial factors, the matrix would consist of all tests with loadings on these three factors. Sometimes tests from factors with other labels (such as Perceptual Speed, Flexibility of Closure, Speed of Closure, etc.) were also included in the reanalysis because of their relevance to the spatial domain or to the particular hypothesis being investigated.

If more than two or three factors are present in the matrix, identifying the later factors becomes increasingly difficult (see Cattell, 1971, p.28). In such cases, it is important to examine both the unrotated and rotated matrices. If factors appear in the rotated matrix that were not apparent in the unrotated matrix, then the hierarchical structure must be constructed by the more laborious procedure of extracting primary and then higher order factors.

In either case, one could argue that this procedure of factoring only a selected submatrix does not allow the "true" factor structure to emerge. This would be a valid criticism if the aim were to reinterpret the entire matrix in the traditonal Thurstone or Guilford manner. However, within a hierarchical model one can profitably examine particular domains, such as the spatial factor and its subfactors in this way.

Another important issue is the stability of the heirarchical solution. A major argument against the hierarchical models of Spearman and Burt was the instability of the general factor. If the first centroid or principal axis represents "g," then the location of this axis should not be entirely at the mercy of the tests included in the battery. The "g" of one analysis could be the verbal factor of another, or more likely, some combination of the two.

Thurstone pointed out that the location of "g" could be ascertained with greater certainty by first determining the primary factor structure and then locating "g" by a simple structure rotation of factors extracted at the higher levels. Cattell (1971) argued that higher order factors could be located with even greater assurance by including a number of primaries in the analysis that are known to be uncorrelated with intelligence. Keeping the higher order factors orthogonal to this "hyperplane stuff" assures a better solution, since correlations between primary factors are less stable than correlations between tests.

However, most of the reanalyses reported here were concerned with the number and nature of the space subfactors, not the proper location of higher order factors. The usual question was: Are there really two or three factors in this matrix, or just one? In such cases, it is reasonable to use the first principal axis as an estimate of the broad group space factor plus all higher order factors, whatever they might be.

Early Work

British and American investigations of spatial ability have followed different paths since the time of Truman Kelley, and perhaps before. The dominant theme of the early British work was the attempt to isolate a group spatial factor independent of "g." However, after the need for a broad group spatial factor was recognized, British workers tended to regard spatial ability as an inferior counterpart to verbal ability, even though both appear at the second level of the hierarchical model (see Burt, 1949). The association of spatial ability with mechanical-practical abilities may have fostered the notion that spatial thinking was somehow more concrete, while verbal skills were more abstract (Smith, 1964). Early studies found spatial tests more useful than verbal tests for predicting success in technical schools, and so spatial tests have long been used for this purpose in both British and American educational systems.

One of the earliest British studies of spatial ability was reported by McFarlane (1925). Using a number of wooden construction tests, the Cube Construction Test, and Healy's Puzzle Box, she found some evidence of a group factor in addition to "g" for boys but not for girls. However, Spearman (1927) argued that her results could be explained by sex differences in experience with construction activities. He preferred to view her "performance tests" as unreliable measures of "g." The controversy continued through the early 30's, with some studies finding evidence for a small group spatial factor, and some finding "g" sufficient (Smith, 1964).

In 1935, El Koussy administered a battery of seventeen "spatial" tests and nine reference tests (verbal, perceptual speed, pitch and loudness discrimination) to 162 boys aged 11 to 13. He concluded that there was no evidence for a group factor in all his "spatial" tests, and that they primarily measured g. However, some spatial tests involved a group factor in addition to g; he called this the "k" factor. A closer look at his spatial tests reveals that all were figural, but not necessarily spatial. El Koussy (1935) also obtained introspective strategy reports from his subjects. He found that most subjects reported using visual imagery to solve tests that loaded highly on his k factor.

Meanwhile, in America, Truman Kelley (1928) tentatively identified two space factors in his studies of the abilities of school children. Some previous correlational work in the United States employed spatial tests, most notably the Minnesota Assembly Test and the Army Beta. However, space tests were ordinarily used as substitutes for verbal intelligence tests (such as the Army Alpha or Otis) when the testee was illiterate or not fluent in English.

Kelley identified one strong space factor ("ε") and a weak second factor ("θ"). He defined ε as the ability to perceive and retain geometric forms. Today the factor would probably be called a memory factor rather than a space factor. The second factor (θ) was defined as the ability to manipulate geometric forms. However, the factor was clearly defined in only one of his four samples.

Thurstone's PMA Study

Thurstone's Analysis

The next milestone in the American work was Thurstone's (1938) PMA study. Thurstone administered 56 tests to 218 volunteers who were either college students or college graduates. He extracted 13 centroid factors from the tetrachoric correlation matrix and then graphically rotated 12 to orthogonal simple structure. Thurstone could label only nine of these factors: Space (S), Perceptual Speed (P), Number (N), Verbal Relations (V), Word Fluency (W), Memory (M), Induction (I), Reasoning (R), and Deduction (D). The factor called Space was defined as "facility in spatial or visual imagery" (p. 80). The tests that loaded on this factor are listed in Table 1. Flags, Lozenges B, and Cubes had the highest correlations with the factor. The more difficult space tests (Form Board, Punched Holes, Copying, and Mechanical Movements) had only minor loadings. Their major loadings were on the uninterpreted Factor XII. Thurstone could not label this factor because the Chicago Vocabulary and Reading II tests also loaded highly on it.

Insert Table 1 about here

The PMA study is of particular interest because it contains a broad, representative battery of tests, and has become a classic in the field. It paved the way for all future factorial work on spatial ability. It has been reanalyzed

Table 1

Tests Loading on the Space Factor
(After Thurstone, 1938)

	Test	Space Factor	Other Factors
20	Flags	.64	(.43XI)
22	Lozenges B	.63	(.32M, .31D)
18	Cubes	.63	(.29R)
27	Pursuit	.58	(.33N)
23	Surface Development	.55	(.29X)
53	Hands	.46	(.47XI, .29N)
19	Lozenges A	.45	(.53XII, .33R)
45	Syllogisms	.43	(.32I, .32V, .29D)
21	Form Board	.42	(.50X, .40XII)
17	Block Counting	.41	(.36R, .35XI)
55	Sound Grouping	.41	(.45V, .38W)
6	Verbal Classification	.41	(.54P, .31I, .30V)
8	Figure Classification	.39	(.40I, .40D)
24	Punched Holes	.34	(.53 XII, .34D)
54	Rhythm	.34	(.60P, .29D)
14	Disarranged Sentences	.30	(.46P, .40V, .32M)
28	Copying	(.27)	(.37I, .36P, .34X)
29	Areas	(.22)	(.48I)
25	Mechanical Movements	(.07)	(.41R, .40X)

by Spearman (1939), Eysenck (1939), Holzinger and Harman (1938), Zimmerman (1953), and Wrigley, Saunders, and Neuhaus(1958).[1] Each used a different factor method and achieved an interpretation flavored both by the factor method and the psychological theories of the investigator.

<u>The Zimmerman Reanalysis</u>

Zimmerman (1953) started where Thurstone (1938) left off, and continued to rotate Thurstone's centroid axes toward simple structure. With the hindsight of the AAF work (Guilford and Lacey, 1947) and Thurstone's later studies (Thurstone, 1944, 1951), Zimmerman was able to identify two space factors rather than the one reported by Thurstone in 1938.

The first factor was the same as Thurstone's Space factor and Zimmerman called it Spatial Relations (SR). Tests and their loadings on the factor in the two solutions are shown in Table 2. The second space factor was a revised version of Thurstone's uninterpreted Factor XII. Tests and their loadings on the factor in the two solutions are shown in Table 3. Zimmerman labeled the factor Visualization (Vz) after a similar factor that was repeatedly obtained in the AAF work (Guilford and Lacey, 1947).

Insert Tables 2 and 3 about here

Tests that defined the Vz factor were more difficult than those that defined the Spatial Relations factor. Further, tests that loaded on the Spatial Relations factor were speeded, while those that loaded on the Visualization factor were relatively unspeeded.

The 15 tests that loaded on one of these two factors are plotted in the SR-Vz factor space in Figure 1. The tests do not cluster near the two factors but are arrayed throughout the factor space. Further, the plot suggests that the factors would be better represented by the oblique vectors SR' and Vz' rather than orthogonal vectors SR and Vz. The correlation between SR' and Vz' is about .64.

Insert Figure 1 about here

The major shortcoming of both the Zimmerman and Thurstone solutions is the large number of subsidiary loadings for each test. The problem is most acute for the Visualization factor. Thurstone could not separate the complex spatial tests from vocabulary and reasoning tests. Zimmerman managed to do so, but inspection of Table 3 reveals that he was not totally successful.

Table 2

The Spatial Relations Factor

(After Thurstone, 1938, and Zimmerman, 1953)

	Test	Thurstone (1938) Space Factor	Zimmerman (1953) Spatial-Relations Factor
20	Flags	.64	.73
22	Lozenges B	.63	.60
18	Cubes	.63	.59
53	Hands	.46	.55
17	Block Counting	.41	.52
27	Pursuit	.58	.51
23	Surface Development	.55	.50
19	Lozenges A	.45	.40
45	Syllogisms	.43	.40
21	Form Board	.42	.32
8	Figure Classification	.39	.22
6	Verbal Classification	.41	.21
55	Sound Grouping	.41	.21
24	Punched Holes[a]	.34	.27
54	Rhythm[a]	.34	.08
26	Identical Forms[a]	.32	.13
28	Copying[a]	.27	.17
29	Areas[a]	.22	.21
25	Mechanical Movements[a]	.07	.13

[a]Included for reference only.

Table 3

The Visualization Factor
(After Zimmerman, 1953, and Thurstone, 1938)

	Test	Zimmerman (1953) Vz	Zimmerman (1953) Other Factors	Thurstone (1938) Factors
24	Punched Holes	.62	.29D, .29C, .27S	.53XII, .34D, .34S
21	Form Board	.62	.40R, .32VF, .32S	.50X, .42S, .40XII
19	Lozenges A	.54	.40S, .31V, .30C	.53XII, .45S, .34R
25	Mechanical Movements	.40	.41R, .34GR, .33D	.41R, .40X, .29XI
20	Flags	.37	.73S	.64S, .43XI, .26N
23	Surface Development	.35	.50S, .25VF	.55S, .29X
49	Word Recognition	.34	.34M, .28LF	.38M, .36XII, .34W
58	Vocabulary (Chicago)	.31	.76V, .46VF	.54XII, .46R, -.39XI
17	Block Counting	.30	.52S, .39R	.41S, .36R, .37X, .35XI
18	Cubes	.30	.59S, .34P, .26R	.63S, .29R, .25XI
28	Copying	.26	.44R, .37GR	.37I, .36P, .34X
29	Areas	.26	.52GR, .29M, .26C	.48I, .29R, .28XI

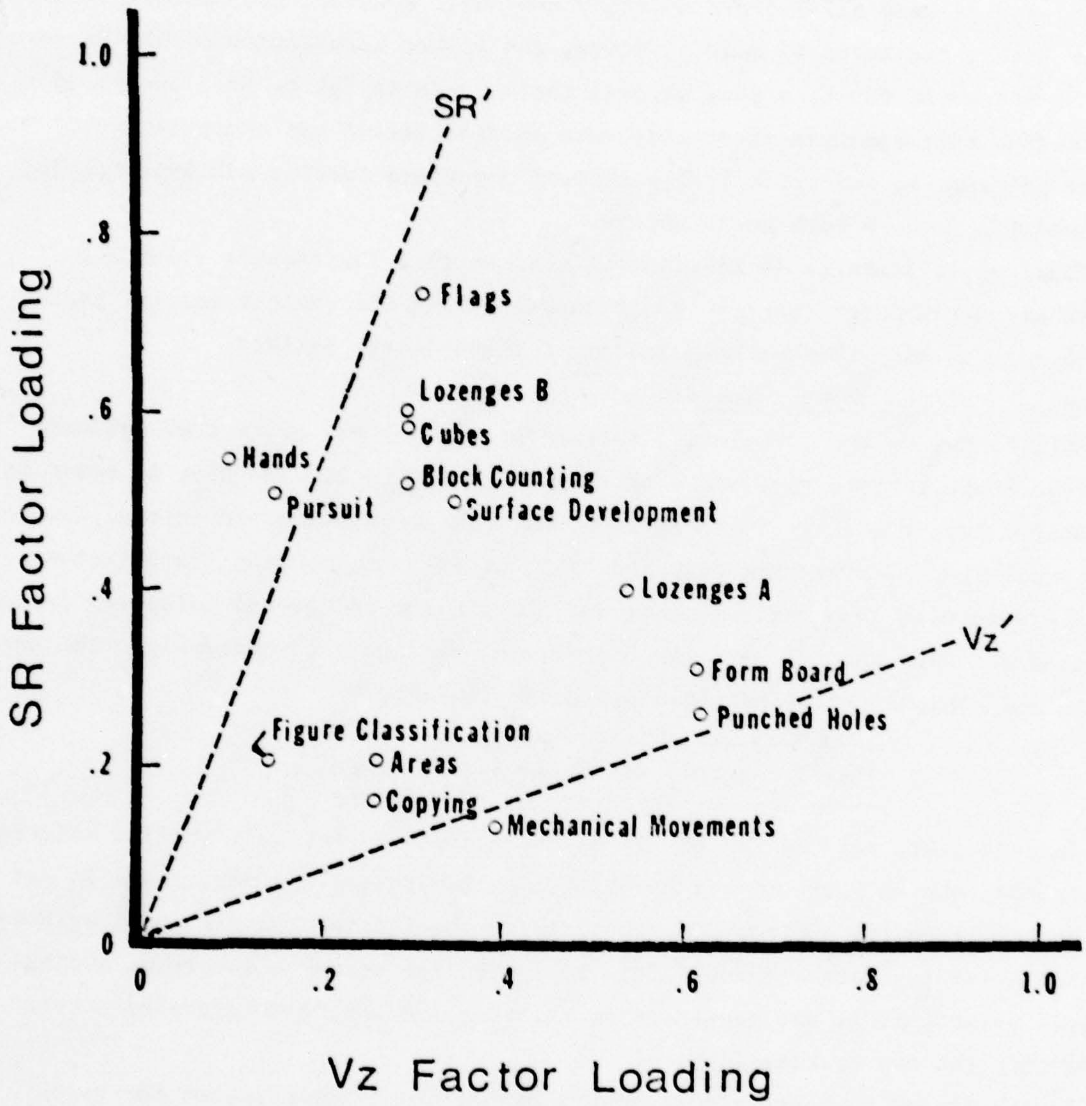


Figure 1. Spatial tests in the SR-Vz factor space (After Zimmerman, 1953).

The root of the problem is that the Vz tests have high correlations with each other and with other complex tests. On the other hand, the SR tests have much lower correlations with each other and with other tests in the battery. This correlation pattern is the major stumbling block for multiple factor theories that attempt to keep all factors on equal footing. However, the pattern is consistent with a hierarchical model. There, the higher correlation of the Vz tests would be accounted for by a more general factor such as "g" or Gf. The residual correlations that remained after this more general factor was extracted could then be examined to determine if the pattern supported further subdivisions into more specific factors such as SR and Vz.

Thus, while Zimmerman's solution is cleaner than Thurstone's solution, neither adequately represents the relationships among the various spatial tests nor the relationships between them and other tests in the battery.

A Reanalysis of the Spatial Tests

Refactoring of the correlation matrix for the 14, PMA tests that defined these two space factors yields the plot shown in Figure 2. The plot is based on a principal factor solution with squared multiple correlations as initial communality estimates. Convergence required eight iterations, and the final factor matrix was rotated to a varimax criterion. The first and second unrotated factors accounted for 48.8 and 5.6 percent of the total variance, respectively. The unrotated and rotated factor matrices are shown in Table 4.

Insert Figure 2 and Table 4 about here

In this plot, rotated Factor II' is the same as Zimmerman's Spatial Relations factor, and rotated Factor I' is his Visualization factor. Again, tests do not cluster neatly on the two factors, but fall at regular intervals on an imaginary arc that spans the factor space. This analysis also makes an important methodological point: It is not necessary to refactor the entire correlation matrix to identify the two spatial factors.

If the SR and Vz factors represent independent abilities, then the tests require these abilities in varying degrees. However, the plot is consistent with other interpretations. These are more obvious in the unrotated factor loadings. Here the tests are roughly arranged in order of complexity, those with positive projections on Factor II are the simplest, while those with negative projections on Factor II are more complex. The continuum may also represent speed to power, with positive projections on Factor II representing speed and negative projections representing power.

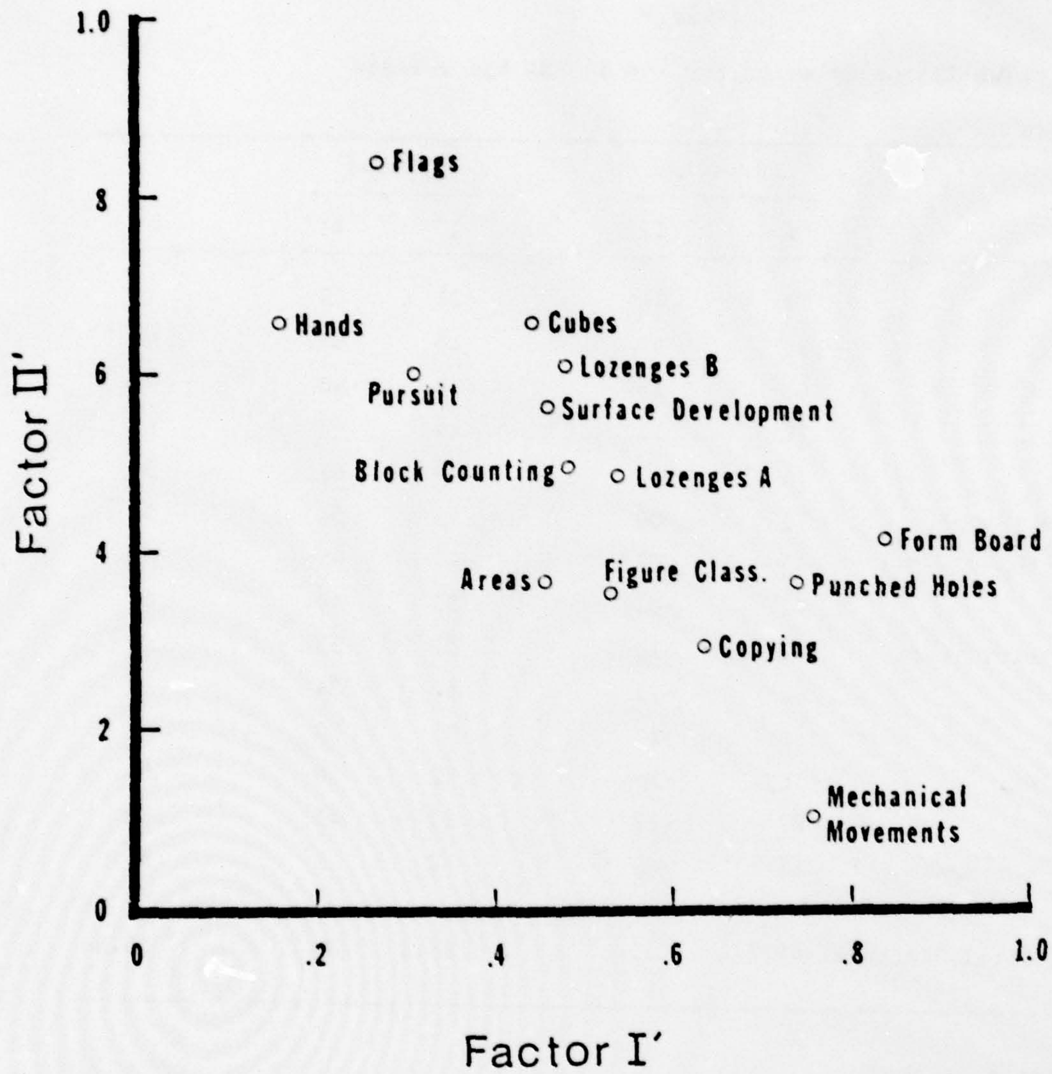


Figure 2. Rotated factor loadings for the 14 spatial tests (After Thurstone, 1938).

Table 4

Two Factor Solution for the 14 PMA Space Tests

Test		Unrotated I	Unrotated II	Rotated I'	Rotated II'	h^2
20	Flags	77	41	27	83	77
53	Hands	56	37	15	65	44
27	Pursuit	63	23	30	60	45
18	Cubes	77	17	44	65	62
22	Lozenges B	76	12	47	61	59
23	Surface Development	71	09	45	56	51
17	Block Counting	67	02	47	48	45
19	Lozenges A	72	-02	53	48	52
8	Figure Classification	62	-10	51	35	39
29	Areas	60	-11	50	34	37
28	Copying	66	-22	63	30	49
24	Punched Holes	76	-24	72	36	64
21	Form Board	87	-27	82	41	83
25	Mechanical Movements	61	-42	73	12	55
	Percent of Total Variance	48.8	5.6	28.1	26.3	54.4

Note. Decimals omitted.

The Holzinger-Harman Reanalysis

Earlier reanalyses of the PMA study, by Holzinger and Harman (1938) and Eysenck (1939), paint a different, yet parsimonious picture of the data. Holzinger and Harman used bi-factor factor analysis while Eysenck used a variant of multiple group factor analysis. Both analyses produced a general factor and a number of independent group factors. Both defined the general factor as the overlap between the group factors. The methods differed primarily in how variables were assigned to groups. Holzinger and Harman used B-coefficients, while Eysenck plotted each column of the correlation matrix, and then assigned variables to groups on the basis of similar contours. The spatial factor was almost identical in the two solutions. Thus, examination of either analysis should yield the same conclusions. The Holzinger-Harman analysis is examined here because it is reported in greater detail, and because the bi-factor method reappears in later analyses.

The tests that loaded on Holzinger and Harman's space factor are shown in Table 5, ranked according to their loading on the general factor. The ranks are about the same as on Thurstone's (1938) first centroid axis. In fact, the correlation between the two sets of "g" values is .968. This compares favorably with the correlation of .965 between the *Thurstone first* centroid and the Holzinger-Harman general factor loadings for all 56 tests that Woodrow (1939a) computed. However, the high correlation does not imply that the two "g" values are interchangeable, as Woodrow (1939a) concluded. The correlation coefficient looks only at relative differences within each group, not at constant differences between them. For this subset of spatial tests, the average first centroid loading is .57, while the average bi-factor "g" loading is only .49. This is a considerable difference, especially when all ten factors in the bi-factor analysis account for only 53.4 percent of the total variance.

Insert Table 5 about here

Loadings on the group spatial factor and the independent specifics are also shown in Table 5. Flags again defined the factor, this time even more forcefully than in Thurstone's solution. There are a few discrepancies between the two solutions, but the overall picture is roughly the same. In fact, the Space factor loadings of the 14 tests correlated .85 with their loadings in the Thurstone Space factor. The most notable difference between the two factors is the presence of Syllogisms, Sound Grouping, Verbal Classification, Rhythm, and Disarranged Sentences on the Thurstone factor and their absence on the Holzinger-

Table 5

General and Space Factor Loadings of the 14 PMA Space Tests
(After Holzinger & Harman, 1938)

	Test	General	Space	Specific
21	Form Board	67	55	50
28	Copying	58	36	73
29	Areas	58	27	77
24	Punched Holes	57	50	65
19	Lozenges A	54	47	70
22	Lozenges B	53	54	65
23	Surface Development	52	48	71
25	Mechanical Movements	52	31	80
18	Cubes	51	58	64
8	Figure Classification	45	42	79
17	Block Counting	40	56	73
27	Pursuit	38	52	76
20	Flags	36	72	59
53	Hands	32	45	83

Note. Decimals omitted.

Harman factor. This could mean that the bi-factor Space factor is "cleaner," but probably just reflects the subjective bias inherent in the B-coefficient method of clustering. Here, tests are selected for clustering on the basis of psychological hypotheses as well as correlation patterns.

Thurstone's Space factor accounted for 7.4 percent of the variance in the battery, while that of Holzinger and Harman accounted for only 6.0 percent of the variance. The Thurstone factor is larger because every test loads on it, even though most loadings are small and psychologically uninterpretable. This is more obvious when the average spatial factor loading of the 14 bi-factor space tests is compared with the average spatial factor loading of the top 14 tests on Thurstone's space factor. In both cases, the average loading was .48. The 14 bi-factor space tests had a slightly lower average loading on the Thurstone space factor of .43. Thus, the bi-factor solution did not remove meaningful variance from the group factors to construct the general factor.

A Reanalysis of the Holzinger-Harman Residuals

The question remains, however, whether the Space factor should be divided into two factors, as the previous reanalysis suggested (see Table 4 and Figure 2). However, there it was not possible to determine how much of the variance on the large first factor belonged higher up in the ability hierarchy, since only spatial tests were included in the analysis. The bi-factor analysis has accomplished precisely this. Now the question is whether the residual correlations that remain after "g" and the group spatial factor have been removed will support another bifurcation. Such an analysis would also reveal how much additional variance might be accounted for by another factorial split.

The residual correlations among the 14 space tests were computed by Holzinger and Harman and are reproduced in Table 6. Principal factors were extracted from this matrix with squared multiple correlations as initial communality estimates. Convergence required six iterations when one factor was retained. The factor matrix is shown in Table 7. The factor is obviously bipolar, with Hands at one pole and Mechanical Movements at the other. When the factor was reflected and the tests ranked according to their loadings, the ranks were almost identical (rho = .99) to those on the second unrotated factor of Table 4. It will be recalled that this matrix was obtained by extracting two factors from the raw correlations of the 14 tests. Thus, this bipolar factor represents the same Visualization-Spatial Relations dimension. But now the distinction appears less important, since this factor accounts for only 4.7 percent of the variance in the 14 tests. Previously, when general and group spatial factors were not extracted from the matrix, the orthogonal Visualization and Spatial Relations factors accounted for

28.1 and 26.3 percent of the variance respectively. Thus, the situation appears different when viewed from a hierarchical perspective. For the 14 spatial tests, the general factor accounted for 25.4 percent of the variance, the group Spatial factor accounted for 24.4 percent of the variance, and the Spatial Relations-Visualization bipolar factor added another 4.7 percent of the variance. In all, 49.8 percent of the variance in these spatial tests was accounted for without the bipolar factor, and 54.5 percent of the variance with it.

Insert Tables 6 and 7 about here

Hypotheses About the SR-Vz Distinction

The problem of deciphering the nature of this Visualization-Spatial Relations factor remains unsolved. It was mentioned that the factor might reflect the complexity of the processing demands placed on the subject. If this were true, then the more complex Visualization tests should have higher general factor loadings than the simpler Spatial Relations tests. In fact, the rank order correlations between the bi-factor "g" loading and the Spatial Relations-Visualization bipolar factor loading was -.72, which supports the hypothesis. The greater complexity of Vz tests may reflect the influence of several factors. For example, subjects must generate their responses on tests with high Vz loadings, while on those with higher SR loadings subjects can simply select their answers from among the alternatives provided. This may explain why the multiple choice Surface Development test is found more toward the SR end of the factor. In fact, Surface Development variance is often split between the SR and Vz factors (Bechtoldt, 1947; Guilford and Lacey, 1947). On the other hand, the subjects must actually draw their answers on Punched Holes and Form Board.

Another possibility is that the Vz-SR distinction reflects speed vs. power. A crude measure of test speededness is the number of items the examinee must complete in a unit of time. A better measure would be the average number actually completed per unit time. However, Thurstone corrected some tests for guessing, so the resulting means are not comparable across corrected and uncorrected tests.

A rough estimate of speededness may be obtained by dividing the total number of items in the test by the number of minutes alotted for the test. This value was computed for each of the 14 tests and then plotted against the bipolar factor loading for that test from Table 7. The plot is shown in Figure 3. The correlation between the two variables is .75, which supports the hypothesis that the Vz-SR factor reflects a speed-power dimension.

Table 6

Residual PMA Correlations After Removing General and Group Spatial Factors
(After Holzinger & Harman, 1938)

Test	8	17	18	19	20	21	22	23	24	25	27	28	29	53
8 Figure Classification														
17 Block Counting	-11													
18 Cubes	12	09												
19 Lozenges A	-04	03	-08											
20 Flags	-04	-06	08	09										
21 Form Board	-02	03	-05	05	-05									
22 Lozenges B	07	-11	-02	10	03	-08								
23 Surface Development	-01	-06	01	03	09	-03	03							
24 Punched Holes	04	-01	-16	08	-04	08	05	01						
25 Mechanical Movements	04	07	00	00	-08	09	-03	-06	15					
27 Pursuit	01	02	12	-07	06	-05	01	-02	-08	-12				
28 Copying	12	02	-04	-05	-13	11	-08	03	-09	07	01			
29 Areas	-06	02	-04	-09	-07	01	-04	-02	05	04	07	05		
53 Hands	-12	10	00	-04	15	-09	12	01	-08	-16	05	-02	11	

Note. Decimals omitted.

Table 7

Factor Analysis of Holzinger and Harman (1938)
Residual Correlations

	Test	Unrotated Factor Loading
53	Hands	.37
20	Flags	.30
27	Pursuit	.22
18	Cubes	.16
22	Lozenges B	.13
23	Surface Development	.09
17	Block Counting	-.01
29	Areas	-.01
19	Lozenges A	-.03
8	Figure Classification	-.10
28	Copying	-.16
24	Punched Holes	-.25
21	Form Board	-.26
25	Mechanical Movements	-.37

Note. Decimals omitted.

The situation may be summarized by examining the general factor loadings and speededness values for the three tests at each end of the Vz-SR bipolar factor. These are the only tests that had even marginally significant loadings on the factor. The results are shown in Table 8. The average g loading of the three tests that defined the Vz end of the bipolar factor was much higher, and the average speededness estimate much lower, than the corresponding averages for the three SR tests.

Insert Figure 3 and Table 8 about here

The Wrigley, Saunders, and Neuhaus Reanalysis

Another rotation of the 13 Thurstone centroid axes was reported by Wrigley, Saunders, and Neuhaus (1958). The analysis was conducted to compare orthogonal quartimax rotation with the other published factor solutions for this matrix.

Quartimax rotation attempts to maximize the sum of the fourth powers of the rotated factor loadings. The result is to concentrate the variance for each test on as few loadings as possible. Thus, quartimax, maximizes the variance of the rows, while the more popular (and more recent) varimax procedure maximizes the variance of the columns in the factor matrix.

For the PMA data, this method produced a factor matrix somewhere between the Thurstone or Zimmerman solutions and the Holzinger-Harman or Eysenck solutions. The analysis yielded a General-Verbal factor that accounted for 27.2 percent of the total variance, a Spatial factor that accounted for 13.5 percent of the variance, and a Numerical factor that accounted for 6.8 percent. The remaining ten factors each accounted for 1.9 to 3.1 percent of the total variance. The General Verbal and Spatial factors are akin to Gc and Gf (Cattell, 1971; Horn and Cattell, 1966; Horn, 1976) although some of the "g" in Gf was placed on the general verbal factor.

The analysis did concentrate the variance for the 14 spatial tests on the spatial factor. This is shown in Table 9 (below). Fully 39.4 percent of the variance in these tests was accounted for by the Space factor. This is almost twice the variance accounted for by the Space factors in the Thurstone or Zimmerman solutions. Another 10.8 percent was on the General-Verbal factor, and the remaining 20.9 percent was scattered throughout the other 11 factors. Thus, orthogonal rotation does not necessarily produce a solution in which most tests are factorially complex. This appears to be the chief virtue of the method. On the other hand, the solution does not separate g from Gf or Gc, nor does it imply any direct

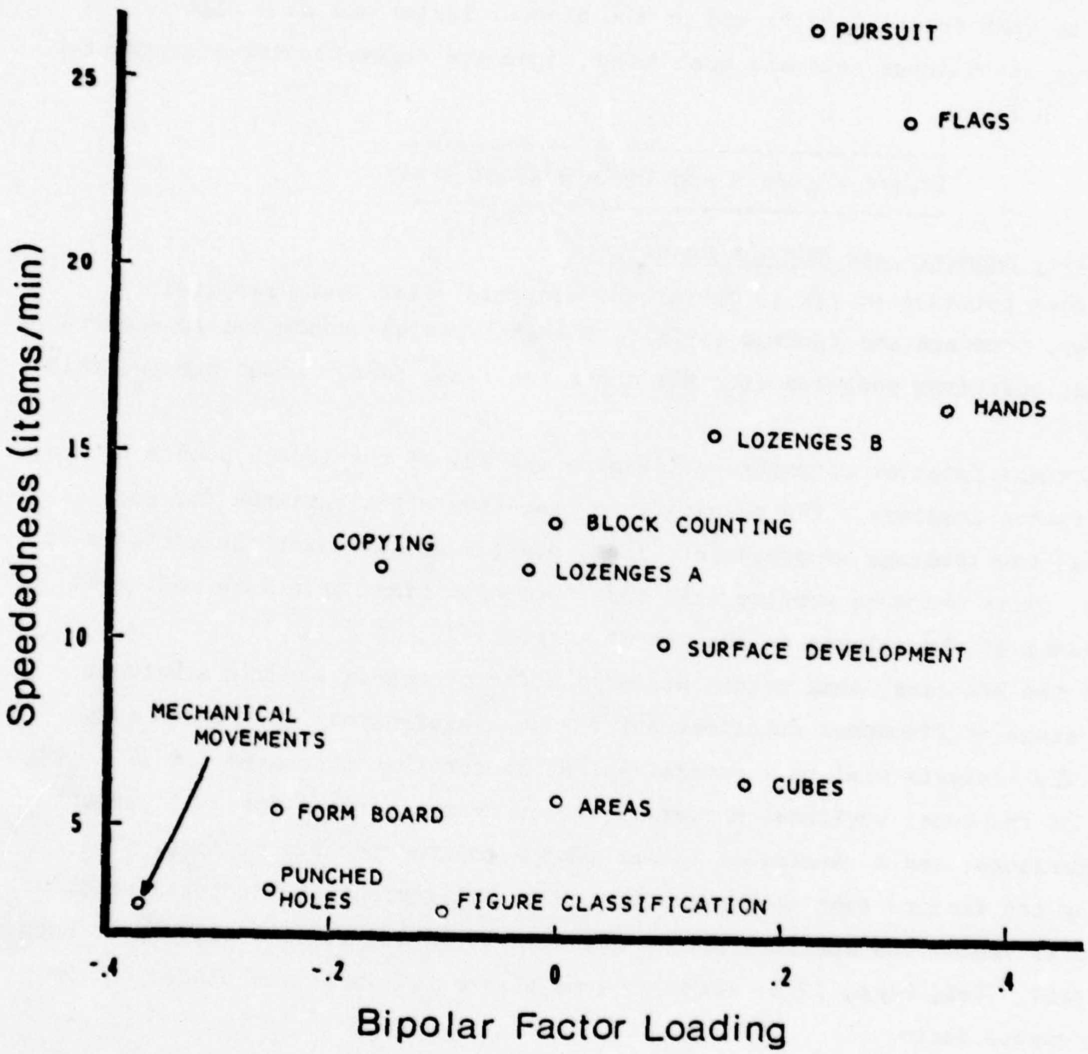


Figure 3. Speededness versus bipolar factor loading for the 14 spatial tests (After Thurstone, 1938).

Table 8

Summary of General Factor Loadings, Bipolar SR-Vz Factor Loadings, and Speededness Estimates for SR and Vz Defining Tests

	General Factor Loading	SR-Vz Bipolar Factor Loading	Speededness Estimate[a]
S-R Defining Tests			
(53) Hands	.32	.37	16.3
(20) Flags	.36	.30	24.0
(27) Pursuit	.38	.22	26.6
Mean	.35	.30	22.3
Vz Defining Tests			
(24) Punched Holes	.57	-.25	3.3
(21) Form Board	.67	-.26	5.2
(25) Mechanical Movements	.52	-.37	2.9
Mean	.59	-.29	3.8

[a]Number of items in a test divided by the number of minutes allotted for the test.

relationship between the ten small factors and the three larger ones.

The Spearman Reanalysis

Before moving on to other studies, brief mention should be made of Spearman's reanalysis of the Thurstone PMA data (Spearman, 1939). After criticizing Thurstone for rotating the g factor out of existence, Spearman proceeded to average the correlations between and within each of ten test groups. These groups were used by Thurstone during the construction and selection of tests for inclusion in the battery. Spatial tests were unevenly split between the Form and Space groups. The Form group was composed of Pursuit, Copying, Areas, and Identical Forms. The Space group contained the other nine space tests, excluding Figure Classification and Hands.

Spearman first computed the general factor and extracted it from the correlation matrix. This procedure is similar to defining the first principal axis as "g" and produces a larger "g" than the Holzinger bi-factor method. He then extracted four group factors: Verbal, Spatial, Numerical, and Memory. Together, the Space and Form groups defined the Space factor, with the former loading .46 and the latter .26. The corresponding general factor loadings were .57 for the Space group and .52 for the Form group. However, further comparisons with the other PMA analyses are impossible since composite variables rather than individual test scores were used in the analysis.

A Final Comparison

Tables 9 and 10 summarize the six analyses of the PMA data. The Spearman reanalysis has been excluded from Table 9 as it is impossible to determine exactly what his analysis does to the 14 space tests that have been used as a common referent in all the analyses.

Insert Tables 9 and 10 about here

Table 9 provides an interesting comparison of how the several analyses decomposed the variance in the 14 space tests. For example, Thurstone's 13 factors accounted for 71.3 percent of the variance in the 14 space tests. However, the Space factor accounted for only 21.3 percent. The remaining 50 percent was scattered throughout the other 12 factors. Zimmerman's re-rotation of the matrix concentrated some of this variance on the Visualization factor. His 13 factors accounted for 70.6 percent of the variance in the 14 tests. The Spatial Relations factor accounted for 20 percent and the Visualization factor for another 13.7 percent, for a total of 33.7 percent of the variance. However, 36.9 percent remained in the numerous small loadings on the other 11 factors.

Table 9

Variance in the 14 Space Tests Explained by Each Factor

Thurstone (1938)		Zimmerman (1953)		Wrigley et al.(1958)		Holzinger-Harman (1938)		Eysenck (1939)	
Factor	Percent of Variance	Factor	Percent of Variance	Factor	Percent of Variance	Factor	Percent of Variance	Factor	Percent of Variance
S	21.3	S	20.0	GV	10.8	G	25.4	G	24.8
P	2.5	P	2.5	S	39.4	S	24.4	S	26.0
N	3.5	N	.9	N	1.1				
V	0.2	V	1.4	P(A)	1.3				
M	3.1	M	3.1	P(B)	1.2				
W	0.9	LF	1.1	M(A)	1.9				
I	4.9	C	6.3	M(B)	2.5				
R	6.6	GR	5.8	F	1.1				
D	5.5	D	2.8	I	4.6				
X	6.9	VF	2.9	AR	3.5				
XI	8.7	R	8.7	SR	1.6				
XII	6.4	V_z	13.7	XII	0.7				
XIII	1.1	MR	1.4	XIII	1.4				
Total 71.3			70.6		71.0		49.8		50.8

Table 10

Contributions of General, Space, and All Factors to the Total Variance

Analysis	Variance Explained			Number of Factors
	General Factor	Space Factor	All Factors	
Spearman (1939)	33.2	2.8	43.2	5
Eysenck (1939)	30.8	6.6	53.4	9
Holzinger & Harman (1938)	30.1	6.0	53.4	10
Zimmerman (1953)		12.1[a]	72.3	13
Thurstone (1938)		7.4	70.7	12
Wrigley et al. (1958)	(27.2)	13.3	72.3	13

[a]Sum of variance explained by SR and Vz factors.

The Wrigley et al. reanalysis concentrated variance on the Space factor, but left a sizable chunk (10.8 percent) on the General-Verbal factor. The remaining 20.8 percent remained in small pieces on the other 11 factors. Note that a total of 51.2 percent of the variance was accounted for by the General-Verbal and Spatial factors. This is almost identical to the total variance accounted for by the General and Spatial factors in both the Holzinger-Harman and Eysenck reanalyses.

On the other hand, the Holzinger-Harman bi-factor solution accounted for only 49.8 percent of the variance in the fourteen tests. However, it is all found on just two factors: the general factor (25.4%) and the group spatial factor (24.4%). Further bifurcation of the space factor into two minor group factors added another 4.7 percent of the variance for a total of 54.5 percent. But the partitiioning of variance is still represented in an orderly way, rather than scattered piecemeal throughout the system. Eysenck's analysis was almost identical, except that slightly more of the total variance was accounted for by his General and Spatial factors.

Thus, the multiple factor analyses of Thurstone, Zimmerman,and Wrigley et al. are a bit misleading. While they tend to account for more of the total variance than the group factor methods, much of the variance in each test lies in the small uninterpreted loadings on factors other than the one on which each test has its primary loading.

Table 10 provides another perspective for comparison of the various analyses. Here, contributions of factors to the total battery of 56 tests are presented. Table 10 reveals that the bi-factor solution allows for both a sizable General factor and a group Spatial factor about as large as Thurstone's. The bi-factor Space factor accounted for 6.0 percent of the total variance, while Thurstone's accounted for 7.4. If Thurstone is followed and only those tests with loadings of .39 or greater are interpreted then his Space factor accounted for only 5.8 percent "interpretable" variance. In either case, the Thurstone and Zimmerman solutions ignore the large general factor but fail to produce a larger group spatial factor. This argues for a hierarchical representation of human abilities.

The Holzinger-Swineford Studies

Another series of bi-factor analyses conducted by Holzinger and Swineford (1939) and Swineford and Holzinger (1942) provide additional support for the hierarchical model. These studies also reveal the importance of subject population and test difficulty for the factor structure obtained.

Holzinger and Swineford (1939)

The first study (Holzinger and Swineford, 1939) was concerned with the stability of the bi-factor solution in two subject samples. The subjects were all seventh and eighth grade students at two junior high schools in Chicago. Students in the Pasteur sample (N = 156) were primarily from working class homes. Both parents were foreign born for roughly half of these students. Students in the Grant-White group (N = 145), on the other hand, were predominantly children of American born parents living in an affluent suburb.

A battery of 24 tests was administered to both samples. The spatial tests in the battery were:

1. Visual Perception Test: This test consisted of 60 items selected from Spearman's Visual Perception Test, Part III. A series of five adjacent figures was presented. The student's task was to indicate which one of four alternatives came next in the series.

2. Cubes: This test was a simplification of one of Brigham's (1932) tests since the latter was found too difficult for children at the elementary school level. The test was similar to Thurstone's (1938) Cubes, which was also an adaptation of a Brigham test. However, it was probably still too difficult, since the average number correct for both groups was about 24 out of a possible 40; random guessing would yield an average of 20 correct.

3. Paper Form Board: This test was a 28 item multiple choice test in which the student indicated which of four alternatives (a square, triangle, hexagon or trapezoid) could be constructed from the stimulus pieces. The test appears easier than the Thurstone or French kit Paper Form Board Tests.

4. Lozenges A: This is the more difficult of Thurstone's (1938) two Lozenges tests. The average number correct was only 18 out of 36, which is exactly at the level of random guessing.

The Grant-White group was also administered a slightly revised version of the Paper Form Board test and Thurstone's Flags test. The Form Board test was revised by adding items in the middle difficulty range, and deleting a corresponding number at the extremes. However, the correlation between the two Form Board tests was only .40.

Three bi-factor analyses were performed. The first two were based on the 24 tests administered to both groups. In the third analysis, the Grant-White data were refactored including the revised Form Board and Flags tests and excluding the original Form Board and Lozenges A tests.

Five factors were extracted in all three solutions: General, Spatial, Verbal, (Perceptual) Speed, and Memory. The General and Spatial factor loadings of the spatial tests in the three analyses are shown in Table 11. There were some differences between the Pasteur and Grant-White factor patterns. The General factor accounted for more variance in the Grant-White sample than in the Pasteur sample, both in the four spatial tests and in the entire battery of 24 tests. Similarly, the group factors were larger in the Pasteur sample. This might in part reflect the fact that the Pasteur group was, on average, six months older.

Insert Table 11 about here

Paper Form Board defined the Spatial factor in the Pasteur analysis, while Lozenges A defined it in the comparable analysis for the Grant-White group. These differences result from the extremely low intercorrelations of the spatial tests in both samples; the average intercorrelation was .33 in the Pasteur group and .35 in the Grant-White group. For Grant-White analysis B, the average intercorrelation was .34.

The residual correlations were quite small, especially in the Pasteur group, with a maximum of -.036 and an average absolute value of only .026. For the Grant-White group, residual correlations were reported only for analysis B. Here, the residuals were slightly higher; the average of the absolute values was .032. The largest residuals were .056 between Flags and Visual Perception, and .053 between the modified Paper Form Board and Cubes. However, neither of these residual correlations are consistent with previous attempts to subdivide the spatial factor in the Thurstone (1938) study. For example, in the bi-factor reanalysis of the PMA study, Form Board and Cubes had a negative residual correlation of -.05, and thus appeared on different poles of the bipolar factor extracted from the residuals.

There are several reasons for these inconsistencies:

1. There were only a small number of spatial tests included in the battery.

2. At least two of the tests (Lozenges A and Cubes) were too difficult for the students. Further, the two versions of the form board test correlated only .40, indicating that this test was extremely unreliable when administered to students of this age. Such unreliability would also help explain the relatively low average intercorrelations of the spatial tests.

3. This type of form board test appears to be easier than the Thurstone or French kit versions. This, coupled with the fact that the Cubes and Lozenges

Table 11

General and Space Factor Loadings for the Six Spatial Tests
(After Holzinger & Swineford, 1939)

	Test	Pasteur (N=156)		Grant-White(N=145)			
		General	Space	General	Space	General	Space
1	Visual Perception	.59	.40	.59	.33	.59	.45
2	Cubes	.25	.43	.36	.31	.36	.28
3	Paper Form Board	.22	.50	.45	.22		
4	Lozenges A	.45	.44	.52	.55		
25	Modified Form Board					.40	.44
26	Flags					.46	.31

tests were comparatively more difficult for these students, may have reduced the complexity (or difficulty) disparity between the tests. Thus, the Vz - SR distinction could not surface.

Swineford and Holzinger (1942)

A subsequent study by Swineford and Holzinger (1942) with 457 ninth graders provides some support for these hypotheses. This time, six spatial tests were included in a battery of 28 tests. Five had been used in the previous study: Visual Perception, Cubes, the revised Paper Form Board, Lozenges A, and Flags. The sixth test, Designs, was another Thurstone test. In this test, the subject must indicate which complex designs contain a simple design resembling a capital Greek letter sigma. The test is similar to the Gottschaldt Figures, except that the target design is the same in all items, and the recognition task appears easier. The test has been used in several analyses, and is often factorially complex, splitting its common variance between Spatial and Perceptual Speed factors (as in Guilford and Lacey, 1947; and between similar factors in Bechtoldt, 1947). However, the test defined the Flexibility of Closure (Cf) factor in Thurstone's study of mechanical aptitude (Thurstone, 1951). There, its highest correlation was with the Gottschaldt Figures (r=.49), which also had its highest loading on the Cf factor. It correlated about as highly with the Gottschaldt Figures test as any other test (r=.28) in the two AAF studies in which it was used (Guildford and Lacey, 1947). Further, the two tests had their major loadings on the same factors in these studies (Guilf rd and Lacey, 1947, Perceptual Battery I & II);as well as in Bechtoldt (1947). Thus, there is good reason to suspect that the Designs test reflects an analytic type of spatial ability (Gf) that plays an important role in later hierarchical theories.

Using a modified bi-factor method of factor extraction, Swineford and Holzinger obtained a general factor, five group factors, and two doublets. The group factors were Spatial, Verbal, (Perceptual) Speed, Memory, and Number. The modified bi-factor technique permitted a test to load on more than one group factor. However, loadings of the six spatial tests were confined to the General and group Spatial factors. These are shown in Table 12, along with the residual correlations.

One factor was extracted from this matrix of residual correlations by the centroid method. Maximum off-diagonal correlations were used as initial communality estimates, and the solution was iterated three times. The average difference between the communalities on the second and third iterations was .005.

The final centroid factor loadings for the six tests are reported in Table 13. Flags defined one end of the factor, while Designs had the highest loading on the other pole. The pattern has some notable consistencies with the bipolar factor previously extracted from the residual correlations in the Holzinger-Harman PMA reanalysis (see Table 7). The major difference is that Designs defined one pole in this analysis, while Mechanical Movements, Form Board, and Punched Holes defined the corresponding pole in Table 7. Further, Paper Form Board had only a slight loading on this bipolar factor (-.07), whereas in the reanalysis of the PMA residuals Form Board had a much stronger negative loading (-.26) (see Table 7). This probably indicates that the Paper Form Board test used in this study is much easier than the PMA Form Board test. If this is true, then the bipolar factor of Table 13 reflects the same complexity dimension that was earlier hypothesized to account for the bipolar factor extracted from the Holzinger-Harman PMA residuals. Of course, this argument assumes that the residual covariance in the Designs test reflects the same type of complex spatial processing as that involved in Thurstone's (1938) Mechanical Movements, Form Board, and Punched Holes tests; or the Gottschaldt Figures test.

Insert Tables 12 and 13 about here

Another possibility is that the residual covariation in the Designs test reflects some other ability dimension, such as Perceptual Speed or Visual Memory. The Perceptual Speed hypothesis is unlikely, as there were no Perceptual Speed tests included in the spatial group factor. Further, one would expect such a test to cluster with the more speeded tests, such as Flags, rather than oppose them on the opposite end of the bipolar factor. The Visual Memory Hypothesis is not a serious threat either, as several investigators (most notably Smith, 1964) have argued that the essence of spatial thinking is the ability to retain and reproduce images of geometric forms in their proper proportions. In this view, tests like Thurstone's Form Board and Punched Holes are good measures of spatial ability because they require the subject to retain and reproduce spatial images. Tests like Flags and Hands are poor spatial measures because there is no necessity to retain and reproduce the spatial image in its correct proportions. Rather, the template of the answer is provided, and the subject merely must verify it.

But this sort of argument ignores that subjects must do more than retain an image in most complex spatial tests. Rather, they usually must transform

Table 12

General and Space Factor Loadings and Residual Correlations for the Six Spatial Tests
(After Swineford & Holzinger, 1942)

Test		Factor Loading		Residual Correlation					
		General	Space	1	2	3	4	5	6
1	Visual Perception	.55	.35						
2	Cubes	.41	.43	.008					
3	Modified Form Board	.29	.46	.038	-.015				
4	Lozenges A	.50	.59	-.021	-.016	-.008			
5	Flags	.45	.43	.059	.077	-.031	-.028		
6	Designs	.49	.23	.011	-.051	.065	.012	-.045	
	Percent of Variance	20.8	18.4						

Table 13

Bipolar Factor for Residual Space Test Correlations
(After Swineford & Holzinger, 1942)

	Test	Factor Loading
5	Flags	.33
2	Cubes	.23
1	Visual Perception	.04
3	Modified Form Board	-.07
4	Lozenges A	-.09
6	Designs	-.22

or manipulate it in some way. Thus, individual differences in the speed or power of such manipulations will also influence test performance. While the arguments of Smith (1964) call attention to an important aspect of spatial ability, a process understanding of spatial ability will require a much more detailed analysis.

The AAF Work

In 1947, Guilford and Lacey reported the results of the AAF factor analytic studies. These studies identified two strong spatial factors called Spatial Relations (SR) and Visualization (Vz), and two tentative space factors, S2 and S3. Thurstone's Flags, Figures, Cards, and Cubes were among the tests that loaded on the Spatial Relations factor. Thus, the factor was thought to be the same one Thurstone called Space in his PMA study (Thurstone, 1938). Guilford and Lacey observed that the Spatial Relations factor "seems to involve relating different stimuli to different responses, either stimuli or responses being arranged in spatial order. It is not clear whether the appreciation of spatial arrangement of stimuli or of responses separately is the key to the factor" (p. 838).

The Visualization factor was defined by the tests like Space Visualization I, which is a paper folding task similar to Thurstone's Punched Holes. Guilford and Lacey felt the factor was "strongest in tests that present a stimulus either pictorially or verbally, and in which some manipulation or transformation to another visual arrangement is involved" (p. 838).

The third space factor (S2) was a specific factor confined to Thurstone's Hands and Flags tests. Guilford and Lacey felt that an appreciation of right hand-left hand discrimination might be an important aspect of the factor. They did not attempt to define the factor further, apparently regarding it as of minor importance. S3 was defined by the test Two Hand Coordination and appeared in only one analysis. In later discussions Guilford dropped factor S3 and listed only Visualization, Spatial Relations, and Right-Left Discrimination (S2) as the three space factors identified in the AAF work (Michael, Guilford, Fruchter and Zimmerman, 1957; Hoffman, Guilford, Hoepfner, and Doherty, 1968).

It is impossible to review every AAF study in detail since the report is large and contains many tests that do not appear in other factor analytic studies. Table 14 lists the space factors and defining tests for the 16 factor analyses reported in the monograph. Tests were included in Table 14 if they loaded .35 or higher on one of the spatial factors.

Insert Table 14 about here

The Spatial Relations factor appeared in one form or another in all 16 analysis. Complex Coordination usually defined or loaded highly on the factor. Instrument Comprehension II and Discrimination Reaction Time also entered prominently in several studies. A composite test of Thurstone's Flags, Cards, and Figures, and his Cubes test also loaded on the factor in two analyses (Perceptual Battery I and II). The Cubes test was also included in the Integration battery, but loaded only .31 on the SR factor in that analysis. Instrument Comprehension II defined the SR factor in the Integration analysis, suggesting that this SR factor was not the same one identified by Thurstone (1938).

This is particularly evident in the analysis of Perceptual Battery II, the only one to include three Thurstone space tests, namely the Flags-Cards-Figures composite, Cubes, and Hands. Hands broke away from the SR factor and defined another factor with the Flags-Cards-Figures test. The latter test split its variance between the two factors. The SR factor was defined by Complex Coordination, although Cubes and the Flags-Cards-Figures composite had the next highest loadings. This was the only analysis where the factor S2 (defined by Hands) appeared.

<u>Reanalysis of the Perceptual Battery II Spatial Tests</u>

The correlation matrix for the nine tests loading .30 or higher on any of the three spatial factors in Perceptual Battery II was refactored using principal factoring and squared multiple correlations as initial communality estimates. Convergence required nine iterations when two factors were extracted. The unrotated and varimax rotated factor matrices are shown in Table 15. The first unrotated factor accounted for 87.3 percent of the common variance, and the second an additional 12.7 percent. The first factor represents the general plus group space factor, while the second bipolar is the familiar SR-Vz dimension. A plot of the varimax rotated factor loadings is shown in Figure 4. Hands, Flags, and Cubes identify the SR dimension as they did in the PMA reanalysis (see Table 4 and Figure 2). Mechanical Principles is the familiar point on the Vz factor. The pattern is remarkably similar to one presented earlier for the reanalysis of the Thurstone PMA Space tests (see Figure 2). Note that the Gottschaldt Figures test clustered with Mechanical Principles near the Vz factor in the plot.

Insert Table 15 and Figure 4 about here

Table 14

Tests Loading .35 or More on Spatial Factors Identified in the AAF Studies
(After Guilford & Lacey, 1947)

Study	Spatial Relations Factor		Visualization Factor		Other Factors
1. December 1942 Classification Battery pp 798-821 N = 3254-4774	Complex Coordination	49	Mechanical Principles A	52	
	Dial & Table Reading	40			
	Two Hand Coordination	39			
	Discrimination Reaction Time	36			
2. July 1943 Classification Battery pp 798-821 N = 3000 unclassified airmen	Bombadier Stanine	58	None		
	Discrimination Reaction Time-D	48			
	Complex Coordination	47			
	Two Hand Coordination	45			
	Dial & Table Reading	43			
3. November 1943 Classification Battery pp 798-821 N = 1900 unclassified airmen	Instrument Comprehension II	50	None		
	Complex Coordination	46			
	Two Hand Coordination	44			
	Instrument Comprehension I	40			
	Discrimination Reaction Time	39			
4. September 1944 Classification Battery pp 798-821 N = 8158 unclassified airmen	Discrimination Reaction Time	47	Mechanical Principles B	51	
	Complex Coordination	47			
	Instrument Comprehension	41			
	Dial & Table Reading	40			
5. Reasoning Tests pp 113-122 N = 202 classified pilots	Instrument Comprehension II	57	Spatial Visualization I-X1	56	
	Complex Coordination	52	Mechanical Principles A	54	
	Planning Air Maneuvers	41	Pattern Comprehension	50	
	Instrument Comprehension I	39	Spatial Visualization II	42	

Table 14 (continued)

Study	Spatial Relations Factor		Visualization Factor		Other Factors
6. Judgement Tests	a		a		
pp 142-155	None		Mechanical Movements	47	
a) N = 689 high school boys			Mechanical Comprehension	46	
	b		b		
b) N = 1024 classified pilots	None		Mechanical Comprehension	41	
			Mechanical Movements	37	
			Mechanical Judgement	35	
7. Foresight & Planning I	Complex Coordination	56	Mechanical Principles A	50	
pp 180-190			Driving Skill	42	
N = .202 unclassified airmen					
8. Foresight & Planning II	Plan a Course	62	Mechanical Principles A	45	
pp 180-190	Complex Coordination	48			
N = 170 classified pilots					
9. Integration Tests	Instrument Comprehension II-B	50	Spatial Visualization I-X2	53	
pp 215-225	Complex Coordination	42	Mechanical Principles	47	
N = 266 classified pilots	Instrument Comprehension I-A	48	Figure Analogies	36	
	Cubes	(31)			
10. Memory Tests I	Complex Coordination	46	None		
pp 261-268	Memory for Planes	43			
N = 179 bombadiers & 527 unclassified airmen	Mechanical Principles	36			
11. Memory Tests II	Complex Coordination	52	None		
pp 261-268	Mechanical Principles	(33)			
N = 238 unclassified airmen + 527 unclassified airmen from 10 (above)					

Table 14 (continued)

Study	Spatial Relations Factor		Visualization Factor		Other Factors	
12. Mechanical Tests	Complex Coordination	52	Mechanical Movements A	51		
pp 333-339	Mechanical Comprehension	36	Mechanical Principles A	49		
N = 153 unclassified			Mechanical Comprehension D	40		
aviation students						
13. Perceptual Battery I	Complex Coordination	52	Mechanical Principles A	50		
pp 408-418	Flags, Figures, Cards	49	Map Distance	47		
N = 392 unclassified	Direction Orientation B	47	Gottschaldt Figures	(32)		
aviation students	Table Reading	42				
	Cubes	42				
	Dial Reading	41				
14. Perceptual Battery II	Complex Coordination	50	Mechanical Principles A	54	Space II	
pp 408-418	Cubes	47	Map Distance	46	Hands	46
N = 392 unclassified	Flags, Figures, Cards	43	Gottschaldt Figures	38	Flags, Figures, Cards	42
airmen	Direction Orientation B	41			Cubes	(25)
	Picture Integration	40				
	Table Reading	35				
15. Carefulness Battery	Two Hand Coordination	62	Directional Plotting (wrong)	56	Space III	
pp 686-695	Complex Coordination	46	Mechanical Principles B	54	Two Hand Coordination	58
N = 354 aviation	Discrimination Reaction Time	(33)	Directional Plotting (right)	45	Rotary Pursuit	53
students					Plotting (right)	46
					Directional Plotting(R)	42
					Plotting (accuracy)	38

Note. Page numbers refer to Guilford & Lacey (1947)

Table 15

Unrotated and Varimax Rotated Factor Matrices for a Reanalysis of the Spatial Tests in Perceptual Battery II (After Guilford & Lacey, 1947)

Test	Unrotated		Rotated		
	I	II	V_z	SR	h^2
Hands	.34	-.33	.06	.48	.23
Flags, Cards, Figures	.63	-.30	.30	.63	.49
Cubes	.69	-.16	.43	.56	.50
Complex Coordination	.43	00	.33	.27	.18
Direction Orientation B	.63	.06	.53	.35	.40
Picture Integration	.68	.16	.63	.30	.49
Mechanical Principles	.49	.21	.51	.15	.28
Map Distance	.35	.16	.37	.10	.15
Gottschaldt Figures	.46	.23	.50	.11	.27

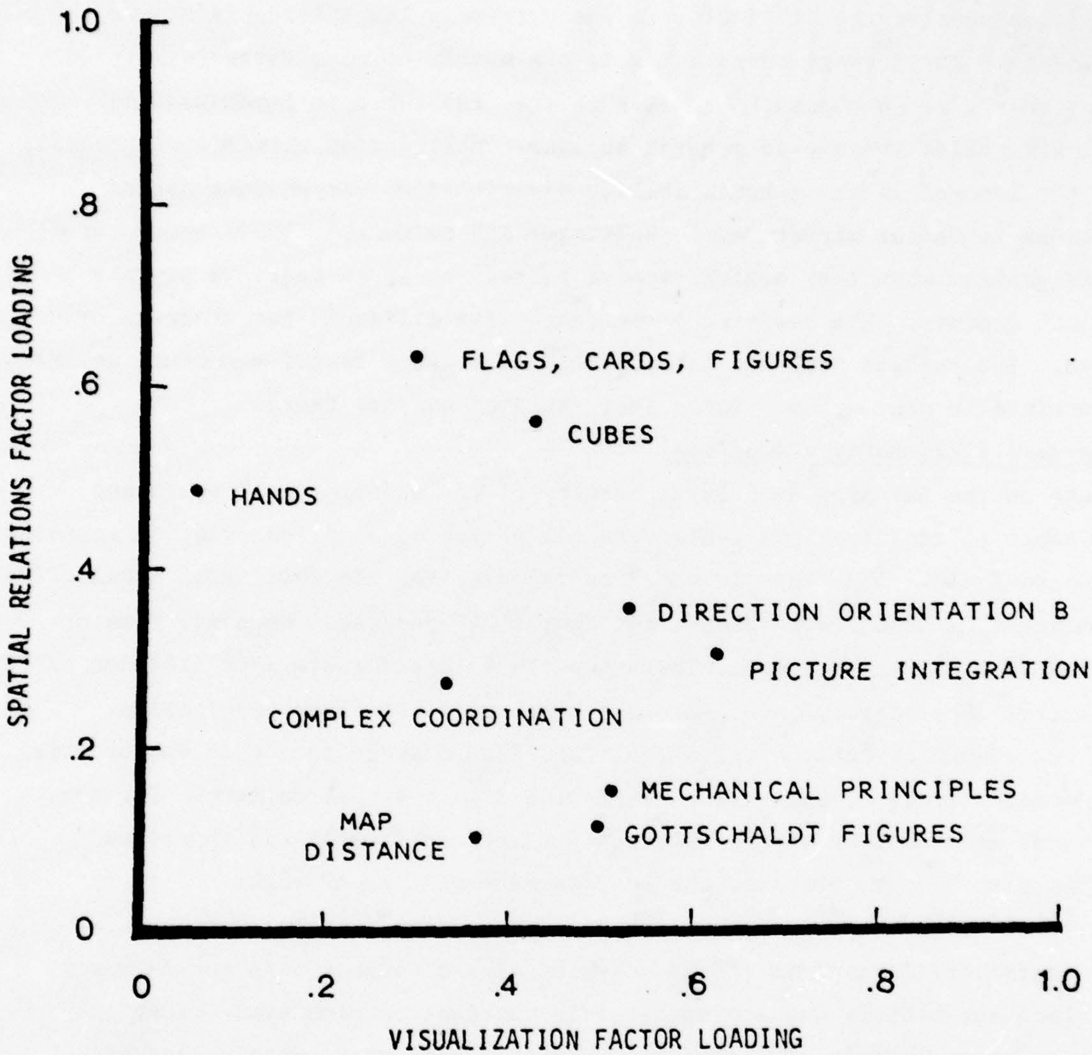


Figure 4. Rotated factor loadings for the reanalysis of Perceptual Battery II (After Guilford & Lacey, 1947).

If three factors are retained, Map Distance, not Hands splits away and defines a singleton. This third factor is probably not spatial, since no other test loads on it.

A large part of the difficulty is the extremely low intercorrelations of the tests. The average correlation in the matrix of nine tests refactored here was only .27. One possible reason is that the subjects (unclassified airmen) were below average in general ability. Restriction of range, especially toward the low end of the general ability distribution, can produce marked distortions in factor structures.[2] Holzinger and Swineford (1939) encountered the same problem when they administered a battery of space tests to seventh and eighth graders. The tests were apparently too difficult for students of this age. The results were low intercorrelations and a factor structure unlike that obtained in many other studies that employed similar tests.

The Sheppard Field Battery Analysis

Late in the AAF program a large battery of 45 "experimental" tests and 20 reference (classification) tests were administered to 8,158 aviation students at Sheppard Field. The "experimental" tests were, for the most part, final or revised versions of tests developed during the AAF program. However, some new tests and adaptations of several Thurstone (1938) space tests were included in the battery. The battery is of interest because it allows an examination of the relationships between tests that loaded on a space factor in one or more of the smaller studies, but were never included in the same battery. Further, it provides another look at the relationship between the AAF and Thurstone space factors. It is, perhaps, the best summary of the AAF work.

Not all of the experimental tests were administered to every recruit. The 45 tests were divided into five sub-batteries of approximately nine tests each. Each sub-battery was administered in combination with every other sub-battery to approximately 400 students. Within sub-battery correlations and correlations between experimental and classification tests were based on about 1,600 students. Correlations between classification tests were based on the full sample of 8,158 recruits. Thus, there is confounding of between sub-battery and between group covariance. This is not a major concern, however, since the examinees were all from the same population (18-19 year old Air Force recruits) and sample sizes were all large.

The correlations between the 65 tests were computed and reported in an appendix to the AAF final report (Guilford and Lacey, 1947), but were not factored at that time. Five years later, Guilford, Fruchter, and Zimmerman

(1952) reported a factor analysis of 39 experimental and seven reference tests. They factored the matrix by a combination of the multiple group and centroid methods. Two independent rotations to orthogonal simple structure were then performed.

Five of the 13 factors that were extracted and rotated are of particular interest here. Tests that had an average loading in the two solutions of .35 or greater on these factors are shown in Table 16. Factor names that seem more appropriate are shown in parentheses if they differ from the Guilford et al. (1952) label.

Insert Table 16 about here

Visualization. The factor called Visualization (Vz) was similar to Zimmerman's (1953) Visualization factor. The defining tests were Spatial Visualization I and II, and Mechanical Principles. In Spatial Visualization II, the examinees read a verbal description of how a solid block of wood painted a different color on each side is cut into smaller blocks. They are then asked questions about the resulting number of blocks of a given size and color. Spatial Visualization I is a Binet type paper folding task similar to Thurstone's Punched Holes. However, the task is probably more complex as the shape of the piece that is cut out changes as the paper is unfolded. The Mechanical Principles test is similar to Thurstone's Mechanical Movements, except that some items use aviation situations.

Spatial Orientation. The factor called Spatial Relations is better described by the label Spatial Orientation (SO). The Spatial Relations label is used in this review to describe the factor defined by Thurstone's Cards, Flags, and Figures tests. The central characteristic of tests that define the SO factor appears to be "empathetic involvement" (Guilford et al., 1952). The observers must first imagine themselves in the situation and then make some judgement about the stimulus array from this perspective. There is often a left-right discrimination component in these tests. The Visualization factor, on the other hand, seems to involve the mental manipulation of an external object without any imagined movement or involvement of the self.

Aerial Orientation, the test that defined this factor, was the source of the Spatial Orientation test in the Guilford-Zimmerman (1948) Aptitude Survey. Each item shows a cockpit view of a shoreline. Five photographs of an airplane in different altitudes are presented adjacent to each stimulus picture. The

Table 16

Tests Defining Selected Factors
in the Shepard Field Battery Analysis
(After Guilford, Fruchter & Zimmerman, 1952)

Factor	Test No.	Test Name	Average Loading
Visualization	3[a]	Spatial Visualization II	.62
	19[a]	Spatial Visualization I	.60
	52[a]	Mechanical Principles	.58
	41	Angle Estimation	.48
	15	Formation Visualization	.42
	21[a]	Object Recognition	.39
	2	Figure Analogies	.39
	11	Visualization of Maneuvers C	.35
Spatial Relations (Spatial Orientation)	24[a]	Aerial Orientation	.62
	11[a]	Visualization of Maneuvers C	.58
	15[a]	Formation Visualization	.44
	51	Instrument Comprehension	.43
	63	Complex Coordination	.35
Object Identification (Spatial Relations	25[a]	Object Identification I	.63
	26[a]	Object Identification II	.56
Planning Speed (Spatial Scanning)	14[a]	Maze Tracing	.54
	12[a]	Planning a Circuit	.48
	20	Map Planning	.35
Perceptual Speed	8[a]	Speed of Identification C	.58
	53[a]	Speed of Identification A	.52
	32	Pattern Assembly	.44
	6	Map Distance	.36

[a]Tests used to define factors in the reanalysis.

examinee must match the cockpit view of the shoreline with the airplane position from which that view would be seen.

Visualization of Maneuvers presents a stimulus picture of an airplane in a starting position. A simple maneuver is described such as a turn or a bank of a certain number of degrees. The examinee must select the alternative picture that portrays the airplane's new position. An important requirement is that all maneuvers be visualized from the pilot's position in the cockpit. Thus a right turn means to the pilot's right regardless of the plane's position in the stimulus picture.

In Formation Visualization, each item presents top and side silhouette views of a formation of two or three airplanes. The examinee must select the picture that shows the formation from a front view. This particular test appears amenable to both Spatial Orientation (empathetic) or Visualization (detached manipulation) strategies. Its loading on the Spatial Orientation factor was about the same as its loading on the Visualization factor. However, there is other evidence that many subjects solve items like those in Aerial Orientation (which was the defining test for this factor) in a non-empathetic way (see Barratt, 1953, and pp. 136 below).

<u>Spatial Relations</u>. The Object Identification doublet is actually closer to the Spatial Relations (SR) factor of the Thurstone (1938) PMA analyses than the factor labeled SR by Guilford et al. (1952).

Object Identification I is similar to Thurstone's Flags, except that the stimulus figures are silhouettes of planes, trucks, guns, tanks, and ships instead of flags. Object Identification II is the second part of this test. Here the stimulus figures are flags as in the original Thurstone test. The factor reflects the high correlation (.68) between these parallel tests.

<u>Spatial Scanning</u>. The next factor in Table 16 was called Planning Speed, since most of the tests that loaded on the factor had loaded on various planning factors in the earlier AAF work. French, Ekstrom, and Price (1963) call the same factor Spatial Scanning (Ss). Scanning seems to be a more appropriate label as "the level of planning required by the tests seems to be a simple willingness to find a correct path visually before wasting time marking the paper." (French et al., 1963, pp. 42-43).

The factor was defined by Maze Tracing. This test presents a complicated maze on which certain pathways are marked by a letter. The examinee must indicate whether the pathway between any two letters is clear or blocked. Planning a Circuit presents an electrical circuit diagram with intersecting, intermeshed wires, a meter at one end, and several sets of two pole terminals at the other end. The task is to determine where a battery should be placed in order to complete the circuit through the meter.

The major question about this factor is whether it, too, represents a new spatial subfactor. If it does, it is difficult to see what the psychological basis of this uniqueness may be. French et al. (1963) suggest that, within the spatial domain, it may represent an ability analogous to that required in rapidly scanning a printed page for comprehension. If so, one would expect some connection between this factor and Perceptual Speed.

Perceptual Speed. The final factor in Table 16 is Perceptual Speed. The factor was defined by Speed of Identification C and A. In Form A, each item presents five stimulus figures and five alternatives. The examinee must indicate the four matching pairs of objects. Form C is similar except that the items are composed of airplane silhouettes. The distinguishing details are not as obvious and in most paired views the alternatives are rotated. However, it does not appear necessary to rotate the alternative in order to match it with one of the stimulus figures. In any case, Form C is more complex. It correlated slightly higher with the Vz and SO tests, and slightly lower with the Ps factor than did form A.

Pattern Assembly also loaded on the Ps factor. This test is an easy, speeded form board test, and thus adds another dimension to the hypothesis that the SR-Vz factor reflects speed-power differences in the tests. It appears that if a Visualization test is made extremely easy it becomes a measure of Perceptual Speed. Thus, Thurstone's difficult Form Board helped define the Vz factor (see Tables 4 and 7). Swineford and Holzinger (1942) used an easier form board test and it fell in the middle of the SR-Vz continuum (see Table 13). Whether a form board test more difficult than the Pattern Assembly test used here and less difficult than the Swineford and Holzinger (1942) version would load on the SR factor is uncertain. Zimmerman (1954) suggests that it would. He concluded that a spatial test may be made to measure Ps, SR, Vz, and Reasoning (in that order) by increasing the complexity or difficulty of the items. However, examination of his data and the relevant research on the speed-power problem suggests that this may not be the case (see p. 151 ff below).

Reanalyses of the Sheppard Field Battery Spatial Tests.

Correlational Analyses. Tests with their highest loadings on each of the five factors of interest were selected. These are noted in Table 16. The centroid through the two or three tests defining each factor was then used to represent that factor. This produces a more extreme separation between factors than would result if all tests that loaded on the factor were grouped together. Thus, the Vz factor was defined by forming a composite of Spatial Visualization II, Spatial Visualization I, and Mechanical Principles. Correlations between this composite and composites for the other four factors were then computed, assuming unit variance for each variable.

The SR composite was formed by averaging the correlations between Object Identification I and II and the other variables in the battery, and then using these average correlations and those for Object Recognition to define the composite. Thus, the two versions of the Flags test were treated as one test and combined with the AAF version of Cubes to make up the SR factor.

The resulting correlations are shown in Table 17. Correlations between several other tests of interest and these oblique factors were also computed and are shown at the bottom of Table 17. The correlations were all positive, and many were quite high. There is obviously a large general factor in the matrix.

When these factors were ranked according to their correlation with other factors, the order was the same for all factors except Ps. Vz was consistently first, followed by SO, SR, Ss and Ps. The Ps factor, on the other hand, had its highest correlation with Ss, followed by SO, SR, and Vz.

Insert Table 17 about here

As expected, the Pattern Assembly (i.e. easy form board) test correlated highest with the Ps factor. However, its next highest correlation was with the Vz factor, although the SR correlation was only slightly lower. To support the Zimmerman (1954) complexity hypothesis, Pattern Assembly should correlate higher with the SR factor than with the Vz factor. The opposite pattern was obtained here.

Position Orientation is an adaptation of Thurstone's Hands test. The hands test helped define the SR factor in the PMA analyses. Hands was not used to help locate the SR factor here because it may in part measure what Thurstone (1951) calls "Kinesthetic" ability and Guilford and Lacey (1947) call left-right discrimination. The high correlation between Position Orientation and the SR cluster is consistent with Thurstone's (1938) analyses. Its next highest correlation was with the SO composite. This is particularly suggestive. What has been called the Kinesthetic factor may represent the degenerate or simplest form of a spatial orientation test. Alternatively, the ability to make rapid

Table 17

Shepard Field Battery Reference Cluster Intercorrelations, and Correlations Between Selected Tests and Reference Clusters (After Guilford et al., 1952)

		Vz	SO	SR	Ss	Ps
	Visualization (Vz)	1.0				
	Spatial Orientation (SO)	.68	1.0			
	Spatial Relations (SR)	.58	.58	1.0		
	Spatial Scanning (Ss)	.55	.52	.46	1.0	
	Perceptual Speed (Ps)	.30	.35	.32	.42	1.0
32	Pattern Assembly	.32	.25	.28	.23	.40
23	Position Orientation	.25	.36	.42	.30	.31
17	Visual Memory	.28	.32	.28	.33	.33

left-right discrimination may be an important component of SO tests. In either case, the relationship between the two types of tests is clouded by individual differences in solution strategy. Thurstone (1951) observed that students appeared to solve his Hands test in different ways. Similarly, Barratt (1953) found that only 31 percent of the subjects in his sample reported solving items on the Guilford-Zimmerman Spatial Orientation test by projecting themselves into the situation. The relationship is also clouded by speed-power differences. Position Orientation is a highly speeded test while the SO defining tests are relatively unspeeded.

Another way to look at the relationships between these test clusters is within the context of a multi-trait, multi-method matrix (Campbell and Fiske, 1959). Such a matrix of average correlations within and between each of the clusters is shown in Table 18. The between cluster correlations are lower than the corresponding correlations between cluster centroids shown in Table 17. This is because averaging the correlations ignores the covariance between tests within a cluster in computing the cluster variance. The advantage of this method, however, is that it provides a way to compare within cluster correlations with the between cluster correlations. The average correlation within each group of tests that measure an hypothesized construct should be higher than the average correlation between members of that group and any other group.

Insert Table 18 about here

Inspection of Table 17 reveals that this principle holds for all the clusters except SR. Tests in this cluster correlate as well (or even slightly higher) with those in the Vz and SO groups than with each other. Recall that the SR group was formed by first averaging the correlations for the two variations of the Flags test (Object Identification I and II) and then clustering this score with the Object Recognition (Cubes) test. However, when the two Object Identification tests were not averaged first, but entered separately, the within SR correlation rose to .47. This is higher than the average correlation between SR and any other clusters. Thus, it appears that the analysis must include tests that are essentially parallel forms in order to define a coherent SR cluster. This is precisely what Thurstone and Thurstone (1941) did in defining PMA space as a composite of Cards, Flags, and Figures.

The same comment applies to the Ss and Ps clusters. Planning a Circuit is a parallel form of Maze Tracing. For the Ps cluster, the two

Table 18

Average Cluster Correlations for the Shepard Field Battery Reanalysis
(After Guilford et al., 1952)

	Vz	SO	SR	Ss	Ps
Vz	.55				
SO	.48	.58			
SR	.41	.41	.40		
Ss	.39	.38	.33	.48	
Ps	.21	.25	.23	.30	.42

Speed of Identification tests are even more obviously parallel. Thus, specific method variance plays a crucial role in defining factors that load primarily on aptitude constructs in the lower branches of the hierarchical model.

Factor Analyses. The question remains, however, whether the SO and Ss factors represent new subdivisions of the broad group spatial factor. To answer this question, the matrix of cluster correlations was factored in two ways.

First, the correlations between the four spatial clusters (excluding Ps) were factored by the centroid method. Maximum off-diagonal correlations were used as communality estimates, and two factors were extracted. The results are shown in the first two columns of Table 19. The mean absolute value of the residual correlations after two factors were extracted was .015, and the standard deviation .01.

Insert Table 19 about here

The first factor (I) represents the General factor (g or Gf) plus the broad group Spatial factor (S). The second factor sets Ss against the other three clusters, particularly SO and SR. Thus, the Ss factor appears to measure something different than the other three spatial clusters. However, it is impossible to say whether this extra component is a new aspect of spatial ability or some other dimension such as Visual Memory or Perceptual Speed.

The Visual Memory hypothesis was rejected since the two tests that defined this factor in the original analysis correlated poorly with all four spatial clusters. On the other hand, the Ps cluster had particularly strong correlation with the Ss cluster (see Table 17) and so this cluster was included in a new analysis.

Thus, the full matrix of cluster correlations in Table 17 was factored by the centroid method. Again, two factors were extracted using maximum off diagonal correlations as communality estimates. The results are shown in the third and fourth columns of Table 19. The Ss and Ps composites clustered together on the negative pole of factor II' while Vz, SO and SR all had positive loadings. It appears, then, that the major portion of the unique variance in Ss that appeared in factor II derives from Ps, not some new spatial subfactor.

Hypotheses about the SO Factor.

The situation was different for the SO cluster. While SO correlated strongly with Vz, it retained some unique, psychologically interpretable variance of its own. The high correlation with Vz may reflect one or more of

Table 19

Factor Analyses of Shepard Field Battery
Reference Cluster Correlations
(After Guilford et al., 1952)

	First Solution		Second Solution	
Cluster	I	II	I'	II"
Vz	.82	.06	.79	.22
SO	.81	.14	.80	.20
SR	.73	.13	.72	.14
Ss	.68	-.21	.71	-.20
Ps			.51	-.34

the following:

1. The processes involved in solving SO tests are, in part the same as those involved in solving Vz tests. Thus, if n components are required to solve Vz items, m of the same components are required to solve SO items (where $m < n$).

2. The processes involved in solving the two types of tests are different, yet individual differences in them are highly correlated in adult males.

3. Some subjects use Vz strategies and processes to solve some SO test items. Barratt (1953) provided some evidence for this hypothesis. He collected verbal reports of solution strategy on a number of spatial tests from 84 college males. The protocols of 58 students indicated they used a Visualization strategy on the Guilford-Zimmerman Spatial Orientation Test (which is based on Aerial Orientation)). Barratt described this strategy as "subjects rotated on moved stimulus and response problems but did not imagine themselves being re-oriented." (p. 24). Only 26 subjects were classified as using an SO stragegy in which they "imagined themselves being reoriented with regard to the problems." (p. 24). Just the opposite held for another SO test, the Industrial Aptitude Spatial Orientation Test. On this test, the protocols of 26 subjects were classified under the Vz strategy while 58 were classified under the SO strategy. Thus it is evident that at least some subjects use a Vz strategy to solve some SO test items. This would account for the high correlation between the two clusters in this battery. Those subjects who use an SO strategy account for the portion of unique SO variance that remains. This assumes that individual differences in either SO processes or strategy are, at least in part, independent of the corresponding Vz individual differences.

4. Vz and SO tests may require the same processes but differ only in the content on which the transformation operates. Thus, while it may appear that reorienting an imagined self in space is psychologically different from mentally manipulating an object in space, the two mental operations may represent the same set of processes operating on different inputs: an image of the self or an image of an object.

5. Vz and SO tests may require the same processes but differ only in the average complexity or the relative speededness of the tests. Table 20 provides some support for the speededness and, by implication, complexity, hypothesis. As before, speededness was estimated by dividing the number of items in the test by the total time alloted for the test. Complete data were not available for two of the tests either in the Guilford, Fruchter, and Zimmerman (1952)

report or the earlier, more detailed exposition of the AAF research program (Guileford and Lacey, 1947).

Insert Table 20 about here

With the exception of the Ss cluster, the average speededness estimates followed exactly the opposite rank order previously observed in the cluster intercorrelations (see Table 17). The estimates for the two tests in the Ss cluster are undoubtedly too low. They fail to include the number of path searches required for the solution of each item, but only the number of items in the tests. Speededness estimates for two other tests also may be inaccurate. Spatial Visualization II is probably less speeded than indicated in Table 20, as the test contains 12 items about which a total of 44 questions are asked. On the other hand, Speed of Identification A is probably more speeded than indicated, since solving most items should require evaluating more than one alternative. However, both of these possible changes in speededness estimates would produce even sharper support for the complexity hypothesis.

Thus, the ranking of the clusters in terms of their general factor loadings (which are proportional to the average cluster intercorrelations) was identical to the power ranking of the clusters, i.e., Vz, SO, SR, Ps. The Ss cluster fell between SR and Ps in terms of its average correlation with the other clusters. A good estimate of its speedness would probably result in the same placement.

Comparison of AFF and Thurstone Space Factors.

Finally, the present analysis affords a unique opportunity to relate the AAF Vz and SR space factors to the more familiar space factors constructed here. The AAF Vz factor appears to be essentially the same as that identified in the reanalyses of the Thurstone PMA data, and represented here by the combination of Spatial Visualization I and II, and Mechanical Principles. This last test defined or loaded highly on the Vz factor in ten studies (see Table 14). Spatial Visualization, following a distant second, defined the factor in two studies.

The identity of the various AAF SR factors is more problematic. The tests that most frequently defined the factor seldom appear in other factor analytic studies, probably because most were individually administered apparatus tests.

In the reanalysis of Perceptual Battery II (see Table 15) the SR factor was defined by familiar Thurstone tests, not by AAF SR tests. For example, the test Complex Coordination, which defined or loaded highly on the SR factor in numerous AAF studies, split its variance rather evenly between the SR and

Table 20

Speededness Estimates for Tests in the Shepard Field Battery Reference Clusters (After Guilford et al., 1952)

	Test	Number of Items	Time	Speededness (Items/min)	Average Speededness	Cluster
3	Spatial Visualization II	44	26	1.69		
19	Spatial Visualization I	60	39	1.54	1.74	Vz
52	Mechanical Principles	40	20	2.00		
24	Aerial Orientation	30	10	3.00		
11	Visualization of Maneuvers C	48	?	?	3.14	SO
15	Formation Visualization	49	15	3.27		
25	Object Identification I	28	7	4.0		
26	Object Identification II	30	6	5.0	4.30	SR
21	Object Recognition	100	25	4.0		
12	Plan a Circuit	42	14	3.00	2.71	Ss
14	Maze Tracing	29	12	2.42		
8	Speed of Identification C	?	?	?	12.00	Ps
53	Speed of Identification A	48	4	12.00		

Vz factors (see Table 15 and Figure 4).

The five tests that most frequently defined or loaded highly on the AAF SR factors (see Table 14) were included in the Sheppard Field Battery. Correlations between these tests and the five test clusters identified in this reanalysis are shown in the first five columns of Table 21. However, these correlations are difficult to interpret directly because of the large general factor. For example, since every cluster (except Ps) had its highest correlation with Vz, a high correlation between one of the AAF tests and Vz may relect the presence of the general factor, and not imply any special affinity between the test and the Vz factor. Therefore, the general-plus-broad-group spatial factor (Factor I' in Table 18) was partialled out of these correlations. The residual correlations are shown at the bottom of Table 21.

Insert Table 21 about here

Dial and Table Reading, which loaded significantly on the SR factor in three AAF studies (see Table 14) had a large positive residual correlation with Ps and a large negative residual with Vz. Thus, the portion of its common variance that is not accounted for by the general and group spatial factors is pitted against Vz (and SO) and aligned with Ps.

Instrument Comprehension, which defined or loaded significantly on the AAF SR factor in five studies, had a large positive residual correlation with SO. Thus, AAF SR factors defined by this test are probably better described as SO factors.

Complex Coordination, which defined or loaded significantly on the SR factors in thirteen AAF studies, is primarily a measure of the broad group spatial factor. This is consistent with the reanalysis of Perceptual Battery II (see Table 15) where Complex Coordination split its variance between the SR and Vz factors, and was thus almost completely accounted for by the broad group spatial factor. The present analysis suggests that its small special factor loading is on the Ps factor.

Two Hand Coordination loaded significantly on the SR factor in three AAF analyses. The residuals in Table 21 reveal that it was almost completely accounted for by the general plus broad group spatial factor.

Finally, Discrimination Reaction Time, which defined the SR factor in one analysis and loaded significantly on it in three others, had some residual linkage with the SR cluster. The slight positive correlation with Ps may be a

Table 21

Raw and Residual Correlations Between Shepard Field Battery Reference Clusters and AAF Spatial Relations Tests (After Guilford et al., 1952)

	Test	Raw Correlation					Residual Correlation				
		Vz	SO	SR	Ss	Ps	Vz	SO	SR	Ss	Ps
45	Dial & Table Reading	40	42	48	43	49	-10	-08	03	-01	17
51	Instrument Comprehension	46	57	38	32	29	00	11	-03	-08	-01
63	Complex Coordination	39	43	37	37	36	-04	00	-02	-01	08
64	Two Hand Coordination	31	26	25	24	16	03	-02	00	00	-02
65	Discrimination Reaction Time	39	46	47	37	35	-07	00	06	-04	05

Note. Decimals omitted.

methodological consequence of its negative correlation with Vz, or vice versa. At any rate, this is the only AAF SR test that clustered even moderately with the Thurstone-type SR factor.

Thus, it appears that most of the AAF SR factors are not the same as the Thurstone SR factor defined by tests such as Cards, Flags, Figures, and Cubes. Instead, these factors are more representative of the broad group spatial factor.

Finally, the AAF investigators had difficulty separating Vz, not from the factor they called SR, but from the various reasoning factors (Guilford and Lacey, 1947, p. 292). This difficulty is easily explained within the hierarchical model. The various reasoning factors were composed of g or Gf tests, and thus should overlap considerably with the complex Vz tests.

Thurstone's Later Work

The Thurstone Perceptual Battery.

At about the same time as the AAF work, Thurstone reported a factor analysis of perceptual tests (Thurstone, 1944). Several factors in that study are of interest here.

Perceptual Factor A. The tests which defined this factor are shown in Table 22 along with their factor loadings. Although Thurstone defined this factor as "speed and strength of closure," inspection of the tests indicates that it is close to the Space factor of his PMA study amd the Spatial Relations factor of Guilford and Lacey (1947).

Insert Table 22 about here

In the Shape Constancy Test, the subject must remember the apparent shape of a square piece of cardboard seen lying flat on a table across the room. The test had only one item. However, the fact that it defined the factor is congruent with the arguments of Smith (1964) on the nature of spatial ability. He contends that "the k-loading (and therefore the Vz-loading) of a test depends on the degree to which it involves the perception, retention, and recognition, (or reproduction) of a figure or a pattern in its correct proportions" (Smith, 1964, p. 96).

A second test under Factor A in Table 22, PMA Space, is a composite of the familiar Flags, Figures, and Cards tests (Thurstone, 1938; Thurstone and Thurstone, 1941). Gottschaldt Figures A and B were both highly speeded in this administration. Part A contained the easier items, and the score was the number

Table 22

Tests Loading on Selected Factors
(After Thurstone, 1944)

		Factor Loading			
	Test[a]	A	E	F	L
42	Shape Constancy	54			
52	PMA Space	51			
37	Gottschaldt Figures A	51	40		32
-46	Block Design	50	33		31
38	Gottschaldt Figures B	44	34		37
23	Brightness Contrast	38			
- 8	Hidden Digits	36			
- 1	Street Gestalt	35		53	
- 7	Dotted Outlines	35			
- 2	Mutilated Words	34		44	
-45	Mirror Test	30			
20	Two Hand Coordination		59		
-44	Hidden Pictures		47		
56	PMA Reasoning		42		
57	Color-Form Memory		38		
12	Peripheral Span-Single			65	
-11	Dark Adaptation			61	
- 3	Peripheral Span-Double			51	
- 5	Social Judgements-Time			39	
60	Sex	-26	36		

Note. Decimals and loadings less than .30 omitted.

[a]Minus sign means reflected error score.

of figures correctly identified in three minutes. For the more difficult part B, the dependent measure was the number of designs completed per minute. The more complex test (Part B) had a lower loading on the factor than the easier test (Part A) just as the more complex spatial tests had lower loadings on Thurstone's Space factor in the PMA study (Thurstone, 1938).

The Block Design test consisted of eight designs from the Kohs test (Kohs, 1923) with two demonstration items from the Wechsler-Bellevue (Wechsler, 1939). The score on this test was the sum of the times for the last five designs. Using latency as the dependent measure for the Block Design test and administering the Gottschaldt tests under speeded conditions may explain their higher than usual loadings on this Spatial Relations factor.

Finally, males outperformed females on this factor, which also supports a spatial interpretation.

Closure Speed. Factor F is the Closure Speed factor; tests that loaded on it are also shown in Table 22. The factor was defined by the test Peripheral Span-Single. In this test, the subject was asked to stare at a fixation point in the center of a blank screen. A capital letter was then flashed on the screen for 40 milliseconds at one of six distances on one of twelve imaginary radii centered at the fixation point. Score on the test was the number of letters correctly identified.

The test with the next highest loading was Dark Adaptation. In this test, the subject was asked to look at a brightly lit screen for two minutes. During this time a slide containing a capital letter was projected at various points on the screen, but the letter could not be seen as long as the screen was illuminated. The subject's task was to identify the letter as rapidly as possible when the light was turned off. Score on this test was the median response time for seven trials.

The next test on Factor F is the Street Gestalt. However, the dependent measure was the number of items with a response time of three or more seconds. Of course, this error score was reflected in the analysis. This score puts a heavier weight on rapid performance than the usual dependent measure of total number correct. Mutilated Words, which also loaded on the factor, was scored in the same manner.

Peripheral Span-Double was similar to the peripheral Span-Single test, except that here two letters were projected on the screen: one at the fixation point and the other at the radius of a circle around it. The subject's task was to press a key if the two letters were the same. Score on the test was the mean response time for test frames that contained identical letters. The Social Judgements time variable also loaded significantly on this factor. Here, the subject was presented pairs of adjectives (e.g., competent - tactful) and was asked to indicate which trait seemed most desirable. The dependent measure was the number of items with a response time of two or more seconds.

The common requirement of all of these tests seems to be the ability to make rapid identification or comparisons using incomplete or distorted information. Thus, the closure label may be misleading. In Dark Adaptation, Peripheral Span Single, the Street Gestalt, and Mutilated Words, subjects must match an incomplete visual image with a memory trace of that image. In Peripheral Span Double, they must do this for the peripheral letter, but then perform a visual match of the peripheral and central letters. It is noteworthy that Peripheral Span-Double is more centrally located and, by implication, more complex than Peripheral Span-Single in the multidimensional scaling of these data (see Figure 5 below).

Flexibility of Closure. Factor E is the Flexibility of Closure factor. Thurstone felt that the chief characteristic of this factor was the ability to break one gestalt in order to form another, or the freedom from what the Gestalt psychologists called Gestaltbindung.

The factor was defined by the test Two Hand Coordination. In this test the subject was required to tap corresponding quartile segments of two non-symmetrically labelled circles at the same time. Quartile number one was centered at nine o'clock on the first circle and at twelve o'clock on the second circle. The other three quartiles followed in clockwise succession on both circles. The dependent measure was the ratio of the sum of the number of taps in each quartile using each hand separately, to the number of simultaneous taps in corresponding quartiles using both hands. Thus, those with high scores on the test performed as well on the more difficult simultaneous task as they did when using each hand independently.

On the Hidden Pictures test, the subject was required to find six human or animal figures that were concealed in a larger distracting picture. Thus, it appears that the test requires the subject to break one gestalt and form another. Score on the test was the time to find the first five of the six hidden figures.

The contents of PMA Reasoning are uncertain. The original factor (Thurstone and Thurstone, 1941) was defined by Letter Series, Letter Grouping, and Pedigrees. Later versions of the PMA used Word Grouping and Figure Grouping to define this factor (Thurstone and Thurstone, 1947). The test was too easy for the college student volunteers in this study, and may have become more like a Perceptual Speed than a Reasoning test. Female superiority on the test supports this hypothesis.

In the Color-Form Memory test, subjects were shown a slide containing four colored designs for 40 milliseconds. They were then asked to name the designs and their colors. Two scores were computed: a ratio of the number of forms recalled to the sum of forms plus colors recalled, and the number of forms plus colors recalled. The ratio score did not correlate with other tests in the battery and was excluded from the factor analysis. Thus, those with high scores on the Color-Form Memory test were able to recall both colors and forms.

The last two tests with only minor loadings on the factor were the difficult Gottschaldt Figures test (part B) and Block Designs.

While it appears that breaking Gestaltbindung is a significant element, the more pronounced communality is the ability to do two things at once. Thus, performance on these tasks may be a function of the degree of hemispheric dominance. Those who are less lateralized may be able to keep both hemispheres working on separate tasks without one hemisphere interfering with or dominating the other. This is particularly evident in the test that defined the factor, Two Hand Coordination. It also seems to characterize Hidden Pictures. In this test, one must simultaneously break one gestalt (an analytic left hemisphere function?) while forming another (a right hemisphere function). Similarly, those who were able to name the colors and retain images of the forms at the same time would perform well on the Color-Form Memory test. Superior female performance on this factor supports this hypothesis, as women tend to be less strongly lateralized. This raises the interesting possibility of using relative performance on the Space and Flexibility of Closure factors to estimate the degree of lateralization.

Factor L. The final factor of interest here had only marginally significant loadings from the two Gottschaldt Figures tests and Block Designs. Thurstone called the factor a residual, and did not attempt interpretation. The factor represents the residual Gf covariation in these three tests that was not captured by the Spatial Relations factor. If other complex spatial tests such as Paper Folding or Surface Development had been included in the analysis, the L factor probably would have become the Vz factor that appeared so often in other analyses.

Reanalysis. The correlation matrix of all the tests listed in Table 22 was scaled using the KYST program (Kruskal, Young & Seery, 1973). Raw correlations were used because Thurstone did not report reliabilities. The initial configuration was generated by the metric Young-Torgerson procedure. The nonmetric configuration was iterated 23 times in three dimensions and 22 times in two dimensions. Stress values (formula 1) were .120 and .175, respectively. The final two dimensional configuration is shown in Figure 5. The approximate positions of the four factors are also shown in this plot.

Insert Figure 5 about here

The clusters shown in Figure 5 were generated by Johnson's (1967) HICLUS program using the diameter method. There were a number of discrepancies between the cluster and factor analyses. For example, PMA Reasoning, Two Hand Coordination, and Color Form Memory formed a cluster, while Hidden Pictures clustered with the Closure Speed group. This is not unusual, as Hidden Pictures sometimes falls on the Closure Speed factor and sometimes on the Flexibility of Closure factor (Botzum, 1951; Pemberton, 1952). This suggests that Hidden Pictures requires both Flexibility of Closure and Closure Speed, or is solved in different ways by different subjects.

The Closure Speed cluster is also different than the Closure Speed factor. In particular, the two tests that defined the factor (Peripheral Span-Single and Dark Adaptation) broke away and formed a sub-cluster with the Peripheral Span tests and Social Judgement Time. Dotted Outlines and Hidden Digits also entered the cluster at later steps in the analysis. These tests would move to the periphery and define the usual Closure Speed factor in a battery of more complex tests. Here, however, the presence of the simple tasks alters both the scaling and factor structure, and, in a way, permits a cleaner psychological interpretation.

The Thurstone factors appear more useful than the clusters, and are more consonant with the multidimensional scaling. A test can load on more than one factor but can only belong to one cluster. The exclusionary nature of the hierarchical clustering algorithm is particularly unstable at the later stages of the cluster analysis.

As in other analyses, the tests that defined the Thurstone factors were more peripheral, while the more centrally located tests tended to load on more than one factor. The particularly close clustering of the two Gottschaldt

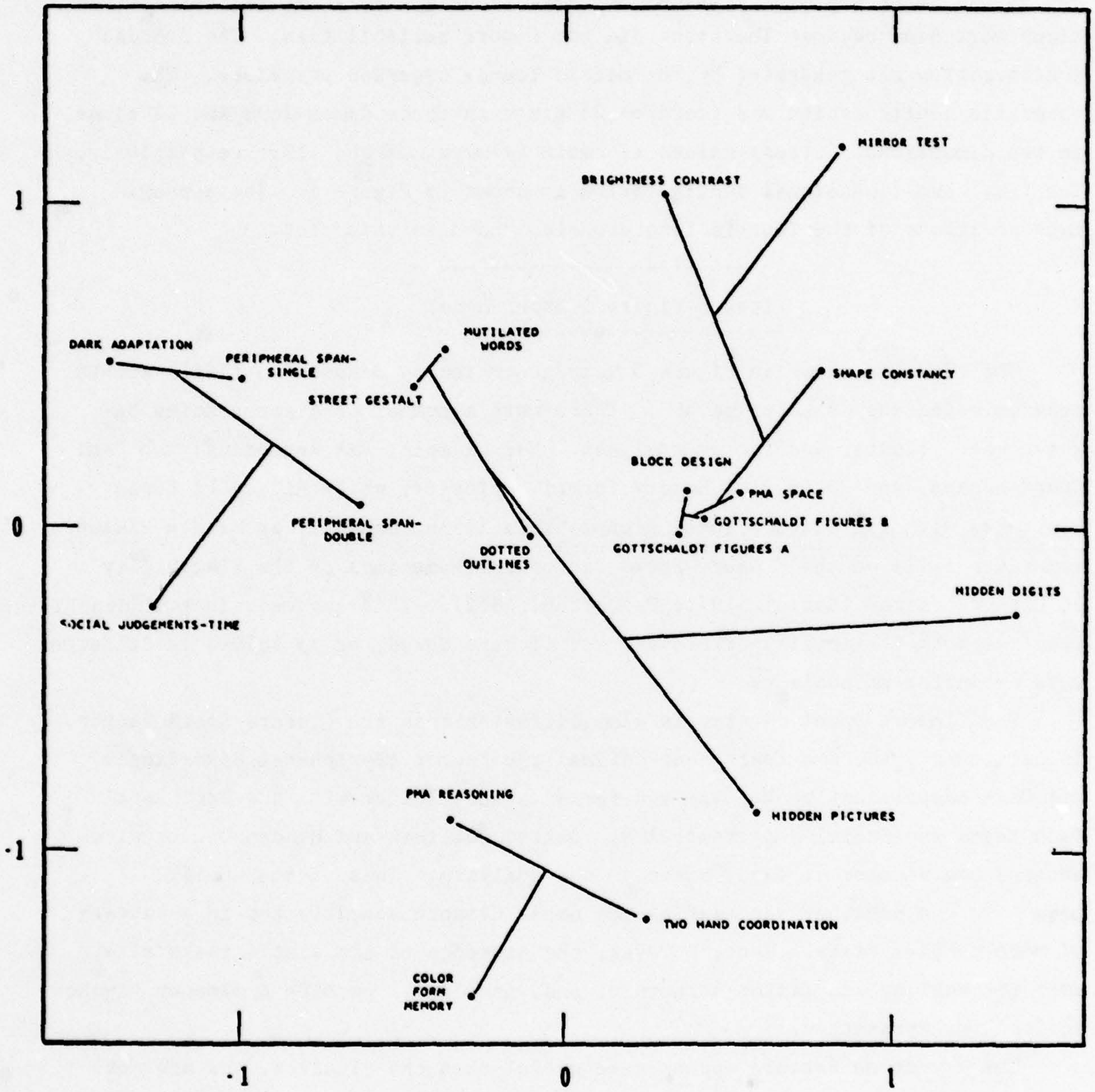


Figure 5. Two dimensional scaling of spatial tests (After Thurstone, 1944).

tests, Block Designs, and PMA space is more the result of low correlations between other variables than a reflection of an unusually strong relationship between these tests. In fact, the highest correlation in the submatrix analyzed here was only .57 between Gottschaldt Figures A and Block Design.

Thus, the reanalysis accepts the Thurstone factors, but with different psychological interpretations. In addition, projecting these factors onto the two dimensional scaling revealed that at least a two level (i.e., bi-factor) hierarchy is present, but overlooked by Thurstone's analysis.

The Thurstone Mechanical Aptitude Battery

Thurstone's results. In 1951, Thurstone reported a study of mechanical aptitude. A large number of familiar spatial tests were included in the test battery, and so the study merits close scrutiny.

A battery of 32 group tests, 25 individual tests and two interest scales were administered to 350 boys. All were juniors in a Chicago technical high school. The main objective of the study was to compare the test scores of two subgroups at the extremes on mechanical experience and interest.

Unfortunately, correlations and factor analyses were reported only for the 32 group tests. Five of these tests were "classified" and are not described in the report. Thurstone extracted ten centroid factors from the correlation matrix for these 32 tests. The solution was iterated once and then rotated to oblique simple structure. A simplified version of the resulting factor pattern matrix is shown in Table 23. Correlations between the factors are shown in Table 24.

Insert Tables 23 and 24 about here

The factors were labeled Induction (I), Space one, two, and three (S1, S2, S3), Kinasthetic (K), Memory two and three (M2, M3), first and second Closure (C1, C2), and residual. Five tests had no significant loadings on any of the factors: Block Counting, Identical Forms, Mutilated Pictures, Picture Squares, and Figure Grouping.

Five of the factors are of particular interest here. Factor S1 was defined by Figures and Cards, followed by Rotation of Solid Figures, and Reversals and Rotations. Thurstone interpreted this factor as representing "the ability to visualize a rigid configuration when it is moved into different positions." (p. 18).

Table 23

Simplified Oblique Factor Pattern Matrix for Mechanical Aptitude Battery (After Thurstone, 1951)

	Test	I	S1	S2	S3	K	M2	M3	C1	C2	Res.	h^2
1	Block Counting									20		53
2	Paper Puzzles			33						32		47
3	Cards		52									74
4	Figures		60									79
5	Hands				22	52						44
6	Copying						21			36		50
7	Bolts	20		20		52						58
8	Gottschaldt Figures							23		30		55
9	Street Pictures			21					49			33
10	Mutilated Words								43			39
11	Designs									38		46
12	Memory for Pictures						53	29				54
13	Visual Memory							39				22
14	Mechanical Movements	21		48								60
15	Surface Development			37	25					26		71
16	Reversals and Rotations		33		22					24		60
17	Lozenges A				31							64
18	Cubes				46							57
19	Identical Forms						20		27			46
20	Mutilated Pictures				21			24	25			38
21	Jig-Saw Pieces										27	25
22	Memory for Geometric Design						47	22				47
23	Picture Squares				21				30			34
24	Letter Series	52										70
25	Letter Grouping	50										65
26	Figure Analogies	30										49
27	Figure Grouping									22		39
28	Mechanical Comprehension Book I			60						21		74
29	Rotation of Solid Figures		40								24	53
30	Block Assembly			27				42			34	58
31	Mechanical Experience			66					22			57
32	Electrical Experience			59								58

Note. Decimals omitted.

Table 24

First Order Factor Correlations
(After Thurstone, 1951)

	I	S1	S2	S3	K	M2	M3	C1	C2	Res.
I										
S1	.38									
S2	-.08	.11								
S3	.46	.31	.10							
K	.23	.46	-.03	.19						
M2	.20	.27	.25	.20	.30					
M3	.56	.41	-.04	.33	.24	.23				
C1	.32	.16	-.31	.09	.14	.13	.27			
C2	.63	.53	-.08	.33	.27	.09	.38	.29		
Res.	.25	.02	-.17	.16	-.02	-.14	.01	.00	.17	

The second Space factor was defined by Mechanical Experience, Mechanical Comprehension, and Electrical Experience, with Mechanical Movements and Surface Development also loading significantly. This factor was interpreted as "the ability to visualize a configuration in which there is movement or displacement among the parts of the configuration." This ability was considered to be the central characteristic of mechanical ability.

The third Space factor was confined to the Lozenges and Cubes tests and Thurstone did not attempt to interpret it. However, in an earlier report (Thurstone, 1950) he speculated that the factor might represent the ability to think about spatial relations in which the body orientation of the observer was an essential part of the problem. However, there appears to be little similarity between the psychological processes tapped by Lozenges and Cubes, and the cluster of AAF tests for which Guilford first proposed this interpretation (see p. 45).

The Kinesthetic factor is also of interest. It was doublet composed of Hands and Bolts. Thurstone arrived at this label by observing students perform various contortions with their hands while solving the tests. He also noted that some students were able to solve the items "in their heads" and did so much more rapidly than those who were constantly referring to their own hands. This suggests that the tests were tapping different abilities in different students. It would also explain why the Hands test sometimes clusters weakly with other Spatial Relations tests such as Cards or Figures (e.g., Thurstone, 1938) and sometimes defines a separate factor (e.g., Guilford and Lacey, 1947).

Finally, the second Closure factor was defined by Designs, Copying, Paper Puzzles, and the Gottschaldt Figures, all with low loadings (.30 - .38). As the label suggests, this factor was thought to represent the same aptitude tapped by the flexibility of closure factor identified in the factor analysis of perceptual tests (Thurstone, 1944). The factor was extremely oblique in this solution, however. It correlated .63 with the Induction factor (which was defined by Letter Series), .53 with S1, and .33 with S3 (see Table 4). With the exception of factor S2, which was defined by the experience tests, the intercorrelations of these factors were quite high. There is clearly a higher order factor in the matrix.

Reanalysis. The reanalysis of these data took many different forms, only a few of which can be mentioned here. The ultimate goal was to construct a hierarchical factor representation of the correlation matrix. The most promising technique appeared to be one outlined by Wherry (1959). The procedure starts by extracting oblique first order factors by the multiple group method and then determining their intercorrelations. Second order oblique factors are extracted from this matrix, and the process is repeated until just one factor can be extracted. The series of factor structure matrices are then transformed into one orthogonical, hierarchical matrix.

Disattentuation. The correlation matrix for the 32 group tests in the Thurstone Mechanical Aptitude Study were first disattentuated and then cluster analyses and multidimensional scaling were performed on the matrix. Thurstone's split half reliability coefficients were used in the disattentuation. These coefficients were undoubtedly too high for the speeded tests, but this underestimates the disattentuated correlation. This was more desirable than using communalities (from the Thurstone solution or a component model of the present solution) that would underestimate the reliability of the more specific tests, and thus overestimate the true correlation.

Level one clustering and scaling. Maximum method cluster analysis was then performed on the disattentuated matrix using Johnson's (1967) HICLUS program. The results are shown in Figure 6. A minimum method analysis was also performed, but did not yield clearly defined clusters.

Insert Figure 6 about here

A nonmetric multidimensional scaling was also performed on both the raw and disattentuated correlation matrices using the KYST program (Kruskal, Young and Seery, 1973). The disattentuated solution was clearer, and more congruent with the corresponding cluster analysis, and so only this solution is reported here. The initial configuration was generated by the metric Young-Torgerson procedure. The configuration was iterated 24 times in three dimensions and 20 times in two dimensions. Final configurations were rotated to principal components. Stress (formula 1) was .159 in three dimensions and .213 in two dimensions. The two dimensional solution was selected because it was more interpretable and more consisent with the cluster analysis. Further, the slight reduction in stress at three dimensions did not warrant retaining an additional dimension.

The results of this analysis are shown in Figures 7 and 8. Figure 7 includes the test names and the level one clustering, while Figure 8 shows a more complete version of the hierarchical clustering from Figure 6 superimposed on the scaling representation.

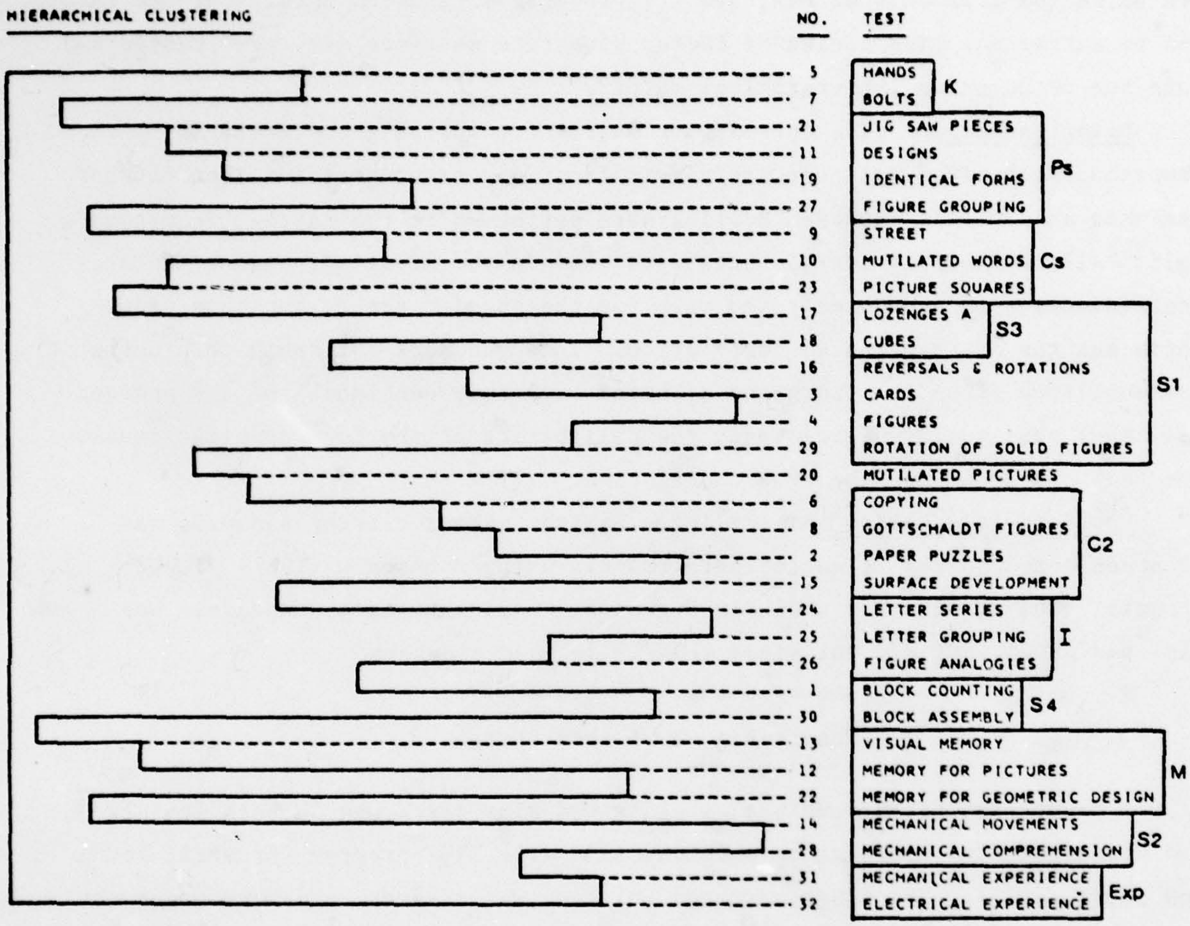


Figure 6. Clustering of the disattenuated correlations by the diameter method (After Thurstone, 1951).

Insert Figures 7 and 8 about here

The cluster analysis suggested that the original set of 32 variables could be partitioned into ten clusters. These clusters are indicated in Figure 6 by the labels Thurstone attached to similar factors in his oblique factor pattern matrix. Several clusters are either new or sufficiently different from the Thurstone factors to warrant comment.

With the exception of the Designs test, none of the tests in the Perceptual Speed (Ps) cluster had significant loadings on Thurstone's factors. The Designs test had the highest loading (.38) on his Flexibility of Closure (C2) factor. It clusters with Ps tests here partly because of the disattentuation process. This test is quite speeded and its reliability coefficient (.96) was undoubtedly inflated. On the other hand, the reliability of the Gottschaldt Figures test was estimated to be .78. Thus, the Gottschaldt Figures test was pulled closer to the other complex, power tests that also had lower, more realistic reliability coefficients. This was evident in a comparison of the two dimensional scalings of the raw and disattentuated correlations. Of course, the Designs test clustered with the Ps tests only because it correlated higher with the Ps cluster than the C2 cluster at that point in the analysis. The exclusionary clustering algorithm prohibits a test from belonging to more than one cluster. However, in the final hierarchical solution, the Designs test emerged with a small loading on the C2 factor.

The clustering of the Mutilated Pictures test was also problematic. The cluster analysis in Figure 6 indicated that it did not cluster neatly with any of the other variables. The multidimensional scaling was equally indeterminate. Consequently, a second diameter method cluster analysis was performed in which this test was clustered with the ten first order clusters. In this analysis, Mutilated Pictures clustered with the Cs tests, and so it was added to that group.

It would have been preferable to let the test stand alone. However, the test would define a "factor" in the final hierarchical matrix. Thus, it is preferable to cluster a test with other tests if possible. This problem does not emerge at higher levels in the analysis. Clusters that are not clustered again at a level each define dummy factors that appear as a column of zeros in the hierarchical matrix. Of course these "factors" are not reported in the final factor structure matrix.

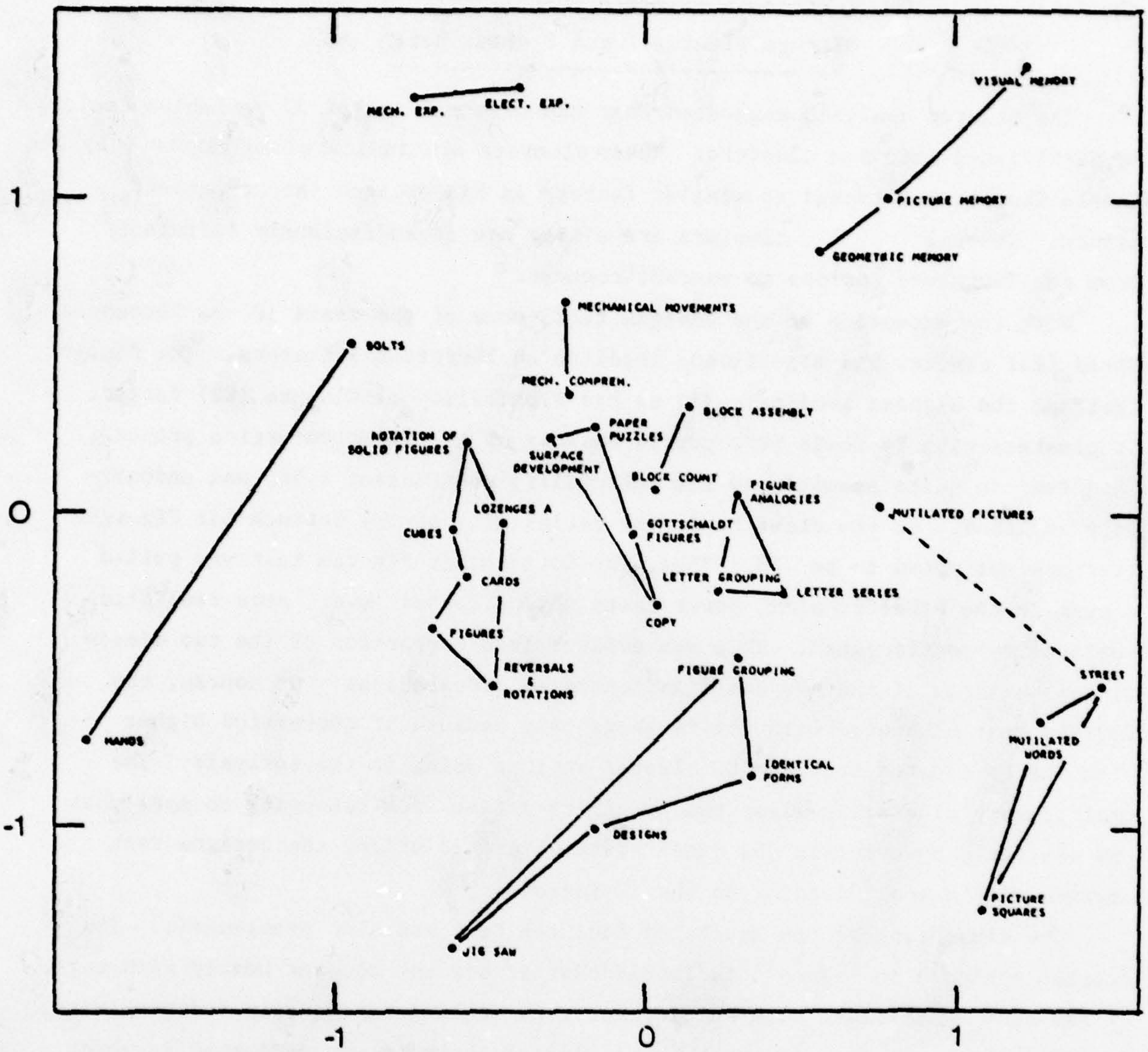


Figure 7. Multidimensional scaling of the Mechanical Aptitude Battery (After Thurstone, 1951).

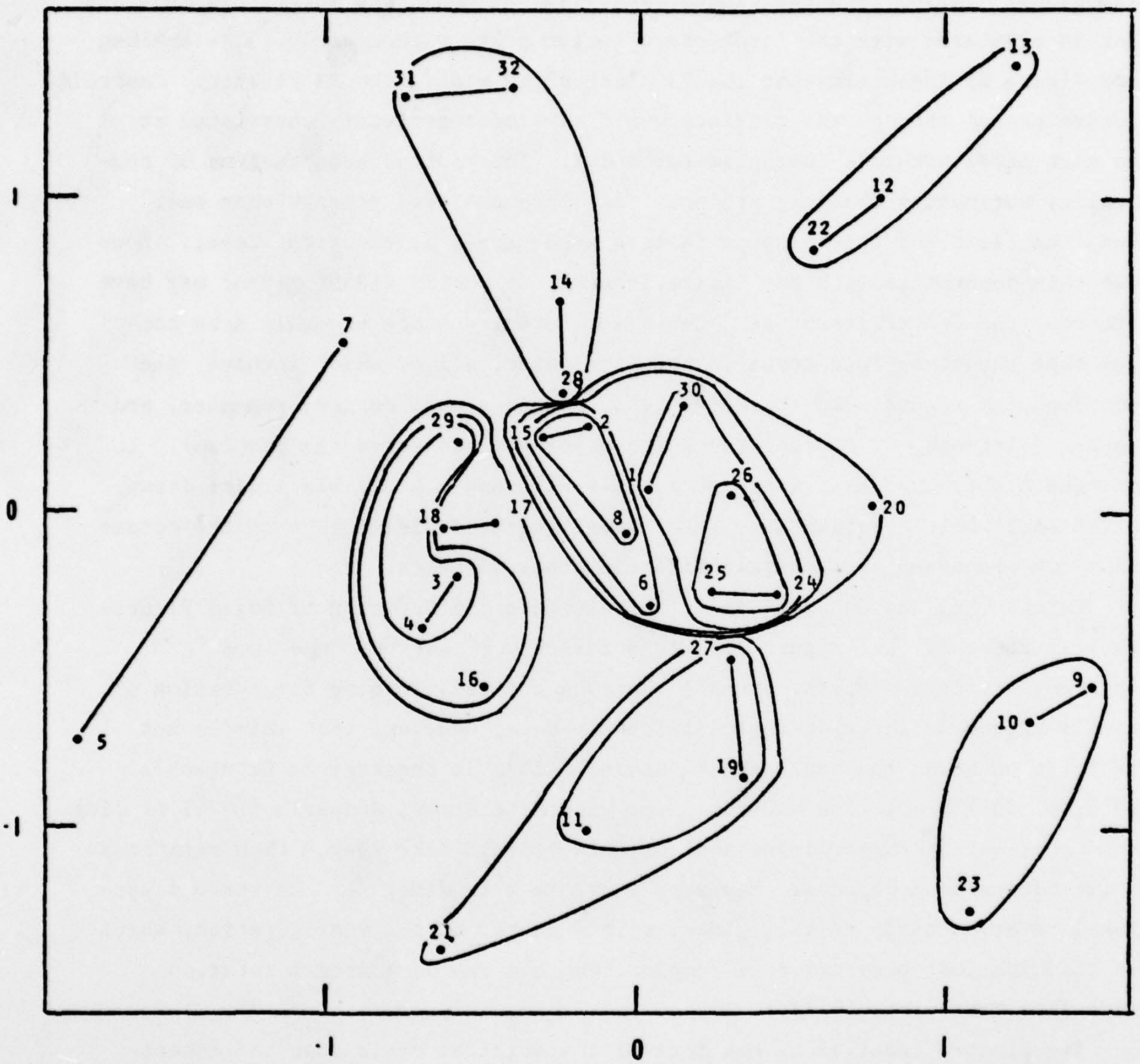


Figure 8. Diameter method hierarchical clusters superimposed on the scaling of the Mechanical Aptitude Battery (After Thurstone, 1951).

Thurstone's third Space factor was merged with his first Space factor in a cluster labeled S1 in Figure 6. This was more a function of the multidimensional scaling than the cluster analysis, although the latter did indicate that S3 clustered with the first space factor tests rather early. The scaling (see Figure 8) indicated that the S3 cluster lay within the S1 cluster. Centroid vectors passed through the clusters would be almost perfectly correlated at the next level within a common factor model. This introduces problems of communality estimation that may produce final communalities greater than one. Thus, the first and third space factors were merged at the first level. However this compromise with the limitations of the Wherry (1959) method may have distorted the factor structure. Cubes and Lozenges A are probably more complex than the other four tests in the S1 cluster, all of which involve the rotation of a figure. In the Cubes test, subjects must rotate, remember, and compare (although, of course, there are other ways to solve the problem). In Lozenges A they must keep track of a small hole and a heavy black line drawn on the card while rotating it. Thus, these tests require more than the rotate and match processes that characterize the other S1 tests.

Further, S3 was embedded in S1 only because the Rotation of Solid Figures test lay above it (see Figure 7). This test lay closer to the two mechanical tests and Bolts, probably because they all involve the rotation of a solid figure in three dimensional space. Note, however, that this is not the facet on which the tests are clustered. This is contrary to Cronbach's (1970, p. 332) prediction and congruent with Metzler and Shepard's (1974) finding that rotations of three dimensional objects did not take longer than rotations of two dimensional objects. However, there is a tendency for the three dimensional rotation tests to fall closer to the center of the configuration, which may indicate that they are more complex than the two dimensional rotation tests (see Marshalek, 1977).

The cluster labelled C2 has more of a spatial emphasis than the corresponding factor in the Thurstone solution. In particular, the Surface Development and Paper Puzzles tests split their variance between the C2 factor and other spatial factors in the Thurstone solution. Again, this was a consequence of the exclusionary clustering algorithim that was modified slightly in the final hierarchical matrix. There, Copying defined the C2 factor even though it was the last test to cluster here. On the other hand, the C2 factor was the only factor on which Copying and the Gottschalt Figures loaded significantly in the Thurstone solution.

The Induction factor was almost identical to the corresponding Thurstone factor. However only one memory cluster appeared here instead of two memory factors as in the Thurstone solution. Memory for Pictures and Memory for Geometric Designs clustered strongly, and were the two tests that defined Thurstone's M2 factor. The third memory test (Visual Memory) clustered only weakly with these two. This test split away and, along with Block Assembly, defined Thurstone's M3 factor. However, in the present solution Block Assembly clustered strongly with Block Counting in a factor called S4. This may have been a consequence of overlapping content, but since Block Assembly was one of the classified tests it is impossible to know for sure.

Finally, Thurstone's second Space factor was split here into a mechanical cluster (S2) and an experience cluster. Even though these two clusters came together early in the hierarchical cluster analysis, there was good psychological reason for keeping them apart. Scores on the experience tests require crystalized knowledge of mechanical and electrical concepts; knowledge that is highly dependent on experience, attitudes, motivation, as well as ability. It is of some interest to see how these tests relate to the mechanical comprehension tests, which utilize mechanical content but require spatial reasoning. However, allowing the two to come together and define a "spatial" factor is misleading, for the two overlap primarily in content. Thus, Thurstone's second Space factor, which was defined by the Mechanical Experience test, was probably more of a mechanical knowledge factor than a space factor.

High order clustering. These ten first order clusters were then clustered again by the diameter method. The results are shown in Figure 9. The first space (S1) and Perceptual Speed tests came together in a cluster called Spatial Relations (SR). This is something of a misnomer as S1 alone is usually what is meant by SR. The important point is that the speeded space tests came together in one cluster while the three clusters of relatively unspeeded, complex tests came together in another cluster (here labelled Vz, again for continuity with previous work).

Figure 10 shows a two dimensional scaling of the ten first order clusters. The S1 cluster was closer to the S4 and C2 clusters than to the Ps cluster. This suggests that all four of the spatial clusters (S1, S2, S4, and C2) could have been clustered into a broad group spatial factor at this level. In fact, if S2, S4 and C2 are clustered into the Vz factor (as shown) but S1 and Ps are not clustered, then S1 clusters with the Vz cluster and not with the Ps cluster at the next level.

Insert Figures 9 and 10 about here

Thus, one is faced with a double problem of not just where to draw the line between clusters, but when. Any such decision is necessarily somewhat arbitrary. The goal here, however, was to construct a complete hierarchical representation of the data, and so the initial clustering that represented two factors at this level was retained.

These seven second order clusters were again clustered by the diameter method. The results are also shown in Figure 9. The clustering indicates that there was really only one cluster at this level. However, the SR and Vz clusters were clustered together, in order to represent the broad group spatial factor in a complete hierarchical model.

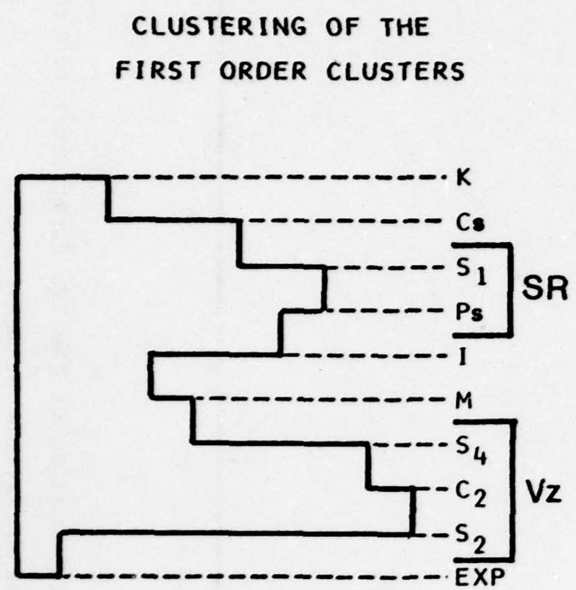
Hierarchical factor analysis. The results of the cluster analysis were used to construct a series of weight matrices for the multiple group factor analysis. The first matrix created the ten first order clusters from the 32 variables; the second created the seven second order factors from these; while the third defined the six third order factors; and the fourth brought the six third order factors into one general factor.

As explained previously, however, variables that are not reclustered at a level do not define factors at that level. In the present case, there were just ten first order factors, two second order factors, one third order factor and one fourth order factor. A common factor model was employed at all levels. Test reliabilities were used as communality estimates at the first level, and the maximum off diagonal correlation at all other levels. If a cluster was not reclustered, its communality was estimated to be 1.0 at that level.

The four oblique structure matrices were then transformed into one orthogonal hierarchical matrix by the Wherry (1959) procedure. This matrix is shown in Table 25. Factor loadings less than .10 were omitted. The table shows a large general factor, labeled Gf, the broad group spatial factor (S), two second order group factors (SR and Vz) and ten first order factors.

Insert Table 25 about here

Tests and clusters near the center of the scaling representation in Figure 7 loaded strongly on the Gf factor. If the verbal domain were better represented in this battery, Gc and G factors would also appear and capture much of the



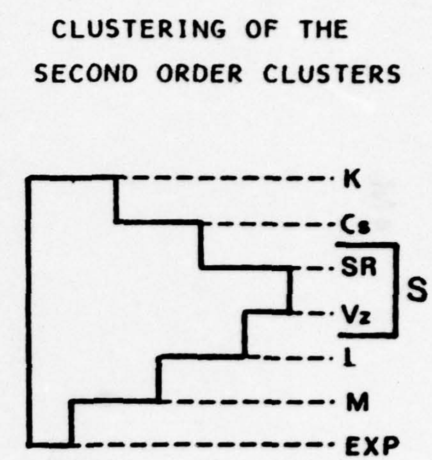


Figure 9. Clustering of the first and second order clusters by the diameter method (After Thurstone, 1951).

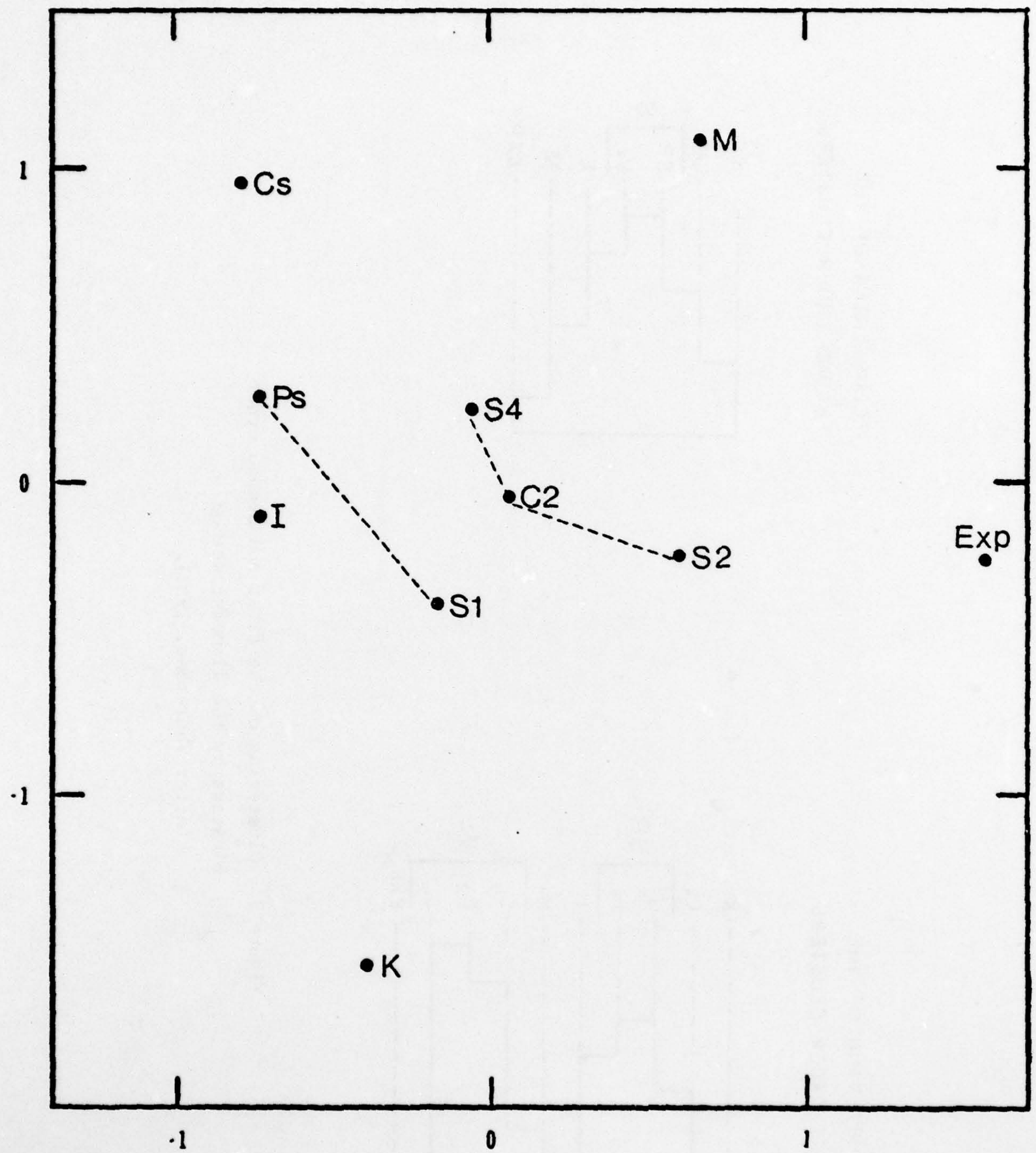


Figure 10. Two-dimensional scaling of the ten first order clusters (After Thurstone, 1951).

Table 25

Orthogonal, Hierarchical Factor Structure Matrix for the Thurstone (1951) Mechanical Aptitude Battery

		Gf	S	SR	Vz	Ps	S1	C2	S4	S2	K	Cs	I	M	Exp.
5	Hands	33									73		-11		
7	Bolts	56									64		11		14
11	Designs	49	25	13		41		10							-19
19	Identical Forms	55	16	11	-10	45						19			-10
27	Figure Grouping	48	17	06		29				11	-10				
21	Jig-Saw Pieces	34	26	13		44							-14		-17
16	Reversals & Rot.	53	32	13			44						-17		-22
4	Figures	56	29	14			51				11		-13		-14
3	Cards	59	29	12			52						-12		-11
18	Cubes	53	20	07			33							-12	
17	Lozenges A	64	22	08			37				16				
29	Rota. Solid Fig.	52	22	08			38								12
15	Surf.Development	69	27		15			36							
2	Paper Puzzles	52	25		15			33			-11				
8	Gotts. Figures	61	25		12			28							
6	Copying	59	24	08	07		-11	46							-11
1	Block Counting	65	21		38				46						
30	Block Ass'n	55	17		30				31					10	
28	Mech.Comprehen.	63	13		19					31					37
14	Mech.Movements	56	09		21					42		-11	12		36
23	Picture Squares	37									-10	48			-13
9	Street Pictures	37										50			
10	Mutilated Words	44										50			
20	Mutilated Pict.	53										38			
25	Letter Grouping	62											54		
24	Letter Series	62											60		
26	Figure Analogies	60								-12			51		
22	Geometric Mem.	48												52	
12	Picture Memory	47												62	
13	Visual Memory	28												44	
31	Mech.Experience	41											-12		69
32	Elec.Experience	46											12		74
Percent of Variation		27.8	3.0	0.5	1.3	2.2	3.6	1.8	1.1	1.2	3.3	3.0	2.9	4.9	4.9

Note. Decimals omitted for factor loadings.

variance in the lopsided Gf that appeared here.

The SR and Vz factors were both small, suggesting that this may not have been a meaningful distinction at this level in the hierarchy. In particular, it suggests that a two or, at most, three level hierarchy would be sufficient for these data. Thus, the four first order space factors could be immediately clustered into a broad group spatial factor, as one of the scaling analyses suggested.

Also note that the K factor did not cluster with the other space tests, and that the mechanical comprehension tests emerged with substantial loadings on the experience factor.

However, substantive generalizations about the nature of spatial ability are difficult to make on the basis of this factor analysis. The boundaries between factors are arbitrary, especially at the intermediate levels. The factor structure can be drastically altered by when and where the lines are drawn between clusters.

Multidimensional scaling was initially employed in this analysis as an adjunct to the cluster analysis, which in turn was an aid for the multiple group factor analysis. The real goal was the hierarchical factor structure matrix shown in Table 25. It appears, however, that the initial multidimensional scaling of the disattenuated correlations was the most promising way to represent the complex web of relationships among the tests.

<u>A componential interpretation</u>. The pattern of test points in the multidimensional scaling representaiton can be readily interpreted in componential terms. As used here, "component" refers to a functionally discrete mental process. For example, mental rotation, matching, and storing in memory are examples of component processes.

The peripheral clusters in Figure 7 may be viewed as components of varying degrees of purity or clarity. More central clusters may then be defined by combining the component processes. Note that only a restricted set of components appears here. First, the set includes only those component processes actually required by the tests selected for inclusion in the battery. Of this number, only those in which there are individual differences within the range of component difficulty required by the tests will surface. Finally, the test will cluster with others in the battery to the extent that its components overlap with those required by other tests. All of this will be blurred by measurement error and individual differences in how students solve test items.

The common denominator of tests in the S1 cluster appears to be mental rotation. The subset formed by Cards, Figures, and Reversals and Rotations formed the tightest cluster within the group. Cubes and Lozenges A may have fallen within this cluster because the primary source of individual differences in these tests lies in the speed and power of mental rotation.

But other components (such as memory) produced individual differences in these tests, and so they formed a subgroup. The rotation component was also strongest in the Rotation of Solid Figures test. However, this test fell near the Mechanical tests and Bolts, suggesting that experience with three dimensional rotation problems may have some important influence on test performance.

The location of the Bolts test midway between Hands (left-right discrimination component), the mental rotation cluster, and mechanical experience cluster, suggests that all three of these components produced individual differences in the test. Similarly, Mutilated Pictures fell at the intersection of the visual memory component and the closure component. Performance on this test (which is similar to the WAIS Picture Completion subtest) appears to depend on the ability to retrieve features of similar images from long term memory and mentally "close" the incomplete image. Of course, the location also suggests that the test was solved in different ways by different subjects.

The common process component in the Ps cluster appears to be speed of matching visual stimuli. The cluster is loose, suggesting that test content or other components influence test performance. The proximity of this cluster to the mental rotation cluster suggests that speed of matching may be an important component of tests in the S1 cluster, particularly those nearest the Ps cluster.

Tests at the center of the figure may be more complex because individual differences in several components influence test performance. The configuration suggests that visual memory was particularly important in the block tests. Individual differences in a number of components influenced performance in tests like the Gottschaldt Figures, Surface Development, and Paper Folding.

Of course, these interpretations are speculative. Multidimensional scaling, like factor analysis, cannot produce something out of nothing. The tests are impure; most, if not all, may be solved in more than one way. Further, all require multiple cognitive operations for solution, and so tests may correlate for a variety reasons. Nevertheless, if there is any informaiton in this sort of correlational research that would provide direction for research on a process understanding of aptitude, multidimensional scaling of the correlations yields the most psychologically rich representation. It provides a rough and usable map of the terrain. The trouble lies not in the multivariate methods but in the tests. Process, content, and complexity are all intertwined. The real task is to identify these dimensions and then develop clear ways of measuring them.

Guilford's Postwar Work

Guilford and his students reported several studies of spatial ability in the period between the completion of the AAF work (Guilford and Lacey, 1947) and the formulation of the Structure of the Intellect (SI) model (Guilford, 1956). These studies attempted to investigate hypotheses about the nature of spatial ability that had surfaced during the AAF work.

Michael, Zimmerman, and Guilford (1950)

In the first study, Michael, Zimmerman, and Guilford (1950) enumerated several hypotheses about the differences between the Spatial Relations and Visualization factors. They hypothesized that the Spatial Relations factor represented "the ability to comprehend the arrangement of elements within a visual stimulus pattern, primarily with reference to the human body." (p. 189-190). Thurstone's Cubes and Flags, and the Guilford-Zimmerman Spatial Orientation tests were hypothesized to be exemplary measures of this factor. Thus, the factor they called Spatial Relations was a composite of the factors labeled Spatial Relations and Spatial Orientation in this review.

The Visualization factor was thought to represent the ability to manipulate visual images. Thurstone's Punched Holes and Form Board, and the Guilford-Zimmerman Spatial Visualization tests were chosen to represent it.

Michael et al. (1950) entertained several hypotheses about the differences in cognitive processes or test requirements that might underly the distinction between the two factors. However, they did not investigate these hypotheses directly in the study, but rather factored the correlation matrix, and used the hypotheses to explain unexpected results. Some introspective reports were gathered, but were not utilized in any systematic manner. The hypotheses were:

1. Response format. The subject must draw his response in Punched Holes and Form Board, whereas all the other space tests were multiple choice format.

2. Speed of response. This hypothesis was previously indicated in the AAF work. The distinction between the two factors may in part reflect a speed-power difference. Spatial Relations tests tend to be given under relatively speeded conditions, whereas Visualization tests tend to be administered under fairly liberal time allowances.

3. Task complexity. This was defined as "the number of steps entering into the performance of an item" (p. 192). More complex tasks may require Visualization.

4. Psychological distance. Again, this was a reiteration of an AAF hypothesis. There, it was hypothesized that the ability to visually maneuver an airplane as if the examinee were in a position outside the cockpit would require Visualization ability, while the ability to imagine the maneuvers as if the subject were sitting in the cockpit would require Spatial Orientation, or Spatial Relations ability.

5. Strategy. Finally, the authors recognized that some subjects use Visualization ability to solve Spatial Relations tests. They reported that the introspective accounts of "many subjects" supported this hypothesis, but did not factor within strategy groups or report other relevant analyses.

Results. Six spatial tests and eight reference tests were administered to 360 students in beginning psychology at Rutgers University. Nine centroid factors accounting for a total of 52.6 percent of the variance in the tests were extracted from the correlation matrix. Factors were then graphically rotated to orthogonal simple structure. Six of the factors were labeled as follows: Visualization (Vz), Verbality (V), Numerical Facility (N), General Reasoning (R), Spatial Relations (S), and Perceptual Speed (P). Two factors could not be labeled, and the ninth was a residual.

Correlations and rotated factor loadings for the six spatial tests are shown in Table 26. The Visualization factor accounted for 10.6 percent of the total variance. It was defined by Spatial Visualization (.62), with Punched Holes and Form Board both loading .52. Spatial Orientation also loaded on the factor (.42) reflecting the correlation of .61 between Spatial Orientation and Spatial Visualization, which was the highest correlation in the matrix.

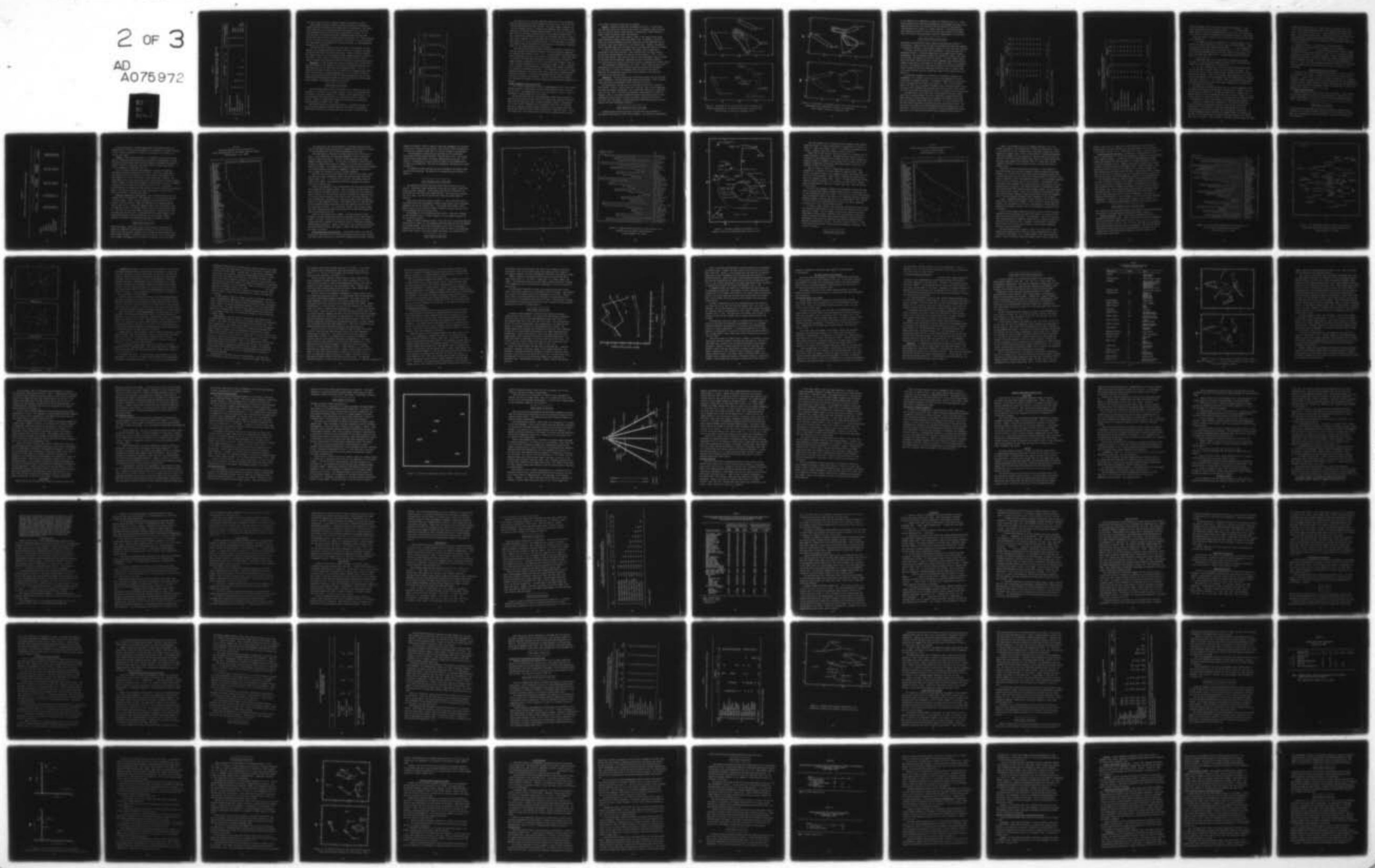
Insert Table 26 about here

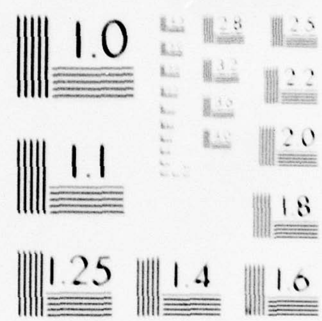
The factor labeled Spatial Relations was defined by Spatial Orientation (.58), and accounted for 9.1 percent of the total variance. Cubes and Flags loaded only .43 and .44, respectively. Spatial Visualization also loaded .44 on the factor.

Form Board and Punched Hoses also "defined" one of the unnamed factors (V2) with loadings of .42 and .36, respectively. The authors speculated that

AD-A075 972 STANFORD UNIV CALIF SCHOOL OF EDUCATION F/G 5/10
SPATIAL ABILITY: A REVIEW AND REANALYSIS OF THE CORRELATIONAL L--ETC(U)
OCT 79 D F LOHMAN N00014-75-C-0882
UNCLASSIFIED TR-8 NL

AD
A075972





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Table 26

Correlations and Factor Loadings of the Six Spatial Tests
(After Michael, Zimmerman & Guilford, 1950)

Test	Correlations						Rotated Factor Loadings		
	5	9	10	11	12	14	SR	Vz	Other Factors
5 Spatial Orientation							58	42	
9 Cubes	40						43	20	(24Ps)
10 Flags	37	36					44	15	(26V2)
11 Spatial Visualization	61	40	35				44	62	(25R)
12 Punched Holes	47	35	38	50			25	52	(36V2, 28R)
14 Form Board	43	37	39	50	54		22	52	(28Ps, 42V2)

Note. Decimals omitted.

this factor might reflect the response component of drawing the answer.

Even though the authors claimed the study supported the Visualization, Spatial Relations distinction, the evidence was not overwhelming. The average intercorrelation of the three Spatial Relations tests was only .38, while their average correlation with the three Visualization tests was .42. On the other hand, the Visualization tests correlated higher with each other ($\underline{r}$=.51) than they did with the Spatial Relations tests ($\underline{r}$=.42). This correlation pattern is similar to many others reported in this review. It can be captured in a hierarchical factor analysis or, even more directly, in a multidimensional scaling of the correlations.

The nine factors accounted for only about half of the variance in these fourteen tests, suggesting that there was a restriction of range in the sample, or that the tests were particularly unreliable. Further, the select samples and liberal criteria for factor extraction that characterize much of Guilford's work tend to yield a matrix of relatively small correlations and a large number of factors that capitalize on minute differences in correlation patterns.

Reanalysis. If the correlations for the six spatial tests in Table 26 are factored separately, the results do not support the hypothesis that the first three tests and the last three tests define separate factors. In fact, if there are two factors in this matrix, the distinction is between the four Thurstone tests and the two Guilford-Zimmerman tests. This is shown in Table 27. Here, two factors were extracted from the correlation matrix for the six tests using principal factoring with squared multiple correlations as initial communality estimates. Convergence required 25 iterations.

Insert Table 27 about here

In the unrotated matrix, Factor I accounted for 44.5 percent of the total variance or 91.3 percent of the common variance. Factor II only accounted for 4.2 percent of the total variance or 8.7 percent of the common variance. Thus, the correlations were reproduced with one factor about as well as with two factors. Further, the small second factor does not support the grouping of tests advocated by Michael et al. (1950).

When three factors were retained and rotated to a varimax criterion, the first factor was defined by Spatial Orientation and Spatial Visualization, the second by Form Board and Punched Holes, and the third by Flags and Cubes.

Table 27

Factor Matrices for the Reanalysis of Michael, Guilford, & Zimmerman (1950)

	Test	Unrotated		Rotated		
		I	II	I'	II'	h^2
5	Spatial Orientation	.76	-.35	.31	.77	.70
11	Spatial Visualization	.74	-.13	.46	.60	.56
9	Cubes	.54	.00	.40	.37	.29
10	Flags	.53	.10	.46	.29	.29
12	Punched Holes	.69	.16	.61	.35	.50
14	Form Board	.71	.28	.70	.28	.57
	Percent of Total Variance	44.6	4.2	53.0	25.8	

The culprit here is the high correlation between the Guilford-Zimmerman Spatial Orientation and Spatial Visualization tests (r=.61). and the low correlation between Flags and Cubes (r=.36). The former slightly elevated, but is not much higher than the .55 Guilford and Zimmerman (1948) reported in the manual for the Guilford-Zimmerman Aptitude Survey. In part, this correlation may reflect the difficult response format of the Spatial Orientation test. In fact, the test was sufficiently difficult that no subject was able to attempt every item. Similarly, only those items that 67 percent of the group attempted were scored on the Spatial Visualization test. On the other hand, the correlation between Flags and Cubes was much lower than the .68 reported by Thurstone (1938) for the college graduates in his PMA study.

The correlation between the Spatial Orientation and Spatial Visualization tests may indicate that Spatial Orientation, which was previously identified as a possible space subfactor (see p . 53), is a more complex aptitude construct than Spatial Relations, yet not as complex as Visualization. On the other hand, it could mean that individual differences in the psychological processes involved in mentally manipulating an object "out there" are the same as those involved in mentally moving the self to a different vantage. Finally, as Barratt (1953) suggests, it may indicate that many subjects solve Spatial Orientation test items by a Visualization strategy. It is impossible to know which of these possibilities obtain on the basis of these data. One thing is certain, however, indivdiual differences in the Guilford-Zimmerman Spatial Orientation test do not define a radically new dimension. They are to a large degree congruent with individual differences in the more familiar Group Spatial and Visualization factors.

Michael, Zimmerman, and Guilford (1951)

In a follow up investigation, Michael, Zimmerman, and Guilford (1951) administered a battery of seven spatial tests and eight reference tests to 151 boys and 139 girls. The students were all in the 12th grade at a junior college in California. The age range was 15 to 20 years.

The spatial tests were the Guilford-Zimmerman (1948) Spatial Orientation and Spatial Visualization tests; Thurstone's (1938) Cubes, Form Board, Punched Holes, and PMA Space tests; and the Spatial Relations subtest of Wrightstone and O'Toole's (1947) Prognostic Test of Mechanical Abilities. The PMA Space test was a composite of Cards, Flags, and Figures, and the Spatial Relations subtest was a multiple choice version of Thurstone's (1938) Form Board. It will be recalled that in the Thurstone Form Board test the examinee must draw lines in the figure to show how the pieces fit together.

Results. Separate analyses were performed for each sex. In both cases, nine centroid factors were extracted and rotated to orthogonal simple structure. Only six could be labeled in each analysis: Visualization, Spatial Relations, Number, Verbal, Perceptual Speed, and Reasoning.

Males performed better than females on most of the spatial tests, while females performed slightly better than males on the two numerical calculation tests, the Guilford-Zimmerman Perceptual Speed test, and PMA Reasoning, but did not exhibit their usual superiority on the verbal tests. However, the spatial tests correlated higher with the verbal tests for females than they did for males.

In spite of these mean differences, the authors found no important differences in the factor analyses. They concluded that "the factor pattern in each test was approximately the same for the two groups" (p. 576). As for the two space factors, which were the object of the investigation, they concluded "in the main, the two hypotheses regarding the nature of Spatial Relations and Visualization were upheld as they were in Michael, Zimmerman and Guilford (1950)."

The conclusions are remarkable on two counts. First, even the most cursory examination of the correlation and factor structure matrices reveals marked sex differences. Second, the hypotheses about the nature of the Spatial Relations and Visualization factors were no more upheld in this study than they were in the previous study.

Reanalysis. The seven spatial and two percpetual speed tests were included in the reanalyses. The perceptual speed tests were Thurstone's (1938) Identical Forms and the Guilford-Zimmerman (1948) Perceptual Speed test.

Multidimensional scalings were first performed on each matrix using the KYST program. Formula 1 stress values were .0396 and .0384 in three dimensions for males and females, respectively; the corresponding values for two dimensions were .1174 and .1206. The final two dimensional representations are shown in Figures 11 and 12. The major differences between these solutions and the three dimensional solutions were, for males, a stronger clustering of PMA Space and Cubes, and, for females, a larger separation between PMA Space and Identical Forms in the three dimensional solutions.

Insert Figures 11 and 12 about here

Diameter method hierarchical cluster analyses were also performed on each matrix using Johnson's (1967) HICLUS program. The results were superimposed

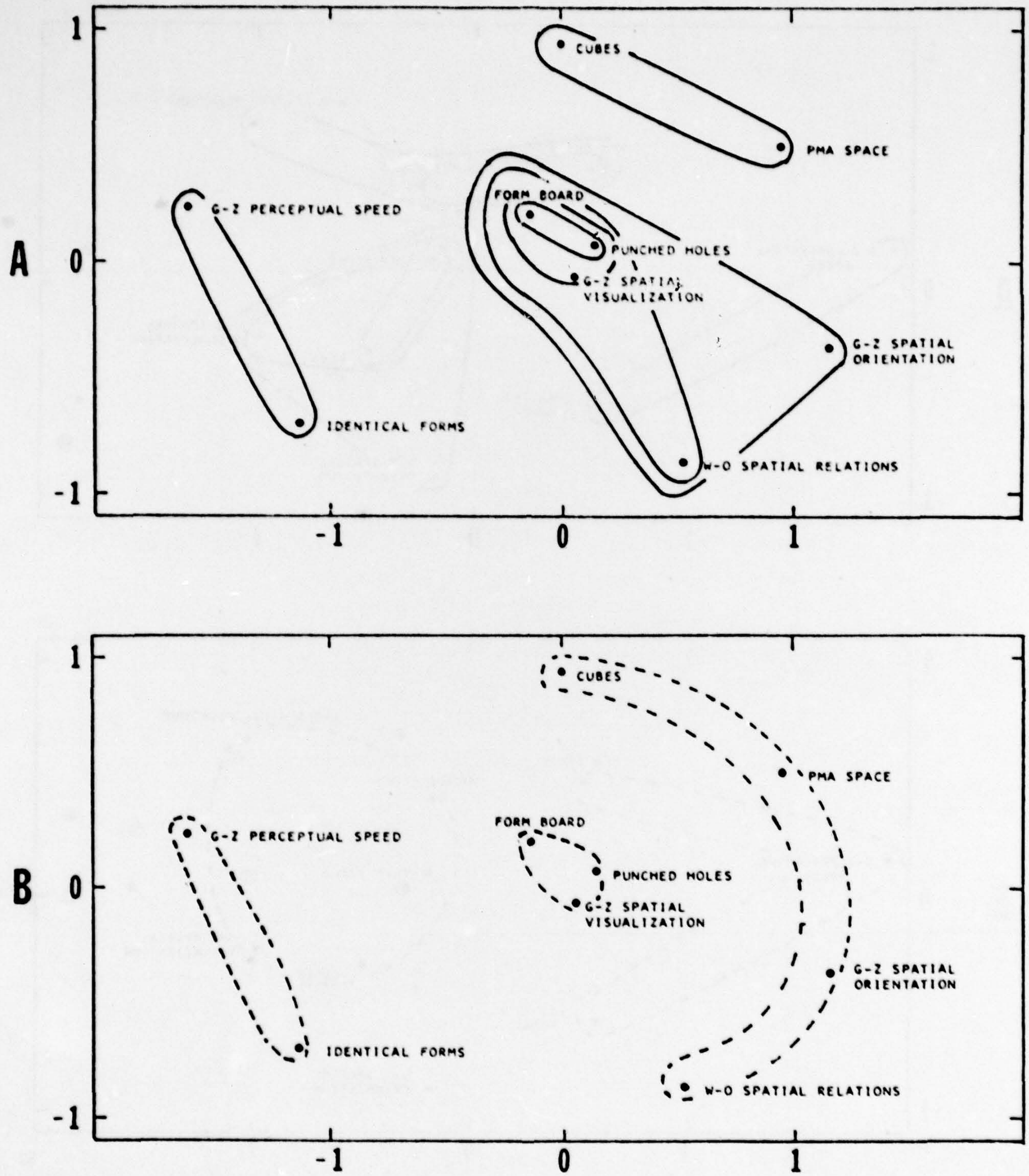


Figure 11. Clustering by (a) diameter method, and (b) rotated factor loadings superimposed on two dimensional scalings of the male data (After Michael, Guilford, and Zimmerman, 1951).

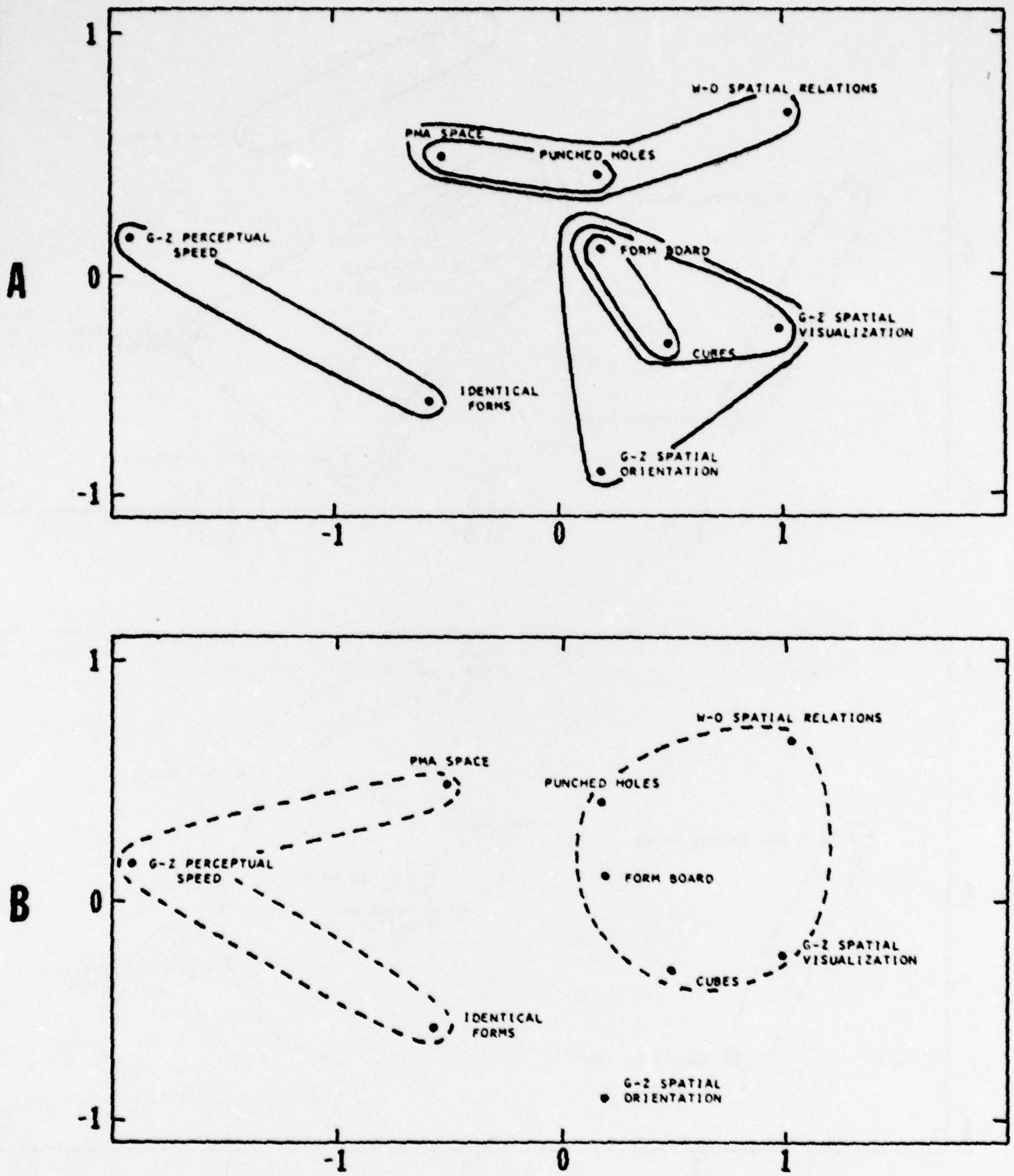


Figure 12. Clustering by (a) diameter method, and (b) rotated factor loadings superimposed on two dimensional scalings of the female data (After Michael, Guilford, & Zimmerman, 1951).

on the respective two dimensional scalings (see Figures 11a and 12a). Since factor analysis often produces more meaningful test groupings than does cluster analysis, three factors were extracted from each matrix by the principal factoring method and rotated to a varimax criterion. Unrotated and rotated factor matrices are shown in Table 28 for males and in Table 29 for females. Three factors were extracted because three were hypothesized (Vz, SR and Ps),

---

Insert Tables 28 and 29 about here

---

and because the hierarchical cluster analyses indicated three clusters in both matrices. Figures 11b and 12b show the groupings when tests were assigned to clusters on the basis of their maximum loading in the rotated factor matrices.

For males, the results were clear and familiar. Figure 11a shows the two dimensional nonmetric scaling representation with diameter method clusters superimposed. The two perceptual speed tests defined one cluster, while Cubes and PMA Space came together to form an SR cluster. The three complex spatial tests (Form Board, Punched Holes and Guilford-Zimmerman Spatial Visualization) formed a strong cluster at the center of the figure that represents the familiar Vz factor. The Wrightstone-O'Toole Spatial Relations and Guilford-Zimmerman Spatial Orientation tests were eventually pulled into this Vz cluster, while in the factor analysis they were pulled more toward the PMA space factor (see Figure 11).

The shifting allegiance of these tests of intermediate complexity (or speededness) is of no great concern, as the line that separates one cluster from another is somewhat arbitrary. The important dimensions are shown clearly in both the unrotated and rotated factor matrices. Thus, in the unrotated matrix, the first factor represents the general plus broad group spatial factor. The second factor separated the perceptual speed tests from the others, while the third set Punched Holes against PMA space in the familiar SR-Vz bipolar factor.

The female data, on the other hand, yielded markedly different results. There were no strong clusters, and so the hierarchical cluster analysis and factor analysis produced disparate test groupings (see Figures 12a and 12b). The factor analysis offered the clearest solution. In the unrotated matrix, the first factor represents the general plus broad group spatial factor. The second factor separated the Guilford-Zimmerman Spatial Orientation test from the others. The third factor was defined by the Guilford-Zimmerman Perceptual Speed test, with Identical Forms and PMA Space loading positively and the

Table 28

Factor Matrices for Males for Reanalysis of Michael, Zimmerman & Guilford (1951)

Test	Unrotated			Rotated			
	I	II	III	Vz	SR	Ps	h^2
Form Board	74	05	15	57	34	37	58
Punched Holes	78	-19	42	85	28	14	82
G-Z Spatial Visualization	74	-05	06	53	45	29	56
W-O Spatial Relations	54	-16	-07	33	45	10	32
G-Z Spatial Orientation	53	-15	-24	20	55	12	36
Cubes	58	01	-11	28	44	28	35
PMA Space	60	-19	-34	18	68	13	52
Identical Forms	48	34	-10	15	24	52	36
G-Z Perceptual Speed	45	59	-01	13	06	73	55
Total Variance	3.42	.59	.41	1.63	1.62	1.16	

Note. Decimals omitted in factor loadings.

Table 29

Factor Matrices for Females for Reanalysis of
Michael, Zimmerman & Guilford (1951)

Test	Unrotated			Rotated			h^2
	I	II	III	S	SO	Ps	
Form Board	81	-17	-04	70	24	37	69
Punched Holes	66	-23	-03	61	11	33	49
G-Z Spatial Visualization	71	-02	-33	73	30	03	62
W-O Spatial Relations	53	-19	-19	57	07	13	35
G-Z Spatial Orientation	70	62	13	24	89	20	89
Cubes	73	12	-15	59	44	14	57
PMA Space	58	-05	20	36	26	44	39
Identical Forms	53	-05	23	30	23	43	33
G-2 Perceptual Speed	32	-19	51	05	04	63	40
Total Variance	3.62	.56	.54	2.35	1.28	1.09	

Note. Decimals omitted in factor loadings

Guilford-Zimmerman Spatial Visualization test loading negatively. This factor may represent either Perceptual Speed or a biploar speed-power dimension. The corresponding labels for the rotated solution would be Space, Spatial Orientation (singleton?), and Perceptual Speed or simply Speed.

There are several hypotheses that may account for this lack of structure in the female data.

1. The space tests were too difficult for the female students, and so scores were largely determined by factors other than spatial ability. There was some support for this hypothesis in the data. After corrections for guessing, the average percent correct was only 29.7, 25.1, 26.0 and 16.2 on the Guilford-Zimmerman Spatial Orientation, Spatial Visualization, PMA Space, and Cubes test, respectively. Corresponding values for males were: 40.9, 41.0, 33.5 and 28.0.

2. Females may tend to solve these tests by nonspatial techniques. The higher correlation between spatial and verbal tests in the female data suggest that these methods may be verbal-analytic.

3. Not only do females tend to solve the tests differently than males, but they may tend to be more eclectic in their solution strategies. This could result from not having clearly defined, systematic methods for solving spatial test problems. There is some evidence that students who do not have well defined methods for solving problems show less differentiation of abilities (see French, 1965, and p . 140 below). This is consistent with the weaker clustering of spatial tests and their higher correlations with verbal tests for the females in this study.

The difficulty hypothesis may explain why the Guilford-Zimmerman Spatial Orientation test split away and defined a separate factor. However, it could be that Spatial Orientation is the important subfactor for females, while Spatial Relations is the corresponding subfactor for males. However, the data do not indicate a bipolar SO-Vz factor in the female data that would correspond to the bipolar SR-Vz factor that appeared in the male data and elsewhere.

Therefore, it appears that a combination of the difficulty and lack of method hypotheses best explain these data. There may be just one loose space factor in the female data. The slight link between the Spatial Orientation test and Cubes is probably not psychologically significant. The Spatial Orientation test is difficult, the response format is confusing, and the test is quite susceptible to alternate solution strategies (see, Barratt, 1953, and p. 136 below). Thus, the factor it defined here was probably just noise.

Finally, there does appear to be a speed factor in the female data. The factor is not simply perceptual speed, since PMA space loads significantly on it. Thus, the study identified a clear, familiar factor pattern for males: a broad group spatial factor, a bipolar SR-Vz factor and a Ps factor. The female data, on the other hand, were more ambiguous. There, only a loose group spatial factor and an unfamiliar Speed or Perceptual Speed factor could be identified.

Finally, it would be of some interest to determine whether sex differences in spatial ability are greater for Vz or SR tests. A reasonable hypothesis is that the difference would be larger on speeded SR tests than on relatively unspeeded Vz tests, since the latter may be more susceptible to alternative solution strategies. However, the male advantage was about equally great for Spatial Orientation, Spatial Visualization, PMA Space, Form Board, and Cubes. Vz, SO, and SR tests are all represented in this list. Even if the male advantage were larger on the SR tests, the results would be ambiguous since the factor structures were so different for the two groups.

Spatial Abilities in the SI Model

The final Guilford study reviewed here derives from the faceted Structure of the Intellect (SI) model. The model posits a three way classification of abilities: content (figural, symbolic, semantic and behavioral); by operation (cognition, memory, divergent production, convergent production, and evaluation); by product (units, classes, relations, systems, transformations, and implications). The full model predicts 120 independent abilities, each defined by a particular combination of operation, content, and product.

The Figural Cognition Battery

Spatial abilities fall in the figural slice of the model. Table 30 shows the 6x5 figural matrix and particular cell abbreviations. Eighteen of the 30 cells were represented in a study by Hoffman, Guilford, Hoepfner and Doherty (1968). These cells are underlined in Table 30. The cognition and evaluation

---

Insert Table 30 about here

---

columns were fully represented, while only four divergent production cells, one memory cell, and one convergent production cell were included in the study. Tests for five reference cells from the semantic slice of the SI model were also represented.

Table 30

Figural Slice of the Structure of the Intellect Model

	Operation				
	Cognition	Memory	Divergent Production	Convergent Production	Evaluation
Product					
Units	CFU	MFU	DFU	NFU	EFU
Classes	CFC	MFC	DFC	NFC	EFC
Relations	CFR	MFR	DFR	NFR	EFR
Systems	CFS	MFS	DFS	NFS	EFS
Transformations	CFT	MFT	DFT	NFT	EFT
Implications	CFI	MFI	DFI	NFI	EFI

Note. Underscored cells represented in Hoffman et al. (1968).

The total battery of 72 tests representing 23 hypothesized factors was administered to 250 architecture students at the University of Illinois, Chicago Circle. Sex was included as a variable in the analysis even though only 13 students were women.

Results. The correlation matrix for 74 variables (72 tests plus sex and year in college) was then factored by the principal factoring method. Squared multiple correlations were used as initial communality estimates and extraction of the 25 factors (23 abilities plus sex and year in college) was iterated until no communality changed more than .02.

The 25 principal axes were then orthogonally rotated by an analytic, procrustean procedure developed by Cliff (1966). The initial target matrix was formed by inserting a loading for each test equal to the square root of its communality on the one factor that it was hypothesized to measure. New target matrices were constructed after each of seven iterations of this procedure. Finally, graphic rotations were performed on "selected pairs of factors, primarily to improve positive manifold and simple structure" (p. 22). Twenty-two hypothesized and one unexpected SI factors were thus extracted from the correlation matrix. The authors interpreted the final factor matrix in terms of the SI model, and claimed it supported that theory.

Principal components of the 72 tests. Other interpretations are not only possible, but more parsimonious. The reanalysis of this battery was conducted in several stages. First, principal components were extracted from the 72 variable correlation matrix. Twenty-one components had eigenvalues greater than or equal to one. However the computer program could rotate only 20, and so these were rotated to a varimax criterion. The results are shown in Table 31. The first factor was the largest, and represents a combination of Gf and the broad group spatial factor. It was defined by the complex spatial tests: Spatial Visualization, Block Visualization, and Paper Folding. Figure Analogies and Figure Series, which are both based on Spearman "g" tests,

---

Insert Table 31 about here

---

also loaded highly. The SR and SO tests (such as Spatial Orientation and Planning Air Maneuvers) had intermediate loadings on the factor, while the simple tests (such as Least Movement, Line Continuations, and Identical Forms) had the lowest loadings. The Hidden Figures test was too easy for this highly select sample, and emerged with a relatively low loading (.29) on the factor.

Table 31

Principal Component Factor Structure Matrix

after Varimax Rotation for the Figural Cognition Battery

(After Hoffman et al., 1968)

		Factor																				
	Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	h^2
69	Spatial Visualization	70																				70
10	Block Visualization	76																				60
31	Paper Folding	70																				74
66	Space Positioning	70													60							77
11	Blocks	60								38												75
42	Problem Solving	63			39																	66
9	Block Rotation	63																				73
21	Figure Analogies	60																				63
19	Sequential Maze Routes	60																				65
27	Figure Series	59															36					60
43	Match Problems II	59																			-30	72
64	Similar Orientations	57					32															60
49	Unnecessary Facts	57										28										66
32	Pattern Arrangement	57												-35								67
36	Judging Figural Balance	54																				60
46	Most Effective Path	54								33												64
45	Monograms		73																			69
43	Sketches		72																			67
18	Designs		69																			71
58	Plot Titles (fluency)		60																			60
42	Make a Figure Test		63																			60
56	Planning Elaboration II		59		35																	58
1	Alternate Signs		59																			60
33	Figure Completion			71																		72
13	Closures			66																		68
29	Hidden Print	38		65																		77
67	Mutilated Words			59					60													66
72	Word Completion				75																	73
71	Verbal Comprehension				75																	74
70	Speed-Shape Recognition					77																70
50	Orientation Memory					75																70
44	Monogram Recall					54																69
22	Figure Classification						64															66
14	Closest Spatial Series						61															55
24	Figure Matching							70														66
12	Circle Continuations								73													68
28	Identical Forms									63												66
2	Angle Estimation										76											67
67	Spatial Comprehension	32										67										69
30	Judging Figural Combinations												67									69
33	Internally Consistent Figures													72								73
63	Remembering Object Orientation					31								55								71
40	Least Movement														67							61
39	Judgment of Size															82						73
54	Perceptual Relation Judgment																77					72
5	Best Figural Separation																	80				65
20	Hidden Figures																		66			68
3	Artistic Interpretations																			68		64
4	Best Figural Class	66																			50	67
31	Identical Figural Relations	60							-38													64
48	Shortest Figural Relations																			-49		58
15	Competitive Planning	31						33											-40		30	69
53	Penetration of Camouflage			32				50					63									70
17	Decorations		45																40			60
55	Planning Air Maneuvers	62															31					70
41	Line Continuations	35							49													62
57	Plot Titles (clever)				46						39											68
20	Figural Class Inclusion						39			38												55
59	Possible Jobs		50		38																	62
60	Poster Judgment	41																				57
61	Prescribed Relations	35																				59
6	Best Figure Pairs	47																				55
8	Best Move Selection	30										46										62
36	Judging Figural Elaboration	65						40														63
7	Best Map Placement	34														36						55
25	Figure Matrix	38					32											35				60
37	Judging Rearrangements	49																				58
68	Spatial Orientation	43									31				32							64
16	Correct Figural Trends	45															31					72
38	Ideational Fluency		66		61																	63
26	Figure Production		50	30																		70
58	Judging Specified Figures	47							31		32											78
	% of Total Variance	14.7	6.2	4.1	4.0	3.0	2.8	2.7	2.6	2.5	2.4	2.4	2.4	2.3	2.2	2.2	2.1	2.1	2.1	2.1	1.9	66.8

Note. Decimals omitted.

The second factor was composed entirely of divergent production tests. Most of the semantic tests had additional loadings on the Verbal factor (factor 4), which was defined by the two verbal comprehension (CMU) tests.

The third factor was Thurstone's Closure Speed factor or DFU in the SI model. It was defined by the test Figure Completion, which is Guilford's version of the Street Gestalt. The test called Closeups had the second highest loading on this factor. In this test, the student is shown close up pictures of common objects (such as a keyhole or a button) and is required to identify the object. This suggests that Closure Speed may involve the recognition of a visual stimulus on the basis of fragmentary or distorted information, and not simply the "closing" of a set of discrete elements.

The fifth factor was a Visual Memory factor (MFS in the SI model), while the sixth was a doublet composed of Figure Classification (EFS) and Closest Spatial Series (CFC). It is difficult to see what, if anything, these two tests have in common. The remaining factors were singletons and doublets. Only two are discussed here.

Factor 8 was defined by Circle Continuations (CFI) with Line Continuations (CFI) having the next highest loading. This is probably a method factor, as the two tests are extremely similar. In Circle Continuations, the student is shown a portion of a circle and then required to determine by inspection which of five dots would be exactly on the circle if the circle were completed. In Line Continuations, a gap appears in a line that passes through two parallel lines, as in the Poggendorf illusion. The student's task is to indicate which of four alternative lines on one side of the gap complete the given line on the other side of the gap. It is noteworthy, however, that the more complex test (Line Continuations) loaded on the spatial factor while the simpler test (Circle Continuations) did not.

Factor 14 also deserves brief comment. The factor defined by Least Movement, with Space Positioning and Spatial Orientation also loading significantly. All three of these tests seem to involve the movement of a spatial configuration with reference to the observer's body, or the factor previously called Spatial Orientation. However, other tests (such as Similar Orientations and Closest Spatial Series) that appear to require this same perspective did not load on the factor.

First scaling and cluster analyses. In the second stage of the analysis, 44 tests best representing 22 SI factors identified in the study were selected. Tests were chosen on the basis of Guilford's recommendations (Hoffman, Guilford, Hoepfner and Doherty, 1968) or, where no tests were recommended, on the basis of the factor solution for this battery. One of the SI factors identified in the study was a singleton (DMT) and so it was not included in this analysis.

A multidimensional scaling was then performed on this correlation matrix using the KYST program (Kruskal, Young and Seery, 1973). The initial configuration was generated by the metric Young-Torgerson procedure. The nonmetric configuration was then iterated 25 times in three dimensions and 16 times in two dimensions. The final two dimensional representation is shown in Figure 13.

A hierarchical cluster analysis was also performed on this matrix using Johnson's HICLUS program. The results of the diameter method are shown in Figure 14.

Insert Figures 13 and 14 about here

The divergent production tests split away from the other tests in the battery. These tests formed subclusters on the basis of content, rather than along the product dimension. DMI and DMU clustered first, followed by the second DMI test, and finally the second DMU test. The clustering was similar for the DF tests. Two DFI and one DFU test formed the bottom cluster in Figure 14.

Other factors are also evident in Figure 14. The two CFI tests (Circle Continuations and Line Continuations) formed a tight cluster, but as previously suggested, this may be a method factor. The two MFS tests formed a Visual Memory cluster; the two CMU tests formed a Vocabulary cluster; and the two CFU tests represented the Closure Speed factor.

The remainder of the clustering was less obvious. There was a tendency for the evaluation tests to cluster together, but the remaining clusters did not follow the SI facets.

Figure 15 shows an enlarged version of the right side of the multidimensional scaling shown in Figure 13. The divergent production tests and the test Judgment of Size were omitted. Tests were then grouped on the basis of a principal components factor analysis of the 44 tests in which 12 factors with eigenvalues greater than or equal to one were retained and rotated to a varimax criterion.

Insert Figure 15 about here

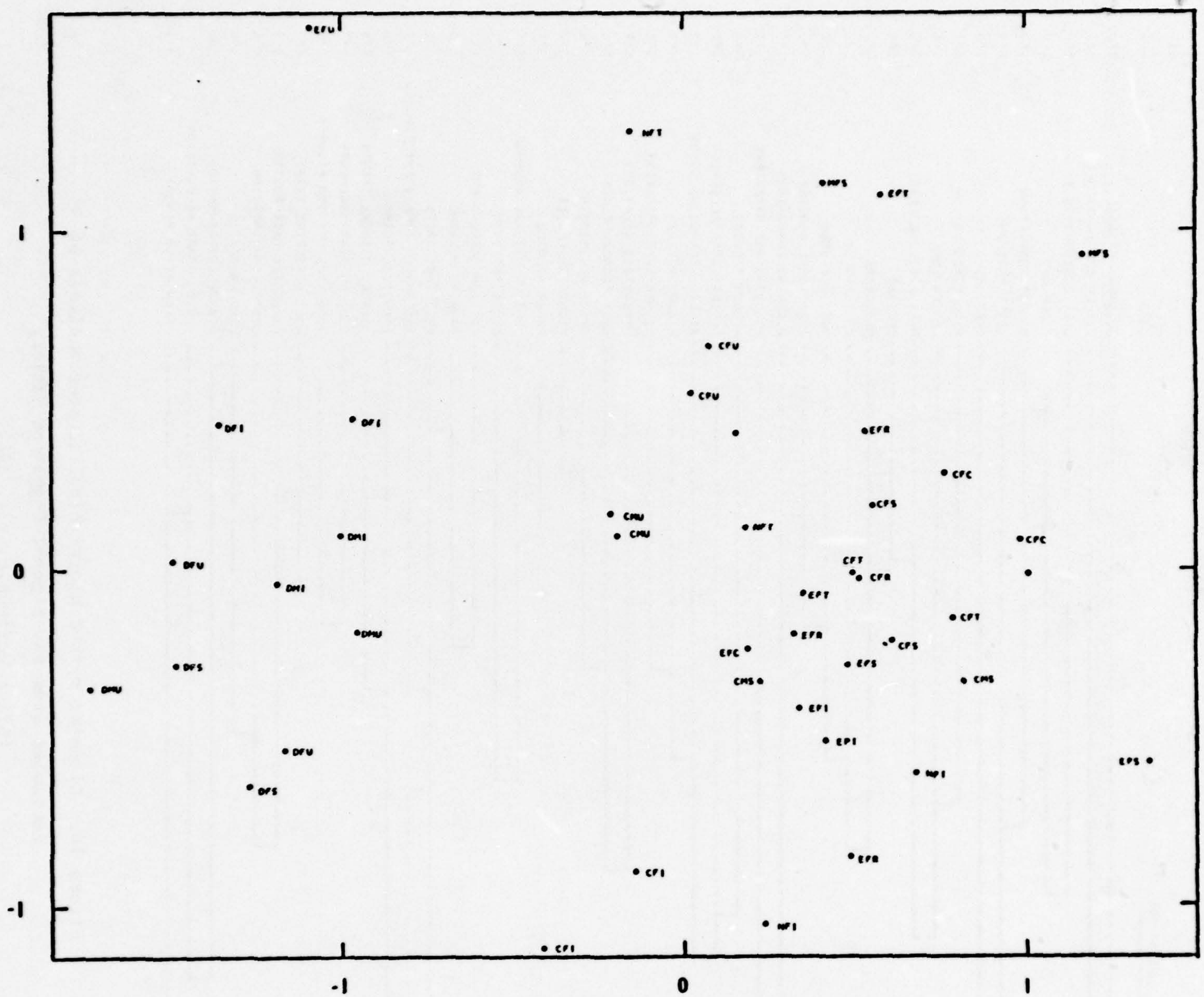


Figure 13. Two dimensional scaling of the 44 variable matrix (After Hoffman et al., 1968).

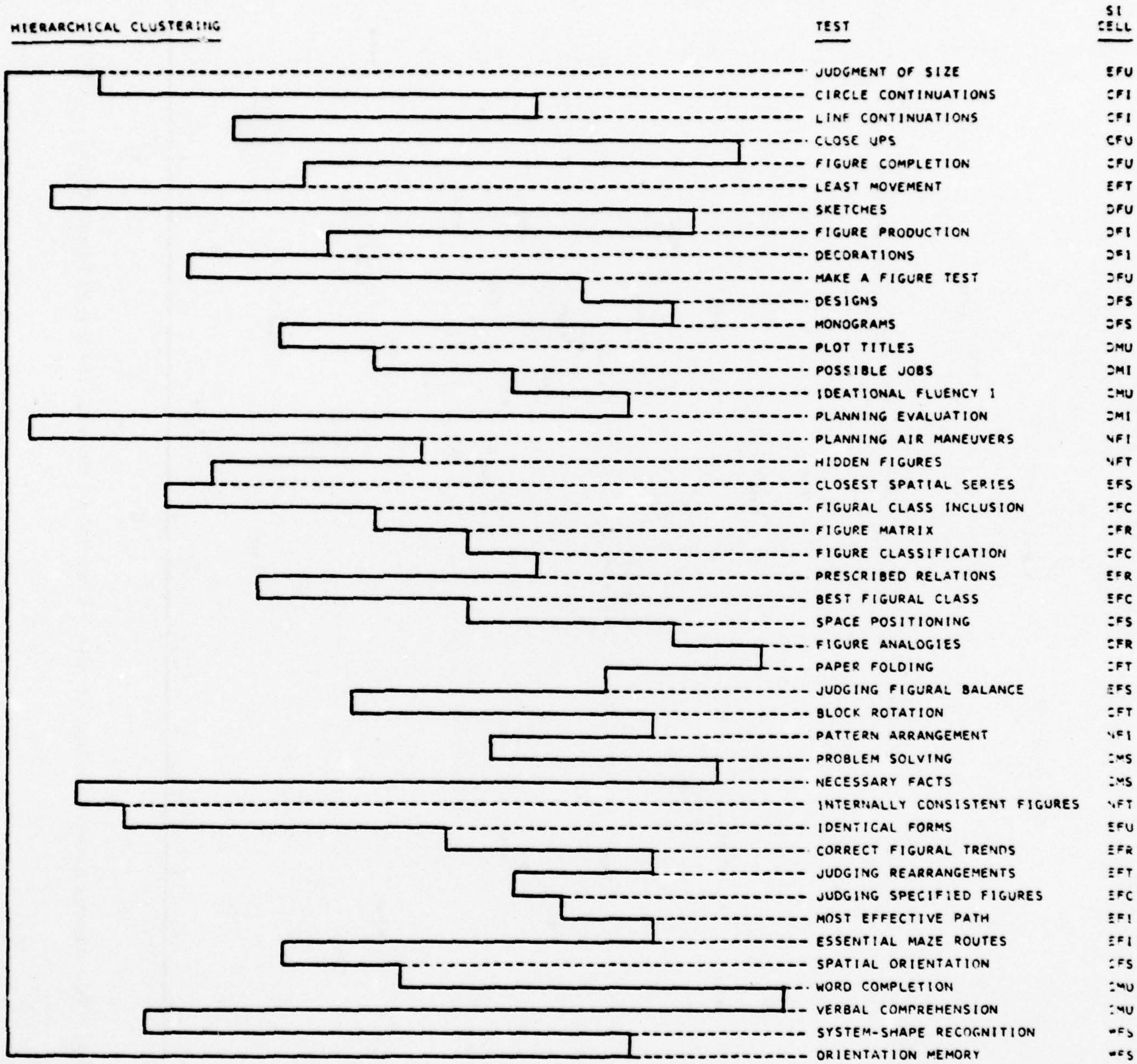


Figure 14. Diameter method hierarchical cluster analysis of 44 variables from the Figural Cognition Battery (After Hoffman et al., 1968).

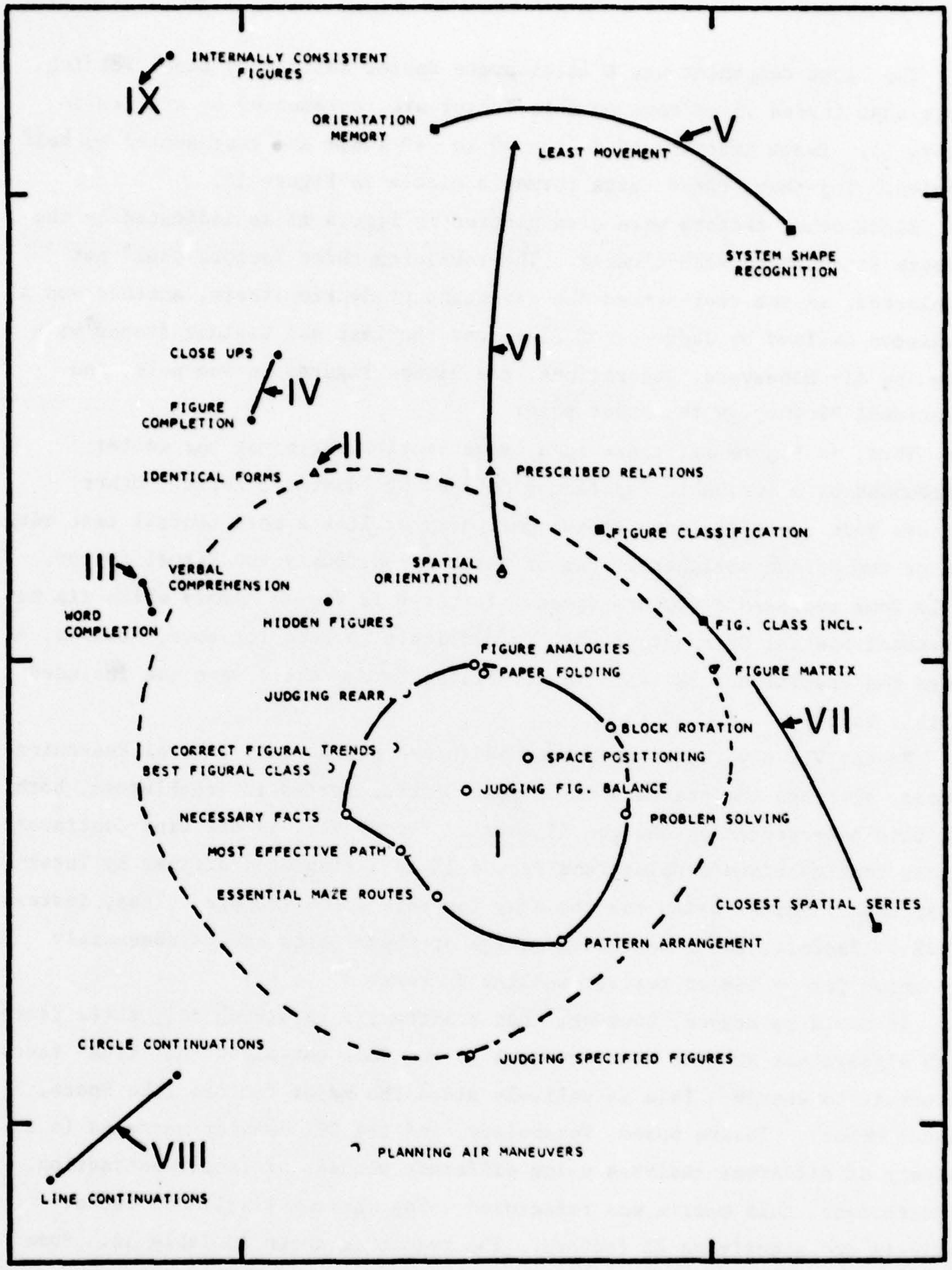


Figure 15. Principal components superimposed on the scaling of Figure 13, excluding divergent production tests.

The first component was a large space factor defined by Paper Folding. Tests that loaded .5 or more on this factor are represented by circles in Figure 15. Tests that loaded in the.30 to .49 range are represented by half circles. Together, these tests formed a circle in Figure 15.

Eight other factors were also plotted in Figure 15 as indicated by the numbers attached to each cluster. The remaining three factors could not be plotted, as one represented the divergent production tests, another was a singleton defined by Judgment of Size, and the last was biploar factor with Planning Air Maneuvers, Decorations, and Hidden Figures on one pole, and Ideational Fluency on the other pole.

Thus, in Figure 15, there is a large Spatial factor at the center surrounded by a (probably Ps) factor defined by Identical Forms. Other factors join specific tests at the periphery or link a more central test with one or two on the periphery. Factor three is obviously the Verbal factor, while four represents Closure Speed. Factor V is Visual Memory while six may represent Spatial Orientation. It is difficult to know for sure, however, since the traditional SR tests (e.g., Flags, Cards, etc.) were not included in this battery.

Factor VII may represent a speeded (hence peripheral) Figural Reasoning factor, although the presence of Closest Spatial Series is troublesome, both for this interpretation and the SI model. Factor VIII is the Line Continuation-Circle Continuations doublet, and Factor IX is a singleton defined by Internally Consistent Figures, which was too easy for this select sample. Thus, instead of 22 SI factors, the interrelationships of these tests can be adequately accounted for by ten or twelve familiar factors.

It could be argued, however, that arbitrarily retaining only those factors with eigenvalues greater than or equal to one does not allow the "true" factor structure to emerge. This is unlikely since the major factors like Space, Visual Memory, Closure Speed, Vocabulary, and the DFI doublet surfaced in a variety of different analyses using different methods of factor extraction. Nevertheless, this matrix was refactored using maximum likelihood factor analysis and specifying 22 factors. The result is shown in Table 32. Some of the factors were familiar, however most were singletons or doublets. Only a few merit comment.

Insert Table 32 about here

Table 32

Varimax Rotated 22 Factor Maximum Likelihood Solution

for the 44 Variable Matrix

(After Hoffman et al., 1968)

		Factor																						
Test	Cell	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	h^2
18 Designs	DFS	74																						63
65 Monogram	DFS	68																						66
62 Make a Figure Test	DFU	67																						57
71 Verbal Comprehension	CMU		75																					68
72 Word Completion	CMU		71																					62
37 Judging Rearrang.	EFT			61																				65
66 Space Positioning	CFS				85																			99
50 Orientation Memory	MFS					68																		56
70 System-Shape Recog.	MFS					64																		46
25 Figure Matrix	CFR						69																	62
12 Circle Continuations	CFI							72																62
41 Line Continuations	CFI							59																51
9 Block Rotation	CFT								90															99
13 Close Ups	CFU									94														99
55 Planning Air Maneuv.	NFI										72													70
32 Identical Forms	EFU											67												64
26 Figure Production	DFI												82											82
33 Intern. Consis. Fig.	NFT													97										99
4 Best Figural Class	EFC														90									99
30 Ideational Fluency	DMU	25	26													86								94
22 Figure Class.	CFC																90							99
62 Problem Solving	CMS		39	26														77						99
34 Judg. Fig. Balance	EFS		27					27											60					69
58 Plot Titles (fluency)	DMU	50																		69				81
39 Judgment of Size	EFU																				64			43
16 Correct Fig. Trends	EFR		35	27							26											56		78
46 Most Effective Path	EFI			26								25											56	70
61 Prescribed Relations	EFR		27		27		32																	46
65 Sketches	DFU	45											42							30				60
14 Closest Spatial Series	EFS			47																				35
60 Least Movement	EFT				34				30									-25						42
51 Paper Folding	CFT					33	42												34					75
38 Judg. Specif. Fig.	EFC							32				36									-29			57
19 Essent. Maze Routes	EFI						29																	48
23 Figure Completion	CFU									45	26	31												64
52 Pattern Arrangement	NFT			25					33		29													48
59 Possible Jobs	DMI	37	32																					49
56 Planning Elab. II	DMI	42	27													32								67
21 Figure Analogies	CFR			40		36													30					72
20 Fig. Class Inclusion	CFC			30																				40
28 Hidden Figures	NFT										28													41
49 Necessary Facts	CMS		32	37			32																	61
17 Decorations	DFI	40	32															-26						45
68 Spatial Orientation	CFS		28		45						27									-30				59
% of Total Variance		6.2	5.7	4.1	3.7	3.5	3.4	3.4	3.3	3.2	2.9	2.8	2.7	2.6	2.5	2.5	2.5	2.4	2.1	2.0	1.7	1.3	1.2	63.8

Note. Decimals omitted.

Factor I was composed entirely of divergent production tests, both figural and semantic, and covering units, systems, and implications. Factor II was defined by the two vocabulary tests, with minor loadings from several semantic and figural tests. Factor III was defined by Judging Rearrangements (EFT) which resembles the easy versions of form board used by Holzinger and Swineford (1939) with Closest Spatial Series (EFS) also loading substantially. The latter test presents four different views of a visual array, and the student is required to select the end view that is further away from its adjacent view.

The fourth factor was particularly interesting. It was defined by Space Positioning, with secondary loadings by Spatial Orientation and Least Movement. All of these tests, particularly Space Positioning, may be solved by projecting oneself into the picture and "walking around" the stimulus. This factor is similar to Factor XIV (Spatial Orientation) in the principal components solution of the correlation matrix for all 72 tests. The remainder of the large space factor obtained in the previous analysis (see Figure 15) was split between Factors VI and VIII. The former was defined by Figure Matrix (CFR), with Paper Folding (CFT) having the second highest loading. This may represent the Gf end of the Vz factor. Factor VIII is defined by Block Rotation (CFT) with small secondary loadings by Least Movement (EFT) and Pattern Arrangement (NFI). This may represent a mental rotation component or the SR factor.

Factor V represents Visual Memory and was defined by the two MFS tests, while Factor VII was the CFI doublet. Factor IX was the Closure Speed (CFU) doublet, this time defined by Close Ups. The remaining 13 factors were all singletons or doublets. However, none of the doublets were consistent with the SI model.

Thus, the only change between this solution and the 12 factor principal components solution discussed previously (see Figure 13) is that the large space factor was split into three or four subfactors. Only two of these subfactors were particularly suggestive, namely Factor IV (Spatial Orientation) and Factor VIII (Rotation or SR). However, a similar Spatial Orientation factor (Factor XIV) was previously obtained in the principal components solution of the entire battery.

The most important point, however, is that the 44 tests used in this analysis were selected because they were the best representatives of the 22 SI factors in the Hoffman et al. (1968) analysis of the same correlation matrix. Thus, this analysis should be strongly biased to obtain these same factors. That they do not emerge here except when they coincide with well known primaries from other systems is a strong challenge to the SI model and to the claims of Hoffman et al. (1968) that their analysis supports SI predictions.

Second scaling and clustering. The decision to reduce the battery of 72 tests to a smaller matrix of 44 tests was primarily a concession to the limitations of the computer programs. In particular, the KYST multidimensional scaling program can represent only 1800 data points, or a lower half matrix of from a 60 variable symmetric matrix. However, the process of deleting tests omitted a number of interesting and important spatial tests. Since this aspect of the analysis was the main concern, a new submatrix of 60 tests was formed, this time eliminating the divergent production tests. These tests defined a separate factor in all previous reanalyses, and, as can be seen in Figure 13, split away from the other tests in the multidimensional scaling of the 44 variable matrix.

The correlation matrix for these 60 tests was then disattenuated using Guilford's reliability estimates or the maximum correlation with any other variable in the battery, whichever was larger. Multidimensional scalings were then performed on this 60 variable disattenuated correlation matrix using the KYST program. The initial configurations were generated by the metric Young-Torgensen procedure. The non-metric configurations were then iterated 22 times in three dimensions and 20 times in two dimensions. The final two dimensional configuration is shown in Figure 17, and the final three dimensional configuration in Figure 18. Stress values (formula 1) were .296 and .224, respectively. Minimum and maximum method hierarchical cluster analyses were

---

Insert Figures 16, 17, and 18 about here

---

also performed on the disattenuated matrix using Johnson's (1967) HICLUS program. The results of the maximum method clustering are shown in Figure 16.

CFI and Cs clusters. Most of the clusters correspond with factors identified in previous analyses. Thus, the two CFI tests (Circle Continuations and Line Continuations) again formed a strong cluster. The four CFU (Closure Speed) tests formed the second cluster in Figure 16.The next two clusters were small and were formed relatively late in the analysis. The first was defined by Artistic Interpretations, which was hypothesized to be an EFT test but emerged with no significant factor loadings in Guilford's analysis, and Closest Spatial Series (EFS).

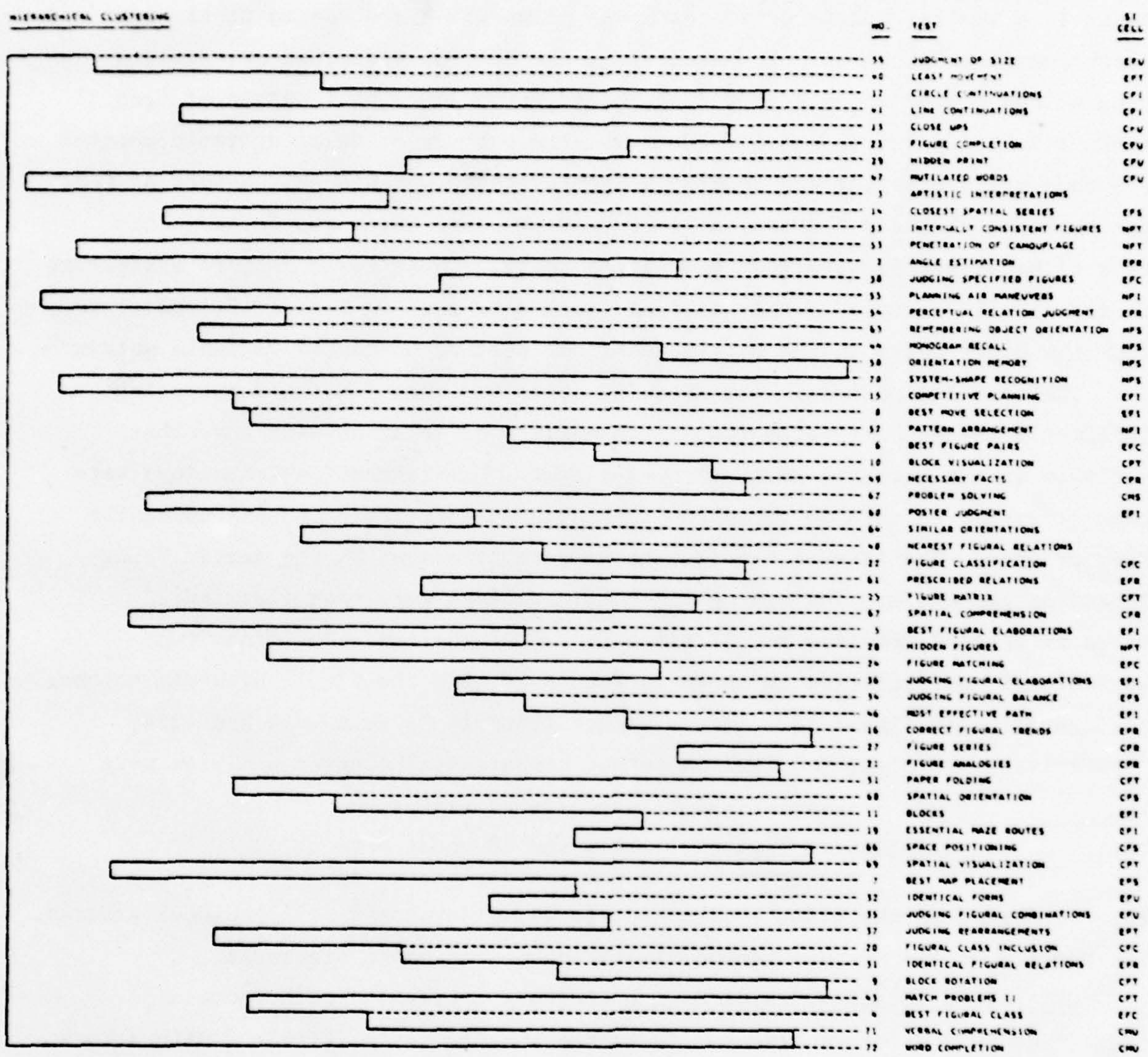


Figure 16. Diameter method hierarchical cluster analysis of 60 variables from the Figural Cognition Battery (After Hoffman et al., 1968).

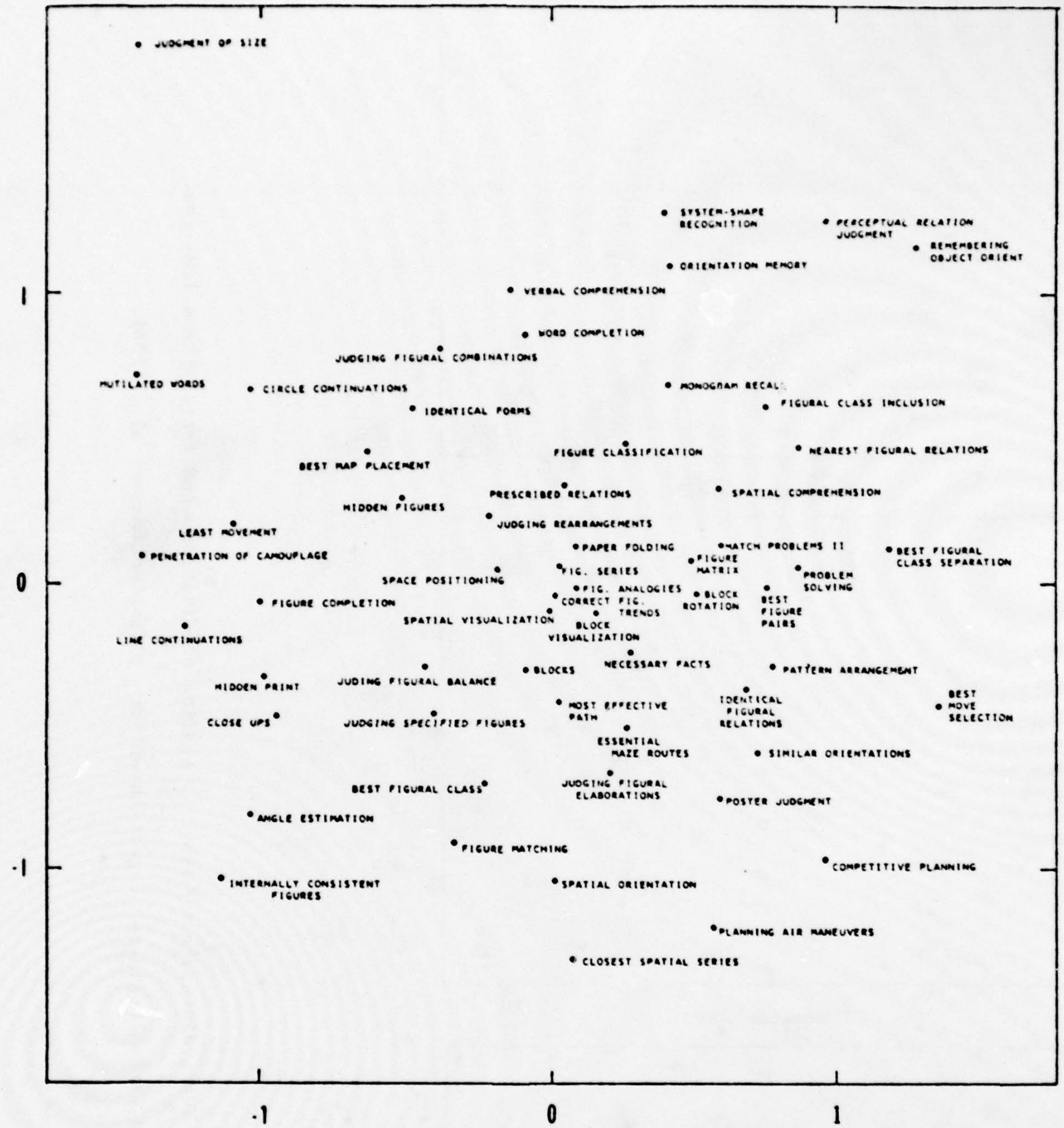


Figure 17. Two dimensional nonmetric scaling of disattenuated correlations from the 60 variable Figural Cognition Battery (After Hoffman et al., 1968).

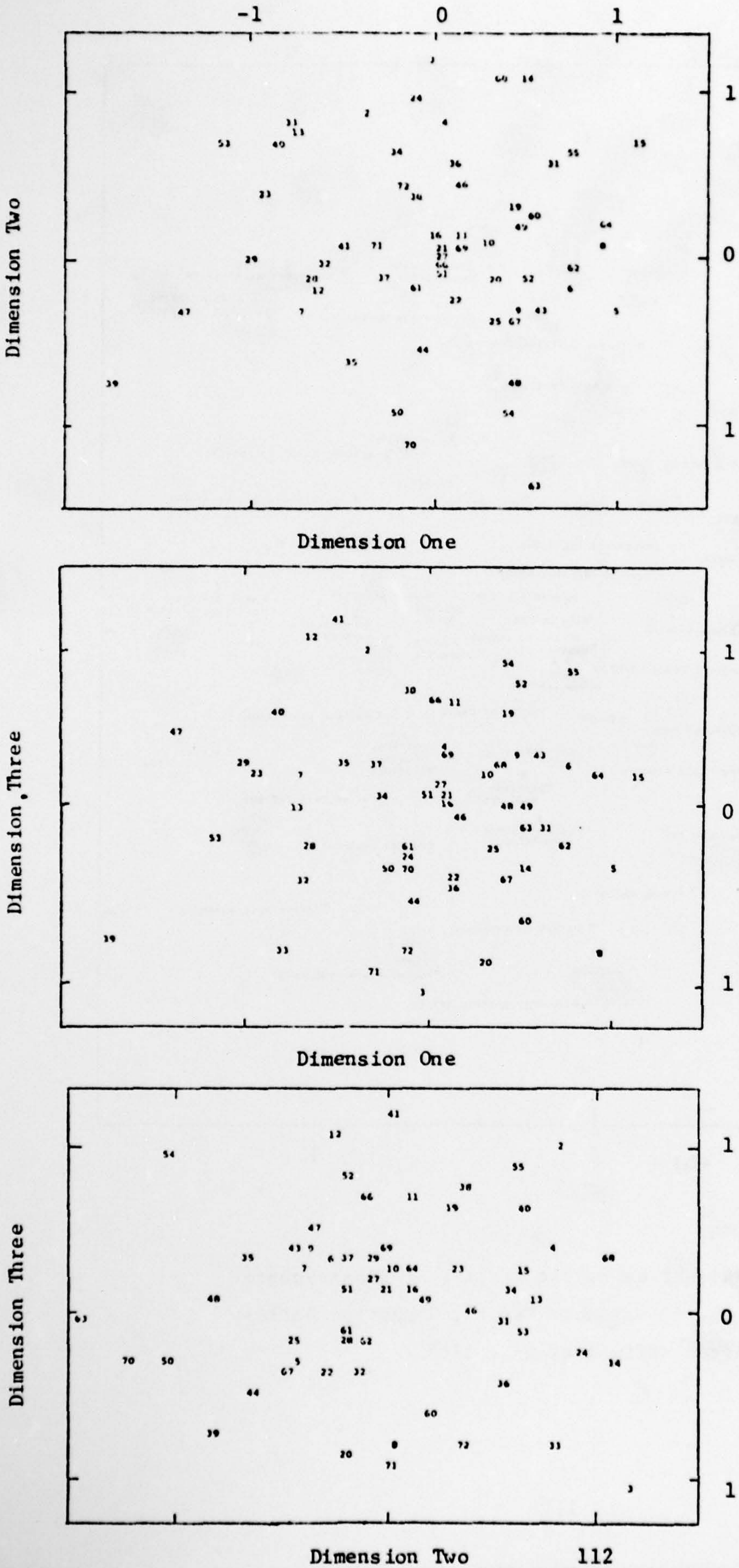


Figure 18. Three dimensional nonmetric scaling of disattenuated correlations from the 60 variable Figural Cognition Battery (After Hoffman et al., 1968).

Flexibility of closure. The next cluster was composed of two NFT tests, Internally Consistent Figures and Penetration of Camouflage. This was the first time that these two tests came together as they did in the Guilford analysis. Both the two and three dimensional scalings of the disattenuated 60 variable matrix located these two tests at the periphery of the same quadrant (see below). Hidden Figures was also hypothesized to be an NFT test and did emerge with a marginally significant loading in the Guilford analysis. Here it clustered elsewhere (at the tail end of the Gf cluster) and was much more centrally located in the multidimensional scalings (see Figures 17 and 18). It would undoubtedly fall even closer to the center if it were more difficult for this sample.

This NFT cluster may represent the ability to break Gestaltbindung that Thurstone attributed to his Flexibility of Closure factor (Thurstone, 1944). That factor was defined by a test called Two Hand Coordination and by Hidden Pictures. The latter is another version of the Penetration of Camouflage test used in this study. Further, it was the easy half of the Gottschaldt Figures test that had the highest loading on Thurstone's Flexibility of Closure factor.

Thus, the ability of break Gestaltbindung is not well measured by complex versions of the Gottschaldt Figures test, such as the Hidden Figures test (French et al., 1963) or the individually administered version of the Embedded Figures test (Wilkin et al., 1971). These tests are usually measures of fluid ability (Gf) and cluster with other complex spatial tasks like Paper Folding or Surface Development (see Snow et al., 1977). The "real" Flexibility of Closure factor is more peripheral in a multidimensional scaling representation or, more specifically, in a hierarchical factor model.

Memory clusters. Three clusters in Figure 16 have been tentatively labeled memory clusters. The first was defined by Angle Estimation (EFR) and Judging Specified Figures (EFC), with Planning Air Maneuvers (NFI) clustering later in the analysis. The common element appears to be short term visual memory for a list of specific visual features.

The second memory cluster was the familiar grouping of System-Shape Recognition (MFS), Orientation Memory (MFS) and Monogram Recall. Remembering Object Orientation (MFS) and Perceptual Relational Judgment (EFR) clustered later in the analysis.

The common requirement of tests in this cluster appears to be short term memory for a larger set of visual features and their interrelationships, particularly their relative positions. It is not clear that the positional information required by most of these tests is the most important test facet. Similar spatial positional memory factors were identified by Christal (1958) and Seibert and Snow (1965). Although there is some evidence that memory for position, color, detail, and form are distinguishable facets of visual memory (Conry and Lohman, 1976; Christal, 1958; Seibert and Snow, 1965), other facets (such as length of presentation, study-test delay interval, and artificiality of the visual display) are probably more important in predicting individual differences in test performance. Most of the tests in this cluster contained a study page and then a test page. The visual image must be retained longer than in tests such as Angle Estimation where both the stimulus figures and the alternatives are drawn on the same side of the paper.

The third memory cluster was composed of Best Map Placement, Identical Forms, Judging Figural Combinations, and Judging Rearrangements. The cluster probably represents the same aptitude complex traditionally known as Perceptual Speed. However, the present designation leads more directly to psychological interpretations. The common denominator here appears to be short term visual memory for a complete image. Thus, the difference between this cluster and the first memory cluster lies in the distinction between a visual feature and a complete image.

Other clusters. The next cluster is particularly interesting, and the interpretation offered here is tentative, yet possibly important for educational research. The cluster was defined by Problem Solving, Necessary Facts, and Block Visualization. The common element appears to be the ability to generate and utilize visual imagery in the solution of verbally stated problems that require verbal solutions. Thus, the generation or manipulation of visual images is not an end in itself, rather the image serves as a mental scratch pad to facilitate representation and solution of the problem. Tests that were only weakly attached to this cluster involved similar sorts of problems but actually provided the figural representation. Thus, in its weakest sense, this aptitude complex might reduce to the ability to utilize figural aids (such as graphs, charts, and schematic drawings) in problem solving. However, the clustering of these tests is only weakly supported by the multidimensional scalings (see Figures 17 and 18). Other facets, such as simple arithmetic reasoning may determine the clustering.

Interpretation of the next cluster is also tentative. The tasks were all easy and seem to involve the ability to reason with figures. This cluster may represent the figural reasoning analog of the SR factor. Thus, while the complex figural reasoning tests like Figure Analogies fell near the center (i. e., "g" or Gf), the simpler, more speeded versions of these tests were more peripheral (see Figure 17).

The next cluster, defined by Correct Figural Trends, Figure Series, Figure Analogies, and Paper Folding, most likely represents Gf or g. The tests were all centrally located in multidimensional scalings, and, with the exception of Paper Folding, are not particularly spatial. However, this version of Paper Folding (which derives from Binet's paper folding task) is more complex than the usual version of this test.

The next cluster is the familiar conglomerate of complex spatial tests that define the Visualization factor. It is noteworthy that the defining tests were Space Positioning and Spatial Visualization. The former test is the prime candidate in this battery for tests that are most easily solved by projecting oneself into the picture and "walking around" the stimuli. Spatial Visualization, on the other hand, is one of the best examples of a test that appears to require a detached mental manipulation (i. e. series of rotations) of the object. This, plus the fact that other Spatial Orientation tests did not cluster together, suggests that while these may represent distinct strategies for solving the tests, both require the same aptitude. On the other hand, the principal components analysis of the entire battery produced a small factor that was tentatively interpreted as representing the ability to project oneself into the stimulus field and "walk around" in it. That such a cluster does not emerge here may result from the exclusionary clustering algorithm.

The next cluster of interest was defined by Match Problems II and Block Rotation. These two tests were extremely close in the three dimensional scaling shown in Figure 18. The cluster reflects the high correlation between these two tests, and probably does not represent a different construct than the one previously identified as Visualization. The interesting feature of this cluster is the correlation that generated it. Block Rotation is one of the better examples of tests that require the mental rotation of a three dimensional object. Match Problems II, on the other hand, does not involve mental rotation. Rather, one must remove a specified number of lines from a given pattern of squares or triangles and leave a fewer, but specified, number of squares or triangles. Further, the student must generate several different solutions for each problem. The important similarity between the tests is that both require the subject to remember an image while performing some transformation on it. In the case of Block Rotation, the task is to remember the relative positions of the sides of the figure while mentally rotating it. In Match Problems, on the other hand, the task is to remember the figure as selected sides are deleted.

This interpretation suggests that Smith's (1964) arguments on the nature of spatial ability are at least partially correct. Mental rotation, while an interesting and special type of mental transformation, is not the most important determinant of spatial ability. Rather, the crucial components of spatial thinking may be the ability to generate a mental image, perform various transformations on it. and remember the changes in the image as the transformations are performed. This ability to update the image may imply resistance to interference, both externally and internally generated. Further, it implies that one of the crucial features of individual differences in spatial ability may lie not in the vividness of the image, but in the control the imager can exercise over the image.

<u>Evaluation of the SI model.</u>

A hierarchical factor analysis was not attempted on this matrix since previous reanalyses have shown the relationship between this factor model and multidimensional scaling representations (see also Marshalek, 1977). Those tests that fell near the center in the scaling representation define higher order factors, while those near the periphery define lower order specifics. It is obvious that a hierarchical structure is present in this matrix, as it has been in all other correlation matrices examined in this review. The nature of this hierarchy is most evident in Figure 15, although Figures 17 and 18 add additional information.

Since the present study represents primarily the figural slice of the model, it is impossible to evaluate the utility of the content facet or the hierarchical ordering of levels of that facet. On the basis of other research, particularly Merrifield (1970), it is reasonable to assume that the semantic-figural distinction is meaningful, since it is congruent with the familiar verbal-spatial distinction. The symbolic factor is probably less distinct, particularly from the figural factor. However, since numbers and letters are termed "symbolic", the facet may represent the large Number factor which emerged in the hierarchical reanalyses of the PMA data (see p. 17), or the Numerical (Gf?) factor that typically falls between the Verbal and Spatial broad group factors (see Snow, Lohman, Marshalek, Yalow and Webb, 1977). Evidence for the differentiation of behavioral content from other content areas is less extensive (see O'Sullivan, Guilford and de Mille, 1965).

Nevertheless, there are indications that some of the SI behavioral abilities are distinct from the other content areas (see Cronbach, 1970), contrary to the negative findings of earlier investigations (Thorndike, 1936; Woodrow, 1939b).

Figural, symbolic, and semantic content facets may define broad second order factors akin to Gv, Gf, and Gc, respectively, or to Space, Number, and Verbal. The behavioral abilities may also define a broad group factor, although it is possible that such a factor would be independent of the other three broad group factors.

The hierarchical ordering of levels of the operation and product facets within the figural content slice of the model is also of interest. Unfortunately, only cognition and evaluation were fully represented in this study. Figure 19 shows a plot of the median general factor loadings of the tests within each of the 23 SI cells purportedly identified in the study. General factor loadings were estimated by the first unrotated factor in the 20 factor principal component analysis of the entire 72 variable matrix.

Insert Figure 19 about here

The most striking separation in Figure 19 was between the divergent production tests and the others. This reflects how the divergent production tests broke away in the initial multidimensional scaling (see Figure 13). The cognition tests were at the other end of the scale. With the exception of the CFI cell (Circle Continuations, Line Continuations), the cognition cells equaled or outranked the others on median general factor loading. The highest general factor loadings were obtained by CFR and CFT. CFR is measured by tests such as Figure Series and Matrices; CFT by Paper Folding and Block Visualization. There is good reason to expect that these tests would have higher general factor loadings on the basis of previous factorial work. The former are versions of Spearman's "g" tests and the latter are Vz tests. The SI model, however, does not predict that some cells will tap abilities that have a broader scope than the abilities tapped by other cells.

Within the figural domain, the rank order of operations was: cognition, evaluation, convergent productions, memory, and divergent production. The placement of convergent production and memory is tentative, since the former was based on only two cells and the latter on one cell. For the product facet, the rank order over the two operations with complete data (cognition and evaluation) was: transformations, relations, systems, classes, implications, and units.

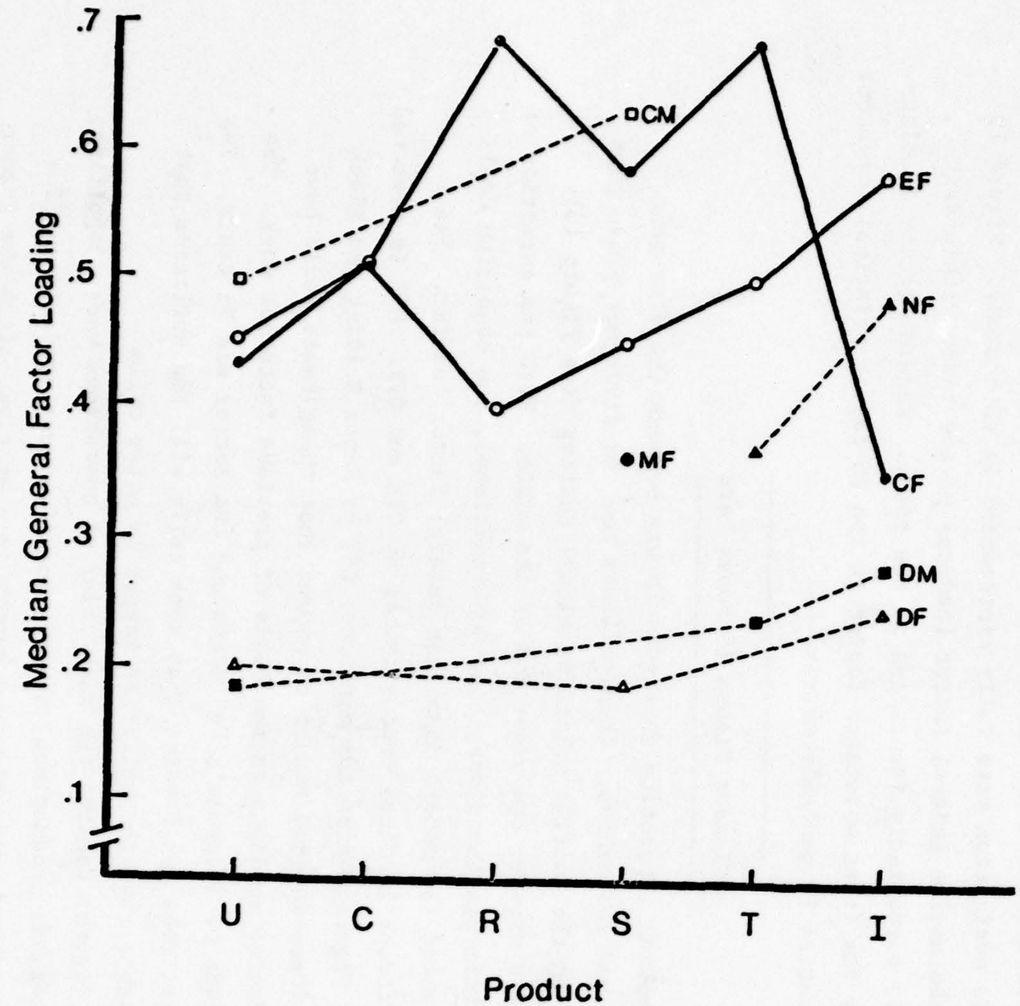


Figure 19. Median general factor loadings of the 23 SI abilities identified by Hoffman et al. (1968).

There is obviously some hierarchical structure within both the operation and product facets. However, the exact nature of this structure is not the major issue here. Rather, it is the simple fact that some sort of hierarchy exists that is most troublesome for the SI model and other theories of parallel abilities. While in his more recent statements Guilford has moderated his views on the possibility of higher order factors (see Cronbach and Snow, 1977, p. 154), earlier expositions of the model emphatically reject the notion of hierarchical structure (see Guilford and Hoepfner, 1971, p. 22).

As Humphreys (1962) has pointed out, hierarchical and facet models are not inherently contradictory. For example, the most reasonable hierarchical coordination of the SI model would place the four content areas as broad group factors, the various operations as narrow group factors under each content area, and the product cells as specific factors beneath each narrow group product factor.

The most troublesome fact for this representation and the SI model is that particular cells, like CFR, CFT, CMR, properly fall at the top of the hierarchy. The SI model predicts that if there are group factors they should be "along the lines of the categories of the SI model" (Guilford, in Cronbach and Snow, 1977, p. 155). That particular cells exhibit this property is contrary to even this more liberal view of parallel abilities within the SI model.

In conclusion, there are a number of problems with the SI model. The levels of some facets are particularly questionable (e.g., Is convergent production different than cognition? Are relations and transformations products or types of cognition?). The most glaring deficit of the model, however, is its inability to account for the fact that some tests correlate with a large number of other tests, while others correlate with only one or two other tests.

Since the SI model is probably faulty, attempts to coordinate it with a hierarchical model are doomed to confusion and contradiction. Building a facet model that translates into the familiar hierarchical model would be worthwhile. However, it would be better to start with something like Eysenck's (1967) three way classification of mental process (perception, memory, reasoning); by test material (verbal, numerical, spatial); by quality (speed to power). Particular levels of Guilford's model could be included, such as behaviorial content or divergent production. Beyond this, however, it would appear more profitable to abandon the SI model than to attempt to coordinate it with other factor models or process theories of intelligence.

The Upper Levels of the Hierarchy

While this review has attempted to maintain a hierarchical perspective, the focus has been on the lower branches of the tree. In particular, analyses have attended to the number and psychological nature of the space subfactors. There has been a deliberate ambiguity in these analyses about the nature of the "general plus broad group spatial factor."

The reason is that none of the studies reviewed sampled a sufficient number and variety of first order factors to make second order analysis possible and enlightening.

Hierarchical Versus Gf-Gc Theory

Spatial ability has always appeared at the second level in British hierarchical models, first clustered with the practical-mechanical abilities (Burt, 1949), later with the practical abilities (Vernon, 1960), and, most recently, alone (Smith, 1964).

The only strong competitor for this model is the pseudo-hierarchical model proposed by Cattell (1963), and later modified by Horn (Horn and Cattell, 1966; Horn and Bramble, 1967) and Cattell (1971). The model is not a true hierarchy because it explicitly denies that there is a unitary structure called general intelligence. In the earliest formulation, the model posited two correlated general intelligences: fluid ability (Gf) and crystallized ability (Gc). Spatial ability fell under the Gf factor (Cattell, 1963; Horn and Cattell, 1967).

In a later study that sampled a broad range of ability and personality primaries, three additional "general" (i.e., second order) factors were identified: Visualization (Gv), Speed (Gs), and Fluency (F). Spatial tests were moved from Gf to the General Visualization factor. Although Cattell (1971) later called this factor a provincial power (pv), the major spatial factors (SR and Vz) were still hypothesized to cluster with Closure Speed (Cs), Flexibility of Closure (Cf), and Adaptive Flexibility (DFT) at the second level.

Later forumulations of Gf-Gc theory, particularly the triadic theory of abilities (Cattell, 1971) rely heavily on the one published study (i.e., Horn and Cattell, 1966) that sampled a sufficient number and variety of first order factors to permit meaningful second order analyses. The study is important, therefore, on two counts. First, it is undoubtedly one of the most comprehensive batteries of well known primaries (in the tradition of Thurstone, 1938; French, 1951; and French, Ekstrom and Price, 1963) yet administered. Second, it underpins much of the recent work on extensions of Gf-Gc theory, particularly the triadic theory of Cattell (1971).

The Horn and Cattell (1966) Study.

Horn and Cattell (1966) administered a battery of tests representing 23 primary ability factors and 8 general personality dimensions to 297 volunteers. Of these, 215 were males, and most were prison inmates. The average age was 27.6 years, the standard deviation 10.6 years, and the range 14 to 61 years. Fourteen of the ability factors were measured by two or more tests, and the remainder by only one test. Scores for those primaries represented by more than one test were obtained by summing the scores for the various tests.

The correlation matrix of tests or test clusters assumed to measure the 31 first order primaries was then computed. Thus, a first order factor analysis was not performed. This matrix was then factored by the principal factoring method, with 25 iterations. Nine factors were extracted, first rotated to a varimax criterion, and then graphically rotated to oblique simple structure.

The personality variables, which were largely uncorrelated with the ability variables, were used to define the hyperplanes in these rotations. Cattell (1971) argues that this "hyperplane stuff" is critically important in any second order analysis to achieve true simple structure.

Results. Of the nine second order factors, three were personality factors and one was an "ability" singleton defined by the Carefulness primary. The remaining five constituted the second order ability factors of interest: Fluid ability (Gf), Crystallized ability (Gc), Visualization ability (Gv), General speed (Gs), and General fluency (F). With the exception of the correlation between F and Gc, the correlations between the five second order ability factors were all positive. The Gv factor was the most oblique; its average correlation with the other four factors was .232. Corresponding values for the other factors were .218 (Gf), .216 (Gs), .10 (Gc), .078 (F).

Reanalysis. A nonmetric multidimensional scaling was performed on the correlation matrix for the 23 primary ability tests and test clusters. The eight personality factors were included in a second analysis, but all fell on the periphery and served only to increase the stress. Thus, while the personality variables may be useful for defining hyperplanes in factor analysis, they were not particularly useful here. Primary factors, their abbreviations, and the tests used to measure them are shown in Table 33.

Insert Table 33 and Figure 20 about here

As in prior analyses, the scaling was performed by the KYST program. Stress values (formula 1) were .083, .116, and .159 in four, three, and two dimensions, respectively. The final two dimensional configuration is shown in Figure 20a. The points in Figure 20a were clustered on the basis of their loadings in the oblique factor pattern matrix of Horn and Cattell (1966).

Hierarchical cluster analyses were also performed on 23 variable correlation matrix using Johnson's (1967) HICLUS program. Both minimum and maximum cluster analyses were performed. Variables that clustered together in all three analyses (minimum method cluster analysis, maximum method cluster analysis, Horn and Cattell (1966) factor pattern matrix) are grouped by a solid line in the multidimensional scaling in Figure 20b. Variables that also entered these clusters in at least two of the three analyses are indicated by a broken line.

The major difference between Figures 20a and 20b is the disappearance of the Gv cluster in Figure 20b. In the Horn and Cattell analysis, this factor was defined by Vz (.50), followed by S (.48), Cf (.45), and DFT (.42), with minor loadings from CFR (.35) and Cs (.31).

As often happens, however, the labels tell very little about the factors. In this case, the tests that defined the various primaries are exceptional in several respects. First, the Vz and Cf primaries were virtually coincident in the multidimensional scaling. This concurs with previous analyses reported here and elsewhere (see Snow, Lohman, Marshalalek, Yalow and Webb, 1977). However, it is unusual for both to be about as distant from the Gf cluster (defined by L, I and CFR) as the Cs cluster. Complex Vz spatial tests have fallen much closer to Gf than the speeded Cs tests in previous analyses. Further, DFT was more peripheral in this analysis than the corresponding test (Match Problems II) was in the analyses of the Figural Cognition Battery (see p. 97). Parenthetically, neither Guilford's analysis (Hoffman et al., 1968) nor the reanalyses of that data indicated that Match Problems was a DFT measure.

Much of the confusion may be attributed to the increased speededness of the Vz, Cf, and DFT tests. The Vz and DFT tests (Form Board and Match Problems) were shorter and more highly speeded here than they were in Thurstone (1938) or Hoffman, Guilford, Hoepfner and Doherty (1968). Further, the Cf factor was represented by the speeded Designs test (see p. 33) rather than by the more

Table 33

Primary Factors, Abbreviations, and Tests
(After Horn & Cattell, 1966)

Primary Factor	Symbol	Tests
Induction	I	Letter Grouping Number Series
Intellectual Speed	SP	Test A(2) Series - Furneaux
Carefulness	C	Test B(1) Series - Furneaux Figure Classify (20-W) Practical Estimates (20-W) Subtracting (9-W) Dividing (20-W) Fractions-Decimals (20-W)
Intellectual Level	L	Test B(2) Series - Furneaux
Figural Relations	CFR	Figure Series Topology Matrices Speed Matrices Power Figure Classify
General Reasoning	R	Problem Solving
Adaptive Flexibility	DFT	Match Arrangements
Spatial Orientation	S	Cards Figures
Visualization	Vz	Form Boards
Associative Memory	Ma	Cued Nonsense Memory Cued Meaningful Memory
Semantic Relations	CMR	Common Word Analogies Abstruse Word Analogies
Verbal Comprehension	V	Vocabulary General Information
Mechanical Knowledge	Mk	Mechanical Information General Information
Formal Reasoning	Rs	False Premises Influences
Experimental Evaluation	EMS	Social Situations
Associational Fluency	Fa	Controlled Associations
Ideational Fluency	Fi	Things Round Ideas
Number Facility	N	Adding Multiplying Mixed Operations
Speed of Closure	Cs	Backward Reading Street Gestalt
Flexibility of Closure	Cf	Designs
Speed of Copying	Sc	Forward Writing Forward Printing
Writing Flexibility	Wf	Backward Writing
Perceptual Speed	Ps	Match Letters & Numbers Rapid Cancellation

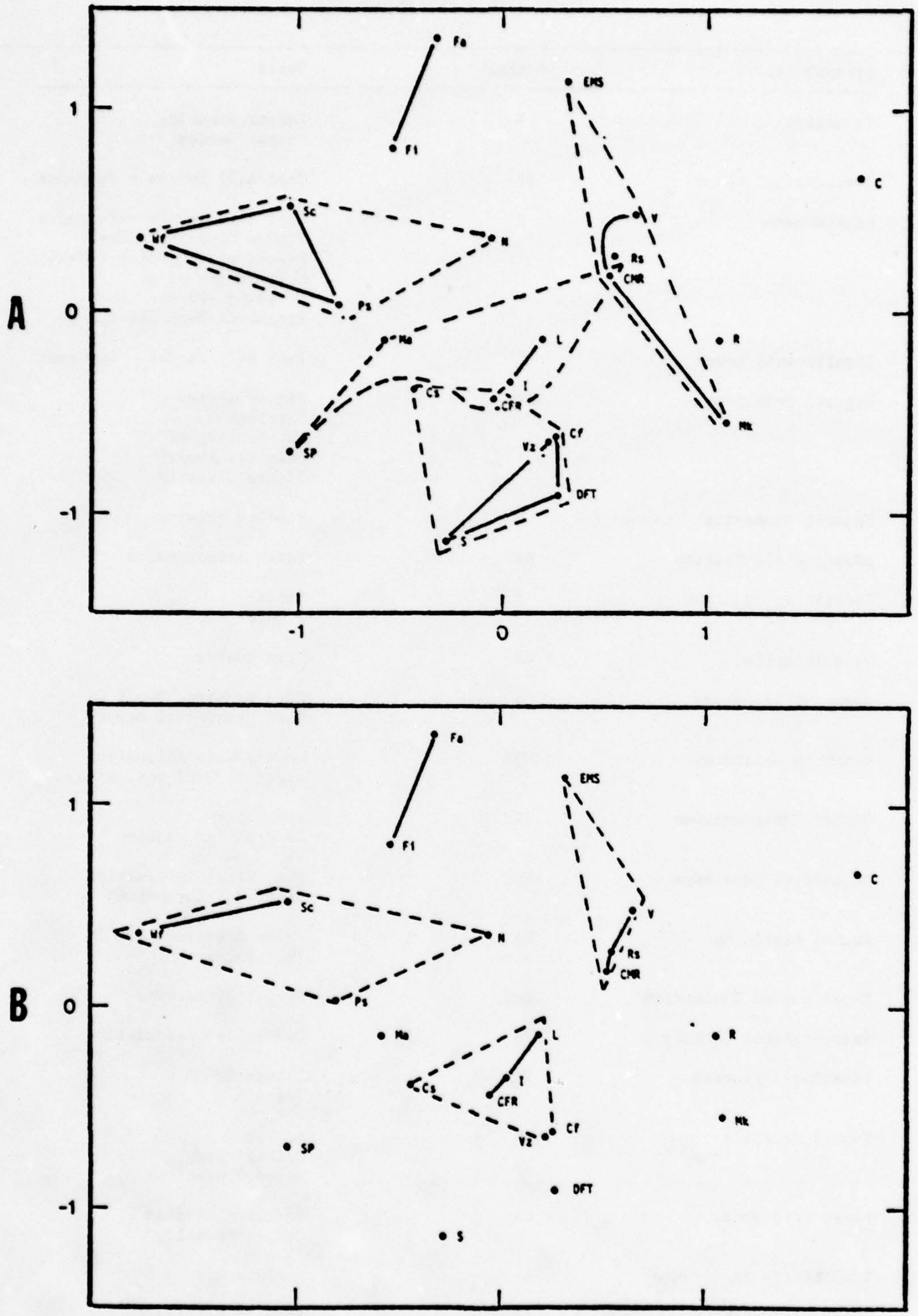


Figure 20. Two dimensional scalings of the Horn & Cattell (1966) data, (A) clustered by Horn & Cattell factor analyses, and (B) cluster in two or three different analyses.

complex, power tests like Hidden Figures (French et al., 1963) or the more difficult version of the Gottschaldt Figures.

The Cs factor, on the other hand, was probably more complex than usual. The factor here was defined by an adaptation of the Street Gestalt test and Backward Reading. The latter test, under the name Mirror Reading, loaded primarily on the Perceptual (Speed) factor, with a minor loading on the Word Fluency factor in Thurstone and Thurstone (1941). Botzum (1951) used the same test in his study of reasoning and closure factors, but under the label of Backward Writing. The test helped define his Cs factor, but also loaded on the Cards-Figures Space factor, which he attributed to the possibility of solving the test by mentally rotating the reversed word. Mooney (1954) used a similar test that defined a factor he called Verbal Closure. The test did not even load on the Cs factor in his analysis. Thus, the Backward Reading test used to define the Cs factor in the present study is, at best, factorially complex. It may measure Perceptual Speed, Space, or Word Fluency in addition to Cs; or may even represent a different type of "closure." This may explain both the location of Cs in Figure 20 and why it clustered with the other spatial tests so early.

The other "general" factors are also suspect. The Gs factor appears to be a Motor or Writing Speed factor. As such, it is more of an overblown primary than a general speed dimension. The F factor is merely a fluency doublet in this study, although the reanalyses of Guilford's work do support a broad Fluency, Divergent Production, or Verbal Productive Thinking factor (Horn, 1976) that is independent of fluid ability and only slightly related to verbal ability (see p. 97 ff).

Gc was not well represented here, and appears to be no more than "a swollen V" (Horn, 1976, p. 443). However, analyses of the Aptitude Project reference battery (Snow, Lohman, Marshalek, Yalow and Webb, 1977) have shown that a broad, verbal achievement-based Gc factor can be separated from Gf at the second order, especially if the complex spatial tests are allowed to represent Gf rather than Gv.

Finally, neither the multidimensional scaling nor the cluster analyses indicated that Associative Memory (Ma) and Intellectual Speed (Sp) should cluster with the Gf factor before the Vz, Cf, and Cs primaries. Indeed, as is evident in Figure 20a, only a severe distortion of the scaling would bypass Cs and bring Sp into the Gf factor.

For the present, then it would appear more parsimonious to speak of two broad group "intelligence" factors, Gf and Gc. Complex spatial tests such as Guilford's Paper Folding fall near the center of the Gf factor, along with tests like the Raven Matrices, Figure Classification, and the like. Less complex, more speeded tests and their factors like Cs, SR, and Ps fall further out in the scaling or further down in the hierarchical model. This is shown both here (see Figure 20b) and in the reanalyses of the Figural Cognition Battery (see Figure 17).

It may be that a broader sampling of visual cognition abilities would necessitate a Gv factor. The reanalyses of Figural Cognition Battery suggested that this may not be so, although all of the figural and spatial tests in that battery were of the paper and pencil variety.

Finally, although Cattell (1963, 1971) argues on theoretical grounds that a general factor combining Gf and Gc at the third level is neither necessary nor meaningful, these analyses indicate the opposite. Such a factor will, as Cattell (1971) notes with dismay, be defined by the Gf factor. In particular, it will capture much of the variance in tests like Matrices, Figure Classification, and Letter Series. What remains in the Gf and Gc factors after G is removed is the familiar verbal-spatial bipolar factor. This is evident here (see Figure 19), and was precisely the result obtained by Snow et al. (1977; see also Marshalek, 1977).

Contrary to Cattell, there are good reasons to expect that tests like the Raven Matrices will be explained primarily by G and, further, good reasons why the verbal-spatial dichotomy is psychologically meaningful. In spite of all the physiological and neurological evidence that Cattell (1971) cited to support his triadic theory of abilities, he failed to recognize the importance of the recent work on the hemispheric lateralization of verbal and spatial functions. Although much of the research in this area is sorely inadequate, there is now a substantial literature supporting the hypothesis that verbal and spatial stimuli are processed with differential efficiency by the two hemispheres. Further, some tests, such as Raven Matrices, may be good measures of general intelligence because they require the active participation and cooperation of both hemispheres (see Zaidel and Sperry, 1973). Thus, there are good biological and psychological reasons for the verbal-spatial distinction, as well as for the concept of general intelligence.

Conclusions

Spatial ability may be defined as the ability to generate, retain, and manipulate abstract visual images. At the most basic level, spatial thinking requires the ability to encode, remember, transform, and match spatial stimuli. Factors like Closure Speed (i.e., speed of matching incomplete visual stimuli with their long term memory representations), Perceptual Speed (speed of matching visual stimuli), Visual Memory (short term memory for visual stimuli) and Kinesthetic (speed of making left-right discriminations) may represent individual differences in the speed or efficiency of these basic cognitive processes. However, these factors surface only when extremely similar tests are included in a test battery. Such tests and their factors consistently fall near the periphery of scaling representations, or at the bottom of a hierarchical model.

Major Spatial Factors

While the processes that these factors hypothetically represent are certainly spatial in nature, they are not usually the referent of the term "spatial ability." While a number of "spatial" factors have been identified, only three survived this review. All of the factors involve mental transformation. They are:

1. Spatial Relations. This factor is defined by tests like Cards, Flags, and Figures (Thurstone, 1938). The factor appeared only when these or highly similar tests were included in the same test battery. Although mental rotation is the common element, the factor probably does not represent speed of mental rotation. Rather, it represents the ability to solve such problems quickly, by whatever means.

2. Spatial Orientation. This factor appears to involve the ability to imagine how a stimulus array will appear from another perspective. In the true spatial orientation test, the subject must imagine that he is reoriented in space, and then make some judgment about the situation. There is often a left-right discrimination component in these tasks, but this discrimination must be made from the imagined perspective. However, the factor is difficult to measure since tests designed to tap it are often solved by mentally rotating the stimulus rather than by reorienting an imagined self.

3. Visualization. The factor is represented by a wide variety of tests such as Paper Folding, Form Board, WAIS Block Design, Hidden Figures, Copying, etc. In addition to their spatial-figural content, the tests that load on this factor share two important features: (a) all are administered under relatively unspeeded conditions, and (b) most are much more complex than corresponding tests that load on the more peripheral factors. Tests designed to measure this factor usually fall near the center of a two dimensional scaling representation,

and are often quite close to tests of Spearman's "g" (such as Raven Matrices or Figure Classification) or Cattell's (1963) Gf.

Types of Spatial Transformations

Two types of mental transformation appear to be involved in tests that load on these three spatial factors. The first is mental movement. Reflecting, rotating, folding, or simply imagining that a stimulus is moved from one position in an array to another position, are all varieties of mental movement.

The second type of mental transformation may be called construction. There are two types of constructions: reproduction (i.e., physical construction) and combination (i.e., mental construction). At the simplest level, reproduction is represented in tests like Thurstone's (1938) Copying, where the subject must correctly copy a stimulus design. At the next level, it is represented by tests like Graham and Kendall's (1948) Memory for Designs, where the design must be reproduced, not just recognized, and the reproduced design must be a veridical representation of the stimulus. Retaining a veridical mental image of a design may be an important component of other complex spatial tasks, such as Hidden Figures (French et al., 1963).

In the mental construction tasks, on the other hand, the subject must actually construct a mental image, usually by reorganizing the stimulus in a new way. The clearest examples of this sort of process are tests like Form Equations (El Koussy, 1935) and Paper Form Board (e.g., Thurstone, 1938; French, Ekstrom and Price, 1963). Mental construction is an important component of many complex spatial tests. For example, in Paper Folding (French et al., 1963), the examinee must construct new holes as he mentally unfolds the stimulus. Finally, mental construction may take the form of mentally deleting parts of a stimulus, as in Match Problems (Guilford and Hoepfner, 1971). This may also be an important component of tests such as Embedded Figures (Witkin, Oltman, Raskin and Karp, 1971) or Hidden Figures (French et al., 1963).

A word of caution, however. The central characteristic of spatial ability may lie in the nature of the internal representation rather than in the speed or efficiency of the various mental transformations applied to the image.

Underlying Dimensions

It is now apparent that one of the basic questions posed at the beginning of this review ("How many space subfactors are there?") cannot be answered with certainty. The important question, then, is "What are the dimensions along which tests and test clusters (i.e., factors) are arrayed?" Particular factors are then seen as reference points on these continua. With this in mind, it is possible to array the various spatial factors as in Figure 21. This representation is a crude distillation of many studies, particularly the multidimensional scaling of Thurstone (1951) (see Figure 7) and Hoffman, Guilford,

---

Insert Figure 21 about here

Hoepfner and Doherty (1968) (see Figure 13).

The Vz factor at the center of the figure represents the complex spatial tests (such as the Guilford-Zimmerman Paper Folding test). The factor is synonymous with the Cf factor, but only when the latter is measured by the more complex Gottschaldt Figures tests (e.g., Hidden Figures) or the WAIS Block Design. Less complex tests of the sort that defined the Cf factor in Thurstone (1944) or Horn and Cattell (1966) would be more peripheral.

The factors at the periphery of Figure 21 are defined by simple, highly speeded tests. Thus, the spokes of the wheel radiating out from Vz represent, simultaneously, a shift from power to speed and from complex to simple. These peripheral factors probably represent individual differences in the speed of various mental processes. Thus, Cs may represent the speed of identification of incomplete or distorted visual information; Ps the speed of matching visual stimuli; SR the speed of executing a particular mental transformation (rotation or reflection); K the speed of making left-right discriminations; and M the speed and effectiveness of storing visual information in short term memory.

The SR factor is more central than the other peripheral clusters, possibly because individual differences in other peripheral clusters, especially Ps and K, influence performance on SR tests. On the other hand, the process of mentally rotating a figure may be more complex than matching or making left-right discriminations.

The SO factor is located close to the center in Figure 21, probably because it is difficult to construct SO tests that are not susceptible to a Vz solution strategy. A "true" SO factor would probably be much more peripheral (e.g., see Figure 13). The connection between SO and the K and M primaries emerged in several studies. Making rapid left-right discriminations (K) may be one component of SO, or it may represent the degenerate or most highly speeded version of the SO factor. The connection between SO and Visual Memory (M) was particularly evident in the reanalysis of the Figural Cognition Battery (see Figure 13). Imagining a reorientation of the self could put considerable burden on visual memory. On the other hand, if the SO tests are solved by

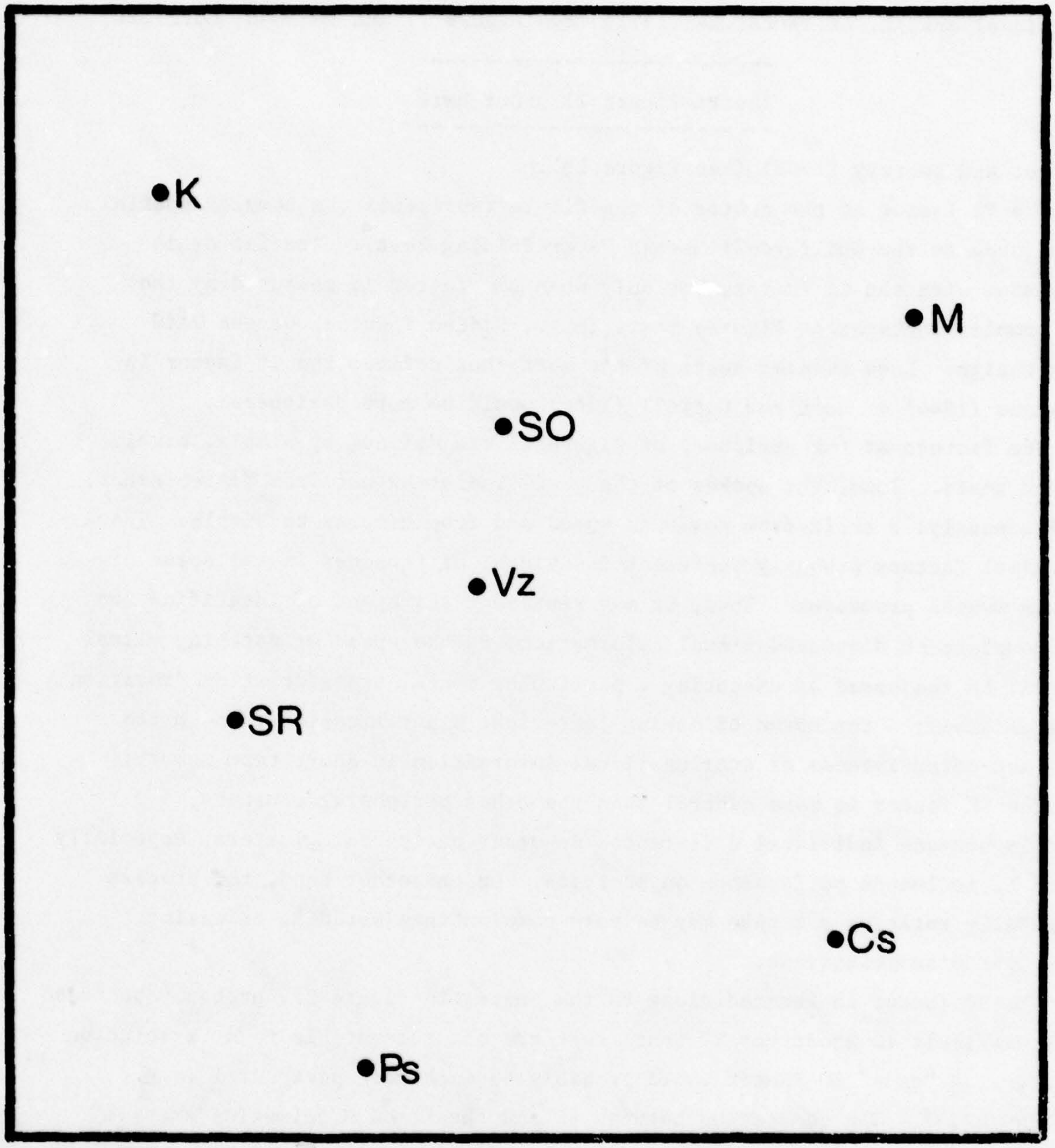


Figure 21. Proposed two dimensional scaling of spatial ability factors.

mentally rotating the entire visual array (a Vz strategy) then this too would put substantial burdens on visual memory.

While the positions of the factors in Figure 21 have some empirical support, they are by no means fixed. Thus, it is probably not wise to read too much into this particular representation. It may be best to speak instead of two independent dimensions, as in Figure 22.

Insert Figure 22 about here

Here, the vertical axis represents the speed-power-complexity continuum, while the horizontal axis represents the nature and perhaps complexity of the cognitive process itself. The ordering of processes along this dimension is based on logical considerations.

This representation, however, lends itself to an explanation of a variety of factorial phenomena. First, factors emerge only when individual differences in the particular processes required by tests can be elicited with sufficient strength to be reflected in the dependent measures that are employed. For example, individual differences in the number of pictures correctly identified is elicited by degrading, distorting, or erasing part of the picture, as in the various Closure Speed tests.

Second, task complexity may be increased by increasing either the number of distinct operations, or the difficulty of each operation. Thus, tasks that elicit individual differences in memory and transformation should be more complex, and thus produce a factor further up in the hierarchy than comparable tasks where individual differences in only one component are elicited. On the other hand, task complexity may be increased by increasing the difficulty of the component that produces individual differences. Thus, Kagan's Matching Familiar Figures test should be more complex, and hence, further up on the hierarchy than Thurstone's Identical Forms. Also, different operations within a class may be inherently more complex than others. Thus, mental rotation may be a more complex process than reflection, even though both would be classified as transformations.

While Figures 21 and 22 provide rough schema of the organization of the important spatial factors, neither representation shows how spatial abilities fit in larger models of human abilities. The reanalyses of the Thurstone (1951), Hoffman et al. (1968), Horn and Cattell (1966) data, together with analyses of another large test battery (Snow et al., 1977) suggest that

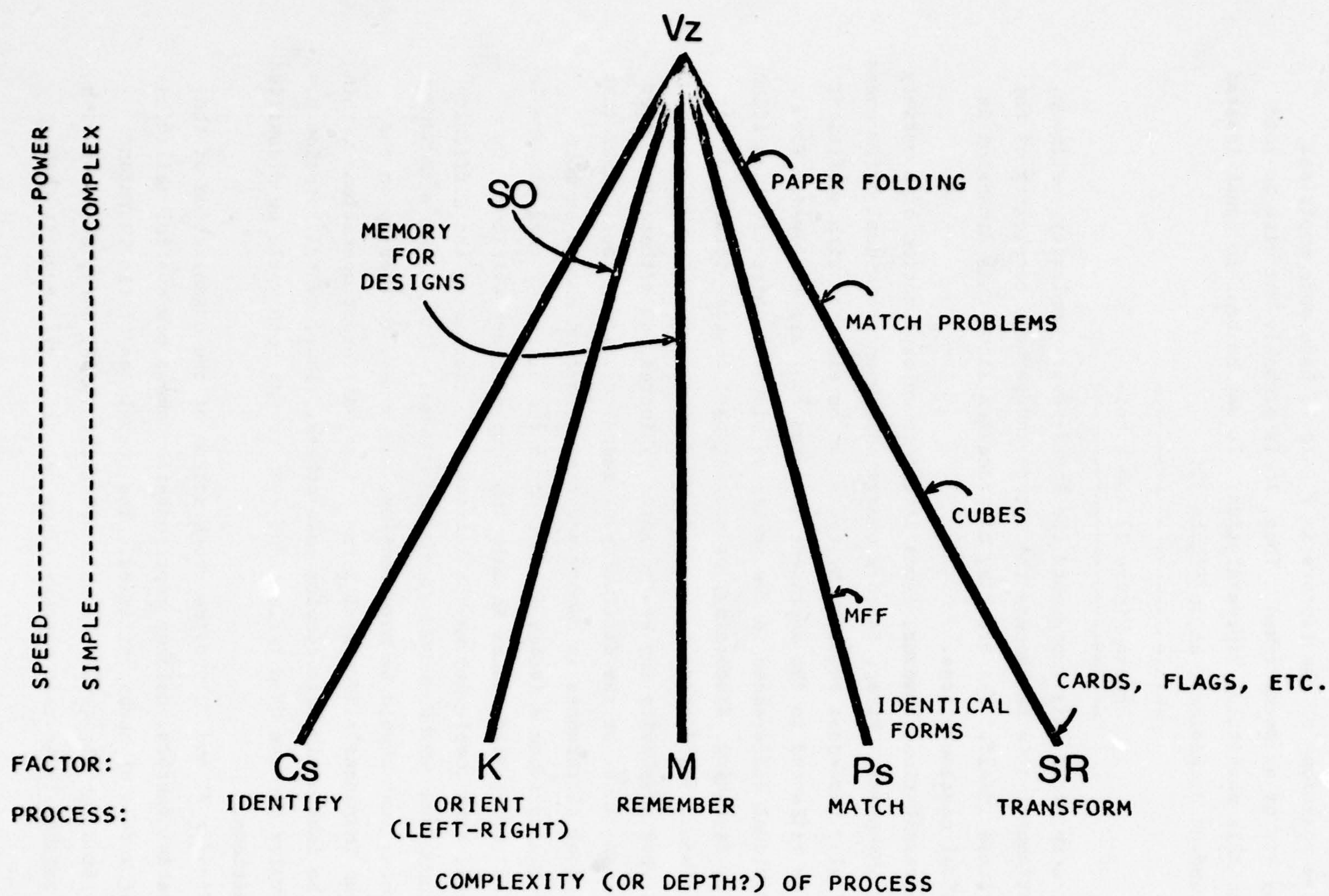


Figure 22. Relating spatial tests and factors to process, speed, and complexity.

general intelligence may be split into a verbal-educational cluster and a spatial-figural cluster. Whether the spatial tests split away from the figural tests and form a broad group factor at the second level depends on what spatial tests are used and how the tests are clustered. Very complex spatial tests have their primary loading on the general factor. If simpler, speeded versions of these tests are used (like that of French et al. (1963)--Paper Folding) the complex spatial tests form a Vz factor that is slightly independent of G (see Snow et al., 1977). If these Vz tests are clustered with Cs, Ps, SR, and SO tests (as in Horn and Cattell, 1966), then a broad group spatial factor emerges which is even further removed from G. However, most of the common variance in the Vz tests still falls on the general factor, while Ps, Cs, and similar tests have their largest loadings on the narrow factors such as Ps and Cs. Only tests of intermediate complexity (like SR and SO tests) have their largest loadings on the broad group spatial factor. The nature of this second order factor changes as different tests representing different factors are included in the analysis. If only the Vz and SR tests are included in the analysis then the broad group spatial factor shifts closer to G. If Cs, Ps, and K tests are also included, the broad group spatial factor becomes more independent of G.

Cattell (1971) argues that the location of higher order factors can be determined with greater assurance by including sufficient primaries and enough "hyperplane stuff" in the analysis to permit oblique, simple structure rotations. While these procedures may be helpful in properly locating higher order factors, there are so many uncontrolled sources of variation in traditional tests that it is doubtful that such factors could ever be fixed with assurance. It would seem more profitable to try to understand the processes involved in spatial thinking than to determine whether such abilities fall under Gf or form a separate second order factor like Gv.

Spearman Revisited

One of the difficulties repeatedly encountered in this review was that primary factors such as Cs, Ps, and M did not cluster together to form narrow group factors like SR or SO. Attempts to fit this sort of complete hierarchical model to the Thurstone (1951) Mechanical Aptitude study (see p. 67 ff) were particularly unsuccessful. There appear to be just two types of "pure" factors: speed factors and power factors. Speed factors are largely independent of one another and of power factors, while the power factors are strongly intercorrelated. Further, the number of potential speed factors is probably infinite, while only three, content based power factors were identified in this review: verbal, spatial, and numerical or symbolic.

In the verbal domain, tests like verbal analogies, vocabulary, and reading achievement represent the power end, while tests for primaries like verbal fluency, ideational fluency, and reading speed fall on the speed end of the spectrum. In the spatial-figural domain, complex tests like Figure Classification, Raven Matrices, and the Guilford-Zimmerman Paper Folding test form the power end, while primaries such as SR, Cs, and Ps form the loose collection of speed factors. For the Numerical-Symbolic content area, tests like Arithemetic Achievement, Letter Series, and Necessary Arithmetic Operations come together at the power end, while speed of computation tests (i.e., Thurstone's Number primary), clerical speed tests Finding A's or Number Comparison, and, perhaps, memory span tests, represent the speed end. The power tests are all highly correlated. If power tests from the three content areas are allowed to form separate factors, the verbal-spatial distinction holds, while the numeric-symbolic factor is engulfed by G. Similar distinctions may be made in other content areas, such as motor (or writing) speed and behavioral-social intelligence. However, these tests and their factors are only minimally related to general intelligence. Suffice it to note that part of the variance in some clerical speed tests like the WAIS Digit Symbol may be attributed to motor or writing speed.

The crucial issue for a hierarchical theory, however, is that the power factors cannot be subdivided into the various speed primaries. Further, the speed primaries are largely independent of one another. What little correlation exists between them may be attributed to overlapping content. Thurstone and Thurstone (1941), Botzum (1951), and Horn and Cattell (1966) all obtained second order factors by combining various speed primaries. However, none of these second order factors were coincident with similar factors defined by power tests in the same content area. Similarly, Horn and Cattell's Gv was much more independent of G than the Vz factor obtained in the reanalysis of that matrix.

But is it reasonable to expect that power in a particular content area may be defined by adding up various speed indices? If there is a shift from power to speed as one moves up the hierarchical model, and if speed of performing simple tasks is largely independent of power with the same types of tasks, then it is impossible to define a general power factor by combining speed primaries. The attempt is akin to the alchemists' efforts to produce gold by combining other chemicals.

Thus, this review comes full circle to Spearman's (1927) contention that there is a general factor, a few content specific group factors, and a (possibly infinite) number of independent specifics. The general factor is defined by power tests, the broad group factors by the various content areas, and the specifics by the various speed tests. Further, as the various factors are presently represented, the model does not form a true hierarchy, since power does not decompose into speed.

The Value of a Common Perspective

In spite of the inadequacies of the hierarchical model, reanalyzing a host of conflicting studies from a common theoretical perspective has revealed a remarkable consistency in factor structures. Most of the confusions in the correlational literature on the number of different spatial factors were traced to different methods of factor extraction and rotation. Other major sources of conflict were related to differences in subject populations, test speededness (or complexity), and individual differences in solution strategy. Low ability samples showed less differentiation of abilities and anomalous factor structures. There were also indications of important sex differences in factor structure. Decreasing the complexity or increasing the speededness of a test also changed the factor structure, making the test and the factor it helped define more specific. Finally, it was hypothesized that individual differences in solution strategy were a major source of confusion, making a test appear factorially complex or causing it to load on different factors in different studies. Solution strategies and the relationship between speed and level on spatial tests are reviewed in the next two sections.

SOLUTION STRATEGIES ON SPATIAL TESTS

Early Work

Although a few early investigations gathered evidence on solution strategies, data were rarely analyzed systematically. Rather, they were used only to help label or interpret factors. For example, El Koussy (1935) obtained introspective reports from some of his subjects on how they solved various tests in his battery. Many reported using verbal imagery to solve tests which loaded on his k factor. On the basis of these introspections, he concluded that the k factor represented the ability to generate and utilize spatial imagery.

Much of the reluctance to investigate individual differences in solution strategies undoubtedly stemmed from the behaviorist taboo on introspective evidence. However, even those who recognized the possibility of strategic differences seemed to regard them as of only minor importance. There appears to have been a blind faith in the power of factor analysis to disentangle the multiple sources of individual differences in test performance. Perhaps the best example of this is a study by Michael, Zimmerman and Guilford (1950) (see p. 84). After careful exposition of several hypotheses about the possible psychological differences between the SR and Vz factors, they simply administered a battery of tests and factored the correlation matrix. Some introspective reports were gathered, but again they were used only to interpret and, at times, rationalize the results.

Barratt

The first systematic attempt to utilize retrospective verbal reports in understanding individual differences in spatial test strategy was reported by Barratt (1953). He administered seven spatial and three verbal tests to 84 college males. The space tests included three SR tests (Flags, Cards, and Figures), one Vz test (DAT Space Relations), and three hypothesized SO measures (Guilford-Zimmerman Spatial Orientation, Industrial Aptitude Spatial Orientation subtest, and a new test called the Barratt-Fruchter Chair-Window test). Unfortunately, other Vz tests were not included in the study to help define that factor.

Four centroid factors were extracted from the correlation matrix, and then rotated to orthogonal simple structure. The three verbal tests defined the first factor; Cards, Flags, and Figures defined the second; the Chair-Window tests and the Industrial Aptitude Spatial Orientation test defined the third factor; and DAT Space Relations and Guilford-Zimmerman Spatial

Orientation test defined the fourth. Although Barratt used slightly different labels, the factors are obviously Verbal, SR, SO, and Vz. As often happens, the Guilford-Zimmerman Spatial Orientation test loaded on the Vz factor, rather than on the SO or SR factors.

Barratt also collected retrospective reports of how each student solved the spatial tests. Subjects were first asked to describe how they solved the problems and then later asked more pointed questions. Analysis of the interview protocols took several forms. The final product was a definition of the problem solving processes tapped by each factor and a list of more specific contrasts on strategy differences for individual tests.

Barratt defined the SR factor as "the ability to turn or rotate a given figure or part of that figure in one plane (or about an imaginary axis) to see if it corresponds to another figure in the same plane" (p. 20). In all, 82 of the 84 subjects used a method that fit this definition. The two discrepant subjects tried to use angles and figural cues only, without rotating the stimuli.

The Vz factor was defined as the "ability to see or observe the spatial relationship of objects involved in dynamic situations, spatial relationships in which the subject has to imagine that the object or objects involved changed their positions in space relative to one another" (p. 21). Between 76 and 83 subjects were classified as using this method on the DAT Space test, depending on the difficulty of the items.

The SO factor was defined as "the ability to determine from where you are looking at an object; i.e., where one is spatially located in relationship to a particular object" (p. 22). On the Industrial Aptitude Spatial Orientation Test, 58 subjects used a method similar to this definition.

In addition to these factor definitions, several specific strategy contrasts were noted for each test. For the Figures test:

1. Of the 82 subjects who used a method similar to the SR definition, 39 rotated the whole figure while 43 rotated only a part of the stimulus on the Figures test. Those who used the latter approach performed significantly better than those who attempted to rotate the entire figure.

2. Those who used abstract symbols scored higher than those who attempted to relate the figure to some familiar or more concrete object.

Four subcategories of solution strategies were reported for the DAT Space test. These strategies were used with different frequencies depending on item difficulty. The categories were:

1. Subject spontaneously folded the pattern and then noted the relationships of the parts (57 subjects on the easier problems; 12 on more difficult items).

2. Subject started with the alternatives first, and then looked at the stimulus figures (17 subjects on easy problems; 20 on hard problems).

3. Subject did not fold or unfold the stimulus pattern or response figures, but looked for other cues such as angle intersections (7 subjects on easy problems; 44 on difficult problems).

4. Subject guessed (1 subject on easy problems; 8 on difficult problems).

Two distinct strategies were used on the Guilford-Zimmerman Spatial Orientation and Industrial Aptitude Spatial Orientation tests.

1. Subjects imagined themselves being reoriented with regard to the stimulus. (Only 26 subjects used this approach on the Guilford-Zimmerman Spatial Orientation test, while 58 used it on the Industrial Aptitude Spatial Orientation test).

2. Subjects mentally rotated the stimulus and response figures but did not imagine themselves being reoriented. (This method was used by 58 subjects on the Guilford-Zimmerman Spatial Orientation test, while only 26 used it on the Industrial Aptitude Spatial Orientation test.)

The major implications of these observations for this review are:

1. Subjects reported using different methods to solve the same test.

2. Within a test, the number of reported strategies increased with item difficulty.

3. More distinct strategies were reported for complex tests (e.g., DAT Space Relations) than for relatively simple tests (e.g., Figures).

4. Even on relatively simple, highly speeded tests such as Figures, subjects reported different problem solving processes.

5. An explanation of why the Guilford-Zimmerman Spatial Orientation test consistently loads strongly on the Vz factor has been offered. More subjects solved this test using a Vz strategy than an SO strategy.

6. There is a tendency to shift from a direct mental manipulation strategy to a more "analytic" strategy using particular stimulus features and logical inference as item difficulty increases.

The Meyers Studies

Two less extensive, but more intensive analyses of verbal reports of spatial test problem solving were reported by Meyers (1957, 1958). In the first study, four college students were given Hidden Blocks and a Surface Development test. They were told the answers after the time limits had expired. The experimenter then asked them to discuss among themselves how they could best improve their scores on the tests. He then left the room and later analyzed the tape of their conversation. Meyers felt that this procedure was more likely to yield an unbiased picture of how the students solved the problems than if they attempted to communicate their problem solving processes to a psychologist.

In the second study, five college students were administered three spatial tests. During the following week each participated in several hour long interviews in which items from three spatial tests similar to those they had taken on the first day were presented.

Observations on the verbal reports were similar in both studies. The following were the major results:

1. Understanding directions. A number of the students failed to understand the test directions, especially on Surface Development and Hidden Blocks. A particularly common shortcoming was the tendency to overlook a key assumption about the nature of the task, e.g., that all blocks are the same size or that the figures can be folded in only one way. Further, some had difficulty deciphering how the numerical and alphabetical symbols were to be used to codify answer choices on Surface Development. Meyers concluded: "with the large groups used in factor analysis, it is not safe to assume that all subjects are attempting to do the same thing" (1957, p. 6).

2. Understanding line drawings. Many comments concerned difficulties in "reading" the line drawings. Further, there were apparently differences in the size of perceptual units between students; some tended to "see" a block where others dealt with more molecular units such as lines or planes.

3. Strategies and difficulty. Students reported solving easy Surface Development items by using mental imagery. However, they quickly shifted to more "analytic" methods as the problems became harder. A similar observation was made by Barratt (1953) in his study.

While interesting, these observations are based on the extensive retrospections of only a few subjects. Quantitative indices were not computed, and so the conclusions represent the "overall impressions" of the investigator. Further, the data were retrospections, about which Bloom and Broder (1950) remark:

> It is very difficult for a person to remember all the steps in his thought processes and report them in the way in which they originally occurred. There is a tendency on the part of the

narrator to edit the report, to set forth the process in a nicely logical order. Things seem to tie together so nicely after the problem has been solved. The narrator will usually omit errors and "dead ends" in his thinking processes. He will not remember the queer quirks and unusual circumstances which surrounded his thinking. Such reports generally present a coherent and well ordered train of thought rather than the incoherent and jumbled process which may have occurred. These retrospective accounts are useful, but it must be recognized that they are rebuilt outlines of thought processes and tend to reveal only the high spots and finished products rather than the raw materials and details in a fantastically complex series of thought steps. (Bloom and Broder, 1950, p. 6)

The French Study

Another investigation of the relationship between problem solving styles and cognitive processes was reported by French (1965). He administered a battery of five "pure" factor tests and ten factorially complex tests to 177 male high school and college students. Students also filled out a questionnaire about their background and general approach to the test problems. They were then interviewed while they solved items similar to those in the test battery. The tetrachoric correlation matrix of the questionnaire, interview, and test variables was then factored by principal components with varimax rotation. This analysis produced 25 factors, 17 of which were considered representative of psychologically distinct problem solving styles or background characteristics. The factors were used to divide the subject pool into 17 pairs of subsamples.

An initial factoring of the 15 test correlation matrix for the entire sample indicated five factors. Tests that loaded highest on the first four of these factors were used as marker variables in the rotations of the 17 pairs of subsample factor analyses. Five factors were extracted in each of these subsample analyses. A targeted, quartimax rotation was then performed on the five factors using the four sets of marker tests. The first four of these factors were further rotated to a patterned oblimax criterion to bring the factors as close as possible to the marker tests. The fifth factor was kept orthogonal throughout these rotations.

The procedure was unnecessarily complex. A simple multiple group analysis using the marker tests to define factors would have yielded essentially the same results.

Of the seventeen pairs of factor analyses performed, only four were discussed in detail. Some of the more important findings were:

1. Division of the sample according to whether the subjects used a rule for solving Cards items produced no noticeable differences in the factor loadings or factor intercorrelations.

2. The correlation between the Space-Visualization and Verbal Comprehension factors decreased in 11 pairs of subsamples where some "systematic" approach was used for a test. However, "systematic" did not mean logical or analytic rather than intuitive or global. Instead, it appears that any reasonably well-defined method of solving problems was called "systematic." Thus, those who had well-defined strategies tended to show greater differentiation of abilities than those who had not developed such specific problem solving skills.

3. The loading of the Cubes test dropped from .52 to .07 on the Space-Visualization factor for those who used an analytic strategy to solve the items. Here, "analytic" meant a positive mark on more than one of the following: (a) geometrical terms used in solving Cubes items, (b) few visualization indications made in solving Cubes items, (c) when asked, reports mentally rotating the cube on two separate axes. Thus, "analytic" did not necessarily mean non-visual.

4. The Guilford-Zimmerman Spatial Orientation test loaded on the Reasoning factor rather than on the Space-Visualization factor for those who used "reasoning" on the test. No further details were given, so interpretation is difficult.

5. There were no important differences in factor structures between those who said they used more or less visualization, except that the Reasoning factor had lower correlations with other factors in the group reporting less visualization. This is counterintuitive, as reasoning should be more influential when visualization is not used.

French concluded that the most pervasive strategic variable was "some kind of reasoned or systematic approach as contrasted to less orderly scanning and visualizing, with reliance on common sense" (1965, p. 26). Further, he observed that the systematic approach may work differently on different tests. A systematic approach could eliminate random behavior and increase both the reliability and factorial purity of a test. This appears to be the case for the eleven contrasts in which a "systematic" approach decreased the correlation between the verbal and spatial factors, producing a sharper differentiation of abilities. On the other hand, especially on spatial tests, a systematic approach could enable a student to derive the correct answer by an entirely different set of processes than those intended by the test constructor. For such individuals, the expected factor loadings of the test would decline or vanish altogether, as they did for those who used an analytic approach to the Cubes test.

While this study suggests the type of strategy differences that may influence factor structures, it is by no means unambiguous. In particular, the difference between systematic strategies and analytic, non-visual strategies warrants more precise differentiation. Further, there is no way of knowing what factor structure differences would have been produced by 17 random splittings of the sample. Nevertheless, the study does suggest that "even simple 'pure factor' tests...do not measure the same things for all people" (French, 1965, p. 26).

Yalow and Webb

A recent investigation reported by Yalow and Webb (1977) further defines the major dimensions of reported solution strategies. Retrospective reports of solution strategy were obtained from 48 high school students on a range of verbal, spatial, and reasoning tests. Eye fixations were recorded while students solved several items from each test. Students were then presented three or four items from each test and asked to describe how they solved each item. The experimenters completed a questionnaire for each test on the basis of these responses. Questionnaires had been developed during pilot investigations with over 100 college students. Experimenters asked students additional questions only if it were not possible to fill out the questionnaire from the student's first description.

Yalow and Webb (1977) reported a preliminary analysis of 13 strategy indices computed across four tests: Vocabulary, Verbal Analogies, Paper Folding, and Paper Form Board. The score on each index was a ratio of the number of times the student reported using a particular strategy to the total number of opportunities to report that strategy. Right and wrong items were analyzed separately. The major results were:

1. High ability students usually knew the answer before looking at the alternatives, while low ability students spent more time evaluating and eliminating alternatives.

2. Low ability students reported more internal verbalization while solving tasks, guessed more frequently, and had less confidence in their answers.

3. Students of intermediate ability reported using specific spatial strategies more frequently than either high or low ability students.

With one exception, the correlations between particular strategy indices for right and wrong items were all positive. Further, the pattern of intercorrelations of the various indices were about the same for right and wrong items, except much lower for the wrong items. Closer examination of these correlations suggested that there were three major dimensions in the 13 indices. The first cluster represented the tendency to analyze the response alternatives. It was defined by the indices for Exhaustive Response Search and Response Elimination. Indices for Checking, Serial Analysis, and Specific Spatial Techniques also correlated with this cluster. The second dimension represented the tendency to construct a response from a careful analysis of the stimulus words or pictures before looking at the response alternatives. This dimension was defined by the indices called Constructive Template Match. and Knowledge of Words. Other indices that correlated with this dimension involved confidence in response, (not) guessing, ability to clearly explain how the item was solved, and lack of verbalization while solving the item.

The third dimension was bipolar, with Impressionistic Solution on one end, and Serial Analysis and Spatial Techniques on the other. This cluster is similar to French's (1965) distinction between the reasoned-systematic approach and the scanning-visualization-common sense approach. However, the present analysis also suggests that it is important to distinguish between a systematic analysis of the problem stem and response alternatives.

Other Studies

A number of other studies address the issue of different solution strategies reflecting different mental processes. Gavurin (1967) administered ten anagram problems under two conditions. In the first condition (N=13), letters could not be physically rearranged, and so subjects had to solve the anagrams "in their heads." In the second condition, another group of subjects (N=14) was allowed to physically rearrange the tiles on which letters of each anagram were printed. The correlation between anagram solving and the Minnesota Paper Form Board test was .54 in the non-manipulation condition, and -.18 in the manipulation condition. Thus, how the anagram test was administered dramatically affected the way items were solved. Unfortunately, Gavurin failed to show that manipulatory condition did not also destroy the relationship between anagram performance and verbal or general abilities. Although the sample sizes were small, it appears this may have happened.

A study by Frandsen and Holder (1969) provides further support for French's (1965) observation that having a systematic approach to a particular problem type makes a difference. They selected 18 pairs of students from a population of 146 undergraduate general psychology students. Pairs were matched on the DAT Verbal Reasoning test, but as disparate as possible on DAT Space Relations. One student from each pair was then randomly assigned to a treatment group, and the other to a control group. Those in the treatment group were taught specific diagrammatic techniques to represent syllogistic, time-rate-distance, and logical deduction problems. Venn diagrams were used to represent syllogisms. Marked lines represented time-rate-distance problems. Diagrams of the facts and conditions were adapted for the deduction problems. Only those low in spatial aptitude who had received the instruction showed significant improvement on tests containing these types of verbal problems. Those high in spatial ability were not affected by the treatment, although there were some ceiling effects for high ability students on both pretest posttest

Some New Data

Some previously unreported data address the issue of strategy differences on spatial tests. The data reported here were collected as part of the administration of a reference battery of tests to 123 Stanford undergraduates. Three spatial tests from the French Kit of Reference Tests for Cognitive Factors (French, Eckstrom and Price, 1963). were included in that battery: Paper Form Board, Paper Folding, and Surface Development. Details of the administration of these and other tests in the reference battery, along with descriptive statistics, correlations, and factor analyses are reported elsewhere (see Snow, Lohman, Marshalek, Yalow and Webb, 1977).

This analysis began with the observation that only some students made drawings or other marks on their tests. For example, on the Paper Folding test, some students drew circles on each stimulus figure that indicated where the holes would be when the paper was unfolded to that configuration. Drawings on the Paper Form Board test indicated how the stimulus pieces could be put together to make the target figure at the top of the page. Drawings on the Surface Development test were more infrequent, and usually indicated the position of one or two planes after they were folded.

Items on each test were scored for the presence of drawings or markings on the item. For Paper Folding, a distinction was made between light, pencil point marks and heavy, clear marks. On Paper Form Board, lines drawn on the target figure at the top of the page were distinguished from drawings made on or beside the item. For Surface Development, each figure was associated with five items. Thus, the item referent for each mark was uncertain.

Marking indices were computed and correlated with each other, total scores on the tests, and reference constructs. Unfortunately, the tendency to mark on one test did not generalize to other tests. Some correlations between marking variables on the three tests were significantly different from zero, but all were small. These correlations, along with the means and standard deviations for each variable, are reported in Table 34.

Insert Table 34 about here

There are several reasons for low correlations. Very few marks were made on Surface Development; in fact, the average was only .6 marks per subject, or one out of every 20 figures. Thus, there was not sufficient variance in the index to generate a correlation with other variables. Form Board, on the other hand, had an average of almost 12 marks per subject, or one out of every four items. However, the instructions for this test suggested that it might be useful to draw pictures. Thus, willingness to draw on this test probably reflects something different than the tendency to mark on test when not specifically directed to do so. Correlations between marking on Paper Form Board and reference constructs, and comparision of correlations between total score on the Form Board test and reference constructs supported this hypothesis. Therefore, only the correlations for Paper Folding are reported here.

The first three columns of Table 35 show correlations between the three Paper Folding marking indices and scores on selected reference constructs. There was a tendency for females, those low on SATQ, and those high on the CPI Good Impression scale to make light marks. Those who made heavy marks tended to score low on Film Memory III, high on the CPI Anxiety Scale, and low on the CPI Well Being Scale. For total marks on the Paper Folding Test (light plus heavy), females, those with low scores on the Visual Number Span test, and those who scored low on the Terman Concept Mastery (a verbal analogies test) tended to make more marks.

Insert Table 35 about here

Several of these correlations are particularly interesting. For example, females, who generally score lower than males on spatial tests, may have achieved scores comparable to males by solving the problems in a different way.

Table 34

Means, Standard Deviations, and Intercorrelations of Marking Indices and Scores on the Three Spatial Tests

Test	M	sd	Correlations[a]									
			1	2	3	4	5	6	7	8	9	10
1. Paper Folding Light Marks	1.72	2.63	-									
2. Paper Folding Heavy Marks	2.11	3.86	02	-								
3. Paper folding Total Marks	3.84	4.73	58	83	-							
4. Paper Folding Total Correct	14.59	3.40	-11	07	00	-						
5. Surface Dev. Total Marks	0.60	1.61	02	26	22	13	-					
6. Surface Dev. Total Correct	46.06	12.97	-19	-01	-12	63	-05	-				
7. Form Board Top Figure Marks	2.90	2.44	00	12	10	21	17	07	-			
8. Form Board Stimulus Marks	8.77	8.29	13	09	14	-04	02	-07	-14	-		
9. Form Board Total Marks	11.67	8.31	13	13	17	03	07	-05	15	96	-	
10. Form Board Total Correct	24.25	7.70	-10	05	-01	57	04	57	13	04	08	-

[a]Decimals omitted.

Table 35

Correlations between Selected Reference Variables and Paper Folding Marking Indices (N=123), and between Paper Folding Total Correct and Selected Reference Variables for Less (N=62) and More (N=61) Marking Groups

	Correlations with Numbers of Marks on Paper Folding (N=123)			Within Group Correlation with Paper Folding Total Score	
Reference Variable[a]	Light Marks	Heavy Marks	Total Marks	Tot. Marks ≤ 1 (N=62)	Tot. Marks ≥ 2 (N=61)
Sex (female=1, male=2)	-27*	-12	-25*	17	-02
Picture Completion	-10	00	-06	35*	23
Street Gestalt	-12	03	-04	36*	21
Harshman Figures	-07	07	01	46*	39*
Film Memory III	19	-23*	-09	33*	-10
Visual Number Span	-20	-17	-25*	04	26
Identical Pictures	-07	15	08	26	44*
Finding A's	-17	01	-09	-02	13
Number Comparison	-09	-17	-19	-21	32*
Paper Form Board	-10	05	-01	69*	40*
Surface Development	-19	-01	-12	72*	53*
Letter Series	-14	-02	-09	43*	51*
Terman Concept Matery	-16	-15	-21*	26	32*
Raven Matrices (Advanced)	-15	04	-05	61*	54*
SATV	-08	-05	-09	07	30
SATQ	-28*	01	-16	43*	57*
WAIS Comprehension	07	19	19	36*	19
WAIS Arithmetic	-16	-01	-10	21	30*
WAIS Digit Span	-18	04	-06	19	25
WAIS Digit Symbol	06	-13	-07	18	27
WAIS Block Design	-12	00	-07	52*	47*
WAIS Object Assembly	-03	02	00	44*	17
Embedded Figures (Errors)	01	-17	-13	-50*	-16
Matching Familiar Figures (Errors)	17	10	19	-53*	-35*
Marks Imagery Questionnaire	20	-07	05	-31*	-14
Marks Picture Memory Test	00	-05	-04	35*	-01
Conry Picture Memory Test	-01	-05	-04	35*	09
Factor Scores					
Gfv	-17	04	-06	73*	70*
Gc	-07	-12	-14	17	22
Perceptual Speed	00	-08	-07	-02	34*
Number	-13	-05	-03	-04	20
Picture Memory	04	-16	-11	26	-24
Memory Span	-09	-10	-14	07	04
Closure Speed	02	01	02	33*	18
California Psychological Inventory					
Anxiety	00	22*	18	09	08
Well Being	17	-23*	-09	-01	06
Good Impression	21*	-03	09	-01	04
Femininity	17	04	13	-17	07

Note. Decimals omitted.

[a]See Snow et al. (1977).

*p less than .01.

Thus, while there was no sex difference in mean scores on this test, females may have done more work on paper and less "in their heads."

It is also reasonable that those who made heavy, detailed markings on the test tended to score higher on the CPI Anxiety scale. Further, it was not just marking in general that mattered here, for the number of light marks did not correlate with anxiety.

The negative correlations between marking and Film Memory III, and marking and Visual Number Span suggests that some of those who marked on the test used the drawings to compensate for poor visual memory.

The last two columns in Table 35 report within group correlations between total score on Paper Folding and the reference variables. The two groups were formed by a median split of the total marking index. Those in the first group (N=62) made one or no marks on the test while those in the second group made two or more marks (N=61). Comparisons were also made between those who made some versus no light marks, and some versus no heavy marks. However, the results were essentially the same.

The major differences between these two sets of correlations were that, for those who made only one or no marks, Paper Folding correlated higher with the other two spatial tests (Paper Form Board and Surface Development), errors on the Matching Familiar Figures test, Marks Picture Memory test, the Conry Picture Memory test, the Picture Memory factor score, the Closure Speed tests, the Closure Speed factor score Film Memory III, and errors on the Embedded Figures test. For the Marks Vividness of Visual Imagery Questionnaire, the correlation was more strongly negative in the low marking group.

For those who made two or more marks, Paper Folding correlated higher with the Perceptual Speed tests (particularly with Identical Pictures, Number Comparison, Digit Symbol, and their factor score), Letter Series, SATQ, and Digit Span Backwards.

Together, these correlations suggest that Paper Folding was more of a spatial test for those who did not mark. For those who did mark, it became slightly more of a Gf test with Perceptual Speed playing a decisive role. Further, there was a hint of a male advantage in the no marking group, but no such difference in the marking group. The stronger negative correlation between Paper Folding and the Marks Vividness of Visual Imagery Questionnaire in the no marking group supports the hypothesis that spatial ability depends on the control the subject can exercise over his image, and not necessarily on the vividness of the image. Further, as suggested here, extremely vivid visual imagery may actually inhibit spatial thinking.

Conclusions

1. There are important differences in solution strategy both between subjects and within subjects over items. Tests often measure different abilities for different students, depending on how problems are solved.

2. Complex, power tests elicit a wider range of alternative solution strategies than simple, highly speeded tests. Vz tests are often solved in more ways than SR tests.

3. Within a test, the more difficult items elicit a wider range of solution strategies than easy items.

4. High ability students report studying the problem stem and constructing an answer before examining the alternatives. They are usually able to give a coherent verbal report of how they solved the item, and they express confidence in their answers. Low ability students, on the other hand, frequently report that they attempt to solve the item by analyzing the alternatives. Further, they report more internal verbalization, more guessing, and less confidence in their answers than do high ability students.

5. Certain tests are particularly susceptible to alternative solution strategies. For example, many Spatial Orientation tests can be solved by a Visualization strategy. On a more general level, multiple choice paper and pencil tests permit a number of alternative solution strategies that are not possible when the student must construct rather than select an answer. Students can also draw or mark on the test, thereby reducing the need to remember more than a single step in the solution of the problem. They can attempt to solve the problem by "working backwards" from the alternatives to the stem, or look for clues in the alternatives that may reveal the correct answer or simply narrow the field. Therefore, a range of alternative solution strategies could be eliminated by using free response rather than multiple choice items. At the very least, alternatives should not be visible when the problem stem is presented.

6. Introspective reports are of limited value. Whenever possible, such reports should be validated against external information. Many processes, especially those that are extremely rapid, cannot be accessed through introspection (see Nisbett and Wilson, 1977). Retrospective reports are even less trustworthy. Such reports are best used as a rough index of strategy rather than as a guide to mental processes. Detailed retrospections are probably quite unreliable. Thus, subjects could be expected to indicate

whether they mentally rotated an object or, instead, mentally projected themselves into the picture. It is unlikely, however, that they would be able to decompose this global behavior into component processes accurately.

7. Perhaps the most promising technique for obtaining valid introspective evidence is to ask subjects to report specific strategy information immediately before (Karpf and Levine, 1971), during (Kroll and Kellicutt, 1972), or after (Paivio and Yuille,1969) they solve an item, usually by anonymously pressing a button. The validity of the self report rises dramatically, although reactive effects might present problems.

8. Individual differences in the ways students solve tests challenge a basic assumption of factor analysis. Factor structures obtained from analyses of such tests may be severely distorted. The most likely outcome is an overestimation of the factorial complexity of a test. Thus, that some SO tests load on both Vz and SO factors may only mean that students solve the tests differently: some using a predominately SO strategy, while others rely on a Vz strategy. Alternately, students may switch between these two strategies while solving different items. However, even in this straightforward example, it is impossible to know whether the test measures two different aptitudes in any one individual, or whether it measures different aptitudes in different individuals. On a more general level, the presence of several tests in a battery that are amenable to alternate solution strategies seriously distorts the factor structure, so that the obtained factor structure may not apply to anyone in the sample. Factoring within strategy groups would undoubtedly produce cleaner factor patterns. There were some indications of this in the within sex analysis of Michael et al. (1951) (see p. 89), and in the finding that Paper Folding correlated higher with the other spatial tests when students did not make pencil marks on the test itself (see p. 148).

9. Individual differences in solution strategy present a major stumbling block for both correlational and experimental investigations of spatial ability. The challenge for future research is to devise experiments that reveal solution strategy for each subject on each item or on each item-type. Only by knowing how subjects solve items can the investigator know what the task measures, or evaluate the generalizability of the processing models that are proposed to describe task performance.

SPEED AND LEVEL

The observation that tests defining the two major spatial factors (Spatial Relations and Visualization) differ in speededness and complexity first surfaced in the reanalyses of Thurstone's (1938) PMA data (see p.29). This speed-level or complexity dimension reappeared in every large correlation matrix examined in this review. Simple, speeded tests usually had low correlations with other tests and fell at the periphery of multidimensional scaling representations or in the lower branches of hierarchial factor models. Factors defined by such tests frequently disappeared when tests were made less obviously similar (e.g., Hoffman et al., 1968 and p.97). This was observed in verbal and numerical tests as well as spatial tests. For example, the highly speeded Flags test defined the Space factor, while more complex tests such as Punched Holes had high loadings on the General factor in Thurstone (1938). Similarly, the Numerical factor was defined by speeded computation tests in Thurstone (1938), while complex arithmetic achievement tests defined the Gf factor in Snow et al. (1977). For memory tests, WAIS Information helped define Gc while the memory span tests were more peripheral in Snow et al. (1977). Finally, verbal fluency and reading speed measures were frequently quite peripheral, while vocabulary and verbal reasoning tests often defined Gc (Snow et al., 1977; Hoffman et al., 1968). Thus, speed-level and complexity differences are pervasive in the factor analytic literature on human abilities.

But these speed level differences suggest that the factor structure of a test may be altered by changing its speededness or the average complexity of items in the test. There was some indication of this in the first section of this review. For example, a difficult form board test helped define the Visualization factor in the reanalyses of the PMA data (see p. 10), a slightly easier form board test fell in the middle of the SR-Vz continuum in Swineford and Holzinger (1942), while a simple, highly speeded form board test helped define the Perceptual Speed factor in Guilford et al. (1952) (see p. 48 and also Zimmerman, 1954). Accordingly, the first porpose of this section is to examine the literature on the effects of altering test speededness or complexity on the factor structure of a test.

The second purpose is to examine the relationship between individual differences in speed and level, particularly on spatial tasks. If speeded tests define different factors than level tests, then individual differences in speed should be at least partially independent of individual differences in level.

The third purpose is to examine the evidence for the existence of general and specific speed factors. Reanalysis of the Horn and Cattell (1966) data suggested that their General Speed or Retrieval Efficiency factor was more a Writing or Motor Speed factor (see p. 125). Lord (1955) also claims to have identified a general speed factor, while others claim to have isolated more specific speed factors such as Speed of Reasoning and Spatial Speed (Davidson and Carroll, 1945; Lord, 1956). Specifically, this section examines the relationship between speed and level factors, the validity of the speed scores that define speed factors, and the evidence for specific and general speed factors.

Finally, implications of these studies for research on aptitudes are presented and discussed.

Speed and Level Defined

Level and power both refer to the maximum level of difficulty of items a person can solve. Although the terms are used interchangeably, level is probably the better term as it connotes less than "power."

Speed refers to the maximum speed a person can perform an operation or solve a task correctly. Rates is the reciprocal of speed.

Experimental Methods

Studies of the relationship between speed and level have followed several paradigms. Early experiments avoided specific speed-accuracy instructions hoping to induce subjects to perform at their "natural" rate (e.g., Hunsicker, 1925). This search for a measure of "natural rate" later became the search for a general speed factor (Horn and Cattell, 1966; Lord, 1956).

The second type of study sought to determine whether individual differences in speed and level both contribute to performance on time limit tasks. A variety of methods were employed in these studies. Some correlated correctness on a time limit test with correctness on the same test with no time limits (May, 1921; Ruch and Koerth

1923; Yates, 1966a, 1966b). Others correlated correctness with the time taken to finish the test (Baxter, 1941; Freeman, 1923), factored correlation matrices containing both level and speed estimates for each test (Davidson and Carroll, 1945; Lord, 1956; Myers, 1952), or regressed speed and level scores on time limit performance (Davidson and Carroll, 1945). Some administered the same test under different time limits (e.g., Davidson and Carroll, 1945) while others kept time limits relatively constant and varied the number of items in each test (e.g., Lord, 1956). Finally, a few studies examined the relationship between speed of performing simple tasks, and level scores on more complex items of the same type (Egan, 1976; Hunsicker, 1925; Tate, 1948).

Studies have employed equally diverse measures of speed, such as number correct on a time limit test (May, 1921; Myers, 1952), total time taken to finish the test (Baxter, 1941; Freeman, 1923), average time for correct items (Egan, 1976; Tate, 1948) and last item attempted on the test (Myers, 1952; Lord, 1956). Further, studies that used time to estimate speed frequently transformed raw time to log time (Tate, 1948; Furneaux, 1961), the reciprocal of time (Davidson and Carroll, 1945) or used some unspecified normalization (Lord, 1956).

Estimating Test Speededness

Several methods have been used to estimate test speededness. The simplest methods define speededness as the number of items presented or solved per unit time. With parallel tests, the more speeded test requires the completion of more items per unit time than the less speeded test (e.g., Lord, 1956; Myers, 1952). But this index does not reveal whether differently speeded tests require different abilities.

Such an index was proposed by Cronbach and Warrington (1951). The index called τ, shows what proportion of reliable variance in a time limit test is independent of the reliable variance in the same test when administered with no time limit. Formally:

$$\tau = 1 - \frac{r_{A_t B_u} \cdot r_{A_u B_t}}{r_{A_t B_t} \cdot r_{A_u B_u}}$$

where A and B are parallel forms of a test, and the subscripts t and u refer to timed and untimed conditions. A more general formula would obtain if the subscripts were changed to t_1 and t_2, referring to any two different time limits.

However, items are roughly ordered from easy to difficult on many tests. Therefore reducing the time limit for a test also reduces the average difficulty of the items that are attempted or solved. In a paced administration, all items are presented, but again more easy items than difficult items are solved when presentation time is short. While a value of τ greater than 0 indicates that the timed test measures something different than the untimed test, the unique variance in the speeded test is not necessarily due to speed of response. The inference is justified only when exactly parallel items are attempted and correctly solved under both conditions. Solution latency provides a better estimate of speed.

Part-Whole Correlation Studies

The first research strategy is exemplified in studies where scores on a time limit test were correlated with scores on the same test after an extended period of time. Studies by May (1921) and Ruch and Koerth (1923) are the most frequently cited examples of this procedure. Spearman (1927) felt these studies supported his hypothesis that speed and power are interchangeable.

In the former (May, 1921), the Army Alpha was administered to 510 army recruits with the usual time limit. They were then given different colored pencils and allowed to work on the test for the same amount of time again. The correlation between regular and double time scores was .97.

Ruch and Koerth (1923) repeated the experiment with college freshman. They selected 72 students who scored in the lowest ten percent and 52 who scored in the higest ten percent on a college entrance test. Students were then administered the Alpha under the usual time constraints; then, as above, with double time allotted; and finally, with unlimited time. The correlation between the usual and double time scores was .97 , while that between the usual and unlimited time score was .94 . On the basis of these high correlations, the investigators concluded that speed of response was not an independent factor of theoretical or practical interest.

But this conclusion overlooks several important characteristics of both studies:

1. The high correlation is in large part a reflection of the part-whole relationship of each subject's scores under the various conditions. Using parallel forms of the test for each condition would yield lower correlations.

2. The degree of speededness of the test is a function of its time limit. Thus, under the "usual time limits," the Army Alpha may be (on the average) primarily a power test.

3. Number correct is not a suitable measure of speed or rate. Total correct on a time limit test accurately reflects average time per item only when all items are of equal difficulty, and there are no errors.

4. The high correlations in the Ruch and Koerth (1923) study are in part a function of the extreme groups design. The high correlation in May (1921) reflects the extremely wide range of scores in the army samples.

Similar limitations, particularly the part-whole relationship of the scores, apply to many other early studies (e.g., Walters, 1927; Ruch, 1924). However, this procedure is often justified, even though it does not illuminate the speed-power issue (see Cronbach and Warrington, 1951). For example, Yates (1963, 1966a, 1966b) has shown that some slow working students are severely penalized on the Raven Progressive Matrices (Raven, 1947) when the test is administered with the usual 40 minute time limit. This suggests that a longer time limit would enhance the construct validity of this power test. It also suggests that speed and level are at least partially independent aspects of performance in some individuals.

Correlating Correctness with Time to Finish

The second category of studies investigated the problem by correlating the score on a test with the time taken to finish the test. Freeman (1923) correlated scores on two examinations with the time taken to finish the tests. Correlations were -.13 and -.12. However, students were not motivated to complete the tests quickly, and so factors such as perseverance, neatness, anxiety, or subsequent commitments also determined time taken to finish the exam.

Baxter (1941) reported a similar study. He gave the Self-Administering Otis to 100 college sophomores. Students were instructed to work for both speed and accuracy, and not to go back over items previously attempted. They were given a different colored pencil at the end of 20 minutes and told to complete the entire 75 items. The experimenter recorded the time taken by each student to complete the entire test. The correlation between this speed measure and the power score was -.06. The speed estimate correlated .75 with the time limit score while the power score correlated .62 with the time limit score. Since speed and power were nearly independent, total contributions to the time limit score could be determined by simply squaring and summing the correlations. Thus, speed accounted for 56 percent of the variance in the time limit score and power 38 percent. Together, the two scores accounted for 94 percent of the variance in the time limit score.

As often happens (e.g., Myers, 1952; Lord, 1956), the time limit score had slightly higher external validity than either the speed or power score , even though all had comparable parallel forms stability coefficients (range .63 - .70 after one month). This is probably because most real life situations do not provide unlimited time, nor do they depend solely on rapid performance. It also suggests that speed and power make independent contributions to the total prediction. Thus the time limit test, which is usually an unknown mixture of the two, predicts best.

Even though the Baxter study represents an improvement over studies like that of Freeman (1923), the dependent measure is still inadequate. Time for right answers, wrong answers, double checks, and guesses are all included in the time score. Perseverance also exercises an unknown influence on the scores.

One of the more carefully conducted earlier studies was that of Hunsicker (1925). She employed a variant of this paradigm, and correlated the time required to complete a sample of easy items with the maximum level attained on a time limit test.

A six level sentence completion test and a five level arithmetic problems test were administered to four student samples (N = 28 to 54). Two samples were junior high school students and two were college student volunteers. Students were tested individually, and the time required to complete the first (easiest) level of both problem sets was recorded by the experimenter. These items were assumed to be of "no difficulty" and to "provide a speed or rate test of rare purity" (p. 16).

Most of the items at this level were easy, however data on the number of students failing each item was not reported. Some were more difficult, such as "How many ounces make a quarter pound?" and "The first .... after June is .... ." Clearly, the items were not of "zero difficulty."

Power was defined as the highest level at which the student solved 50 percent of the items correctly, with those levels below showing a higher success rate and those above a lower success rate. Scores for students who did not fit the system were adjusted by a complicated algorithm.

Median raw and disattenuated correlations between rate and level scores for the two tests are shown in Table 36. Stepped up split half reliability coefficients are entered in the diagonal of the matrix.

Insert Table 36 about here

Table 36

Median Raw and Disattenuated Correlations
Between Speed and Level
(After Hunsicker, 1925)

Test	1	2	3	4
Sentence Completion				
1. Speed	78	44	53	[a]
2. Level	45	86	[a]	56
Arithmetic Problems				
3. Speed	65	[a]	81	45
4. Level	[a]	70	38	82

Note. Raw correlations below diagonal, reliabilities on diagonal, and dissattenuated correlations above diagonal.

[a]Not reported.

The correlations of greatest interest are those between rate and level. They indicate that, within each task, rate and level shared approximately 17 percent of their total variance or 24 percent of their true variance. Across tasks, the rate measures shared 42 percent of their true variance, while the level measures shared 49 percent of their true variance.

It is difficult to generalize these results, however, because students were not encouraged to perform rapidly on the speed parts of the test. The experimenter wanted a measure of each student's "natural rate" and so instructions were carefully worded to avoid encouraging either haste or persistence. Further, the "zero difficulty" items were actually of moderate difficulty, especially for the younger students, although there were no differential correlation patterns between age groups.

Davidson and Carroll (1945) reported a factor analytic investigation within this paradigm. They administered a battery of verbal, reasoning, arithmetic computation, perceptual speed, and reading speed tests to 91 undergraduate psychology students. Most tests were subtests of the Revised Alpha Examination. Speed scores were defined as the time taken to work from the beginning to the end of the test, attempting every item once. Speed scores for four tests were converted to reciprocals. Level scores were defined as the number of items correctly answered when the student was allowed to take all the time he desired to try every item and check his work. Time-limit scores were defined as the number of items answered correctly within a prescribed time limit. All scores were grouped in ten or fewer class intervals before the correlations were computed.

With a few exceptions, these three scores were obtained for all the tests in the battery. Level and time limit scores for three tests were eliminated due to ceiling effects. The usual time-limit score on the perceptual speed test (Scattered X's) was dropped because it did not correlate with other tests in the battery, even the speed scores.

The remaining level and speed scores were then factored by the centroid method. Six factors were extracted and then time limit scores were projected into this factor space. Factors were then rotated to oblique simple structure. Only four factors were labeled: Speed of Computation, Level of Reasoning, Speed of Reasoning, and General Speed. The level and speed of reasoning factors were negatively correlated ($r = -.42$).

Time limit scores related differently to the speed and power factors, some more to speed and others to power factors. This was most evident in a multiple regression analysis that was also performed and reproduced here in Table 37. The relative contributions of speed and power to the time limit scores varied markedly across tests. The contribution of speed was greatest in verbal and addition tests, while level predominated in the reasoning tests.

Insert Table 37 about here

Reanalysis of the Davidson and Carroll data

Principal components were extracted from the 19 variable correlation matrix of 11 speed scores and 8 level scores. Time limit scores were not included because their intercorrelations were not reported, and because level and time limit scores were experimentally dependent. The five components with eigenvalues greater than or equal to 1.0 were retained and rotated to a varimax criterion. The results are shown in Table 38.

Insert Table 38 and Figure 23 about here

The first component was similar to Davidson and Carroll's (1945) General Speed factor. Here it appeared to be more of a Verbal-Reading Speed factor. The General Speed label is inappropriate since spatial and figural tests were *not included* in the battery. Further, the Perceptual Speed test that was included in the battery failed to correlate with these speed scores.

The second, third, and fourth components were similar to the Davidson and Carroll (1945) Level of (Verbal) Reasoning, Speed of Computation, and Speed of reasoning factors, respectively. The fifth component is similar to their unlabeled factor E. Here it appeared to be more of a singleton defined by Phrase Completion.

Several nonmetric multidimensional scalings were also performed on this matrix using the KYST program (Kruskal et al., 1973). Initial configurations were generated by the metric Young-Torgersen procedure. Nonmetric configurations were then iterated 22 times in three dimensions and 14 times in two dimensions. Stress values (Formula 1) were .119 and .198 in three and two dimensions, respectively. The final two dimensional configuration is shown in Figure 23. The clusters in Figure 23 were generated by the BMDP average method hierarchical clustering program (Dixon, 1975).

Table 37

Zero-order Correlations, Beta Coefficients, and Multiple Correlations in the Prediction of Time Limit Scores (T) from Speed (S) and Level (L) Scores (After Davidson & Carroll, 1945)

	Zero Order Correlations			Relative Contributions		
Test	r_{TS}	r_{TL}	r_{SL}	Speed $\beta_{TS \cdot L}$	Level $\beta_{TL \cdot S}$	$R_{T \cdot SL}$
Alpha Examination						
Addition	81	34	22	77	17	83
Arith. Reasoning (2 min.)	48	80	44	16	72	81
Arith. Reasoning (4 min.)	32	71	44	01	71	71
Common Sense	83	30	19	80	15	84
Same Opposites	73	62	33	60	42	84
Disarranged Sentences	79	52	31	70	30	84
Number Series	67	77	48	39	58	84
Verbal Analogies	83	39	34	79	12	84
Directions	65	57	42	50	36	72
Disarranged Morphemes	60	78	56	23	65	80
Letter Grouping	45	73	14	36	68	81

Note. Decimals omitted.

Table 38

Rotated Principal Components for the Davidson & Carroll (1945) Data

Test	Component I	II	III	IV	V	h^2
Speed Scores:						
5 Addition			80			66
6 Arith. Reasoning			40	63		59
7 Common Sense	78					75
8 Same-Opposite	74					62
9 Disarranged Sentences	66	31				57
10 Number Series	35	25	65			64
11 Verbal Analogies	64	27	29			61
12 Directions	49	46	42			67
13 Disarr. Morphemes	44	36		56		68
14 Letter Grouping				74		62
35 Reading Speed	74					63
Level Scores:						
16 Arith. Reasoning		57	34	40		62
18 Same-Opposite	54				49	60
20 Number Series		80	32			78
21 Verbal Analogies	34	46		-29		45
22 Directions		81				71
23 Disarr. Morphemes	30	69			27	73
24 Phrase Completion					86	76
25 Letter Grouping		50			51	54
Total Variance	3.79	3.24	1.92	1.80	1.46	

Note. Decimals omitted for factor loadings.

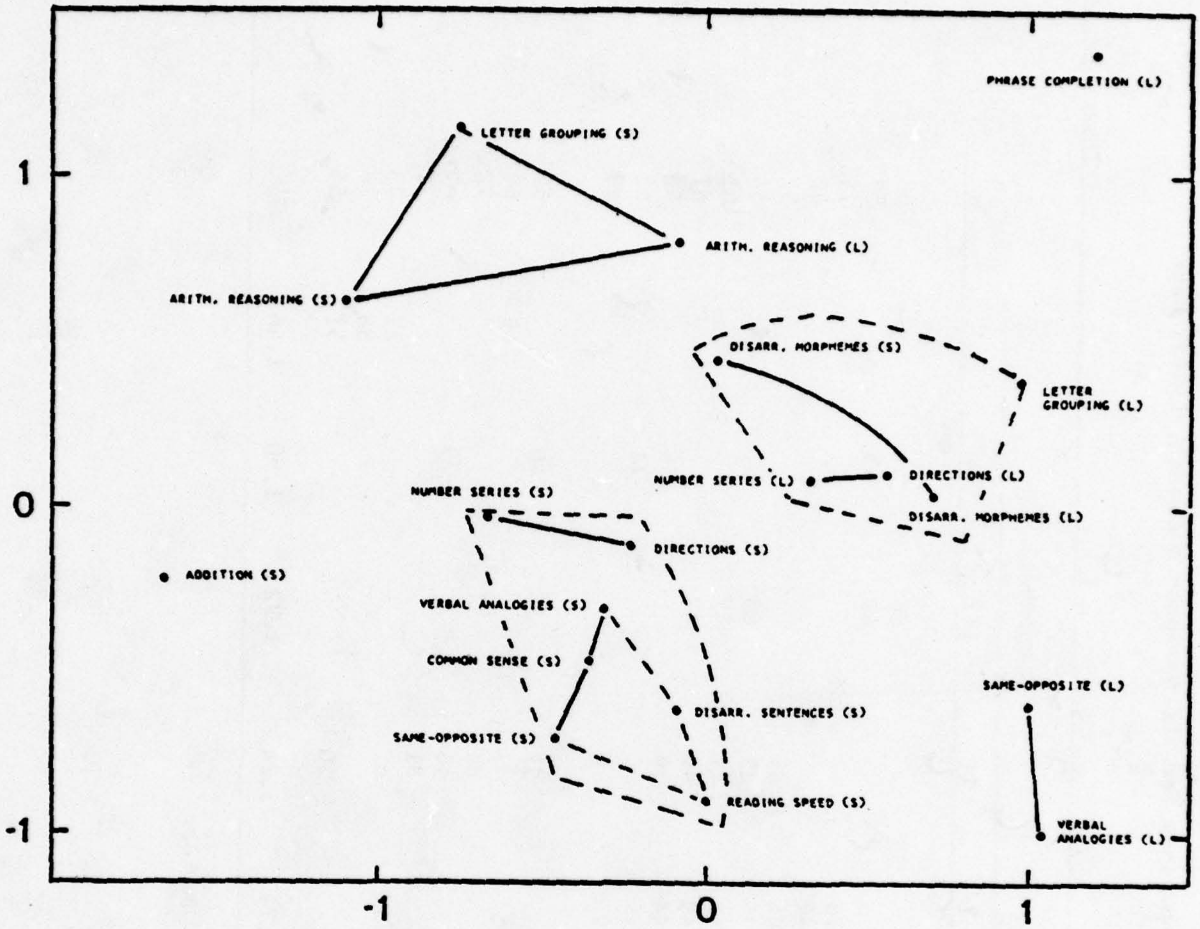


Figure 23. Average method clusters superimposed on two dimensional scaling of the Davidson & Carroll (1945) data.

Speed and level scores fell on opposite sides of the scaling in Figure 23, suggesting that speed and level are at least partially independent aspects of performance. However, the speed scores were obviously inadequate, and the location of verbal analogies, which is usually a good measure of general verbal ability, suggests that there may have been ceiling effects in some of the level scores that were retained.

In spite of these limitations, the study does suggest that speed and level may be highly correlated in the verbal domain, yet still reliably independent (see also Morrison, 1960).

This makes sense especially on vocabulary tests. Subjects who know the meaning of a word should be able to identify the appropriate synonym quickly. With slightly more time they may be able to use other cues such as root derivations, or the like. However, additional time beyond this should be of little value. Either one knows the stimulus word and recognizes one of the alternatives or not. Within the group of those who know the answer there are undoubtedly individual differences in the speed of accessing and comparing word meanings but such differences would be in the order of milliseconds, and thus would not be reflected in the dependent measure employed in this type of study.

However, the work of Hunt and his colleagues suggests that individual differences in the speed of these operations are also related to the differences between medium and high verbal subjects (Hunt, Frost and Lunneborg, 1973; Hunt, Lunnebord and Lewis, 1975). Perhaps such differences would be more strongly related to fluent production, especially within sex (see Bock, 1973).

Items Attempted as Speed

The third type of research strategy employed the number of items attempted within a given time period to estimate speed. Myers (1952) reported one of the better studies of this sort. He administered three forms of a figure classification test to 600 midshipmen at the U.S. Naval Academy. The 100 figure classification items were presented on ten pages with ten items per page. These pages were divided into five 12 minute parts with either one, two or three pages to a part. The three forms differed only in the grouping of the pages into the parts of various lengths.

In the power tests (10 items in 12 minutes), 97 to 100 percent of the examinees completed the items. In the speed tests (30 items in 12 minutes), only 30 to 41 percent of the examinees responded to all items. Scores on the first page of each of the five parts were used to define one factor. Scores on the last page of the two three-page parts together with the number of items attempted on these parts were used to define a second factor. These two factors were then extracted from the correlation matrix of test and criterion scores. The first factor was defined as the ablity to answer problems correctly, and the second factor as the tendency to answer problems quickly. Parallel analyses were made for all three forms and all exhibited similar structures. But the speed factor is spurious, because the scores used to define it are experimentally dependent. It is likely that those who answered more items correctly on the third page would be those who attempted more items. It is difficult to see how one would explain the data if this were not the case. Unfortunately, correlations were not reported, and so reanalysis is impossible.

Another factor analytic investigation of the effect of test speededness was reported by Lord (1956). A battery of nine reference tests and 18 experimental tests was administered to 649 Naval Academy cadets. Grades for five classes and conduct ratings were also available. The experimental tests represented three content areas; vocabulary, spatial ability,and arithmetic reasoning. Two tests in each area were relatively unspeeded level tests, one was moderately speeded, and three were highly speeded. Tests were parallel in content, and differed primarily in the number of items, although time limits tended to be shorter for the speed tests. Two estimates of speed were obtained: number correct on each of the three speed tests in each area, and the last item attempted on one of the speed tests. These experimentally dependent scores were excluded during factor extraction but included for rotation of axes.

Ten factors were extracted from the 33 variable correlation matrix by the miximum likelihood method. Factor loadings for the six experimentally dependent variables were then estimated by projecting these variables into the factor space. Axes were then rotated to achieve "psychologically meaningful oblique factors" (p. 42). Level and speed factors were identified in each area, except arithmetic reasoning. Speed factor correlations suggested a second order general speed factor.

Inspection of the correlation matrix reveals no general speed factor. Table 39 reports average correlations between level, speed, and last item attempted for the experimental tests and the six reference tests and factors. Correlations for the moderately speeded test in each area were omitted for clarity.

Insert Table 39 about here

Table 39 reveals that the correlation between each of the two experimental level tests in each area was the same as the average correlation between the

Table 39

Average Speed, Level, and Reference Test Correlations
(After Lord, 1956)

Test[a]		Vocabulary			Spatial Relations			Arith. Reason.			Reference			LIA		
		R	L	S	R	L	S	R	L	S	N	Ps	F	SR	V	AR
Vocabulary																
2	Reference	-														
3-4	Level	72	(67)													
6-8	Speed	69	66	(80)												
Spatial Relations																
10	Reference	18	13	15	-											
11-12	Level	15	13	11	51	(72)										
14-16	Speed	14	12	13	54	72	81									
Arithmetic Reasoning																
18	Reference	41	30	33	35	28	27	-								
19-20	Level	33	29	28	28	32	31	58	(54)							
22-24	Speed	34	28	36	31	28	32	61	54	(63)						
Other Reference																
26-27	Number (N)	05	03	18	04	-04	-02	31	21	34	(65)					
28-30	Percept. Speed (Ps)	04	03	19	10	09	12	16	08	17	36	(40)				
1	Word Fluency (F)	26	23	31	06	11	10	18	16	19	23	19	-			
Last Item Attempted (LIA)																
17	Spatial Relations	06	07	13	24	26	44[b]	13	10	17	09	14	12	-		
9	Vocabulary	39	33	57[b]	08	04	11	29	18	30	35	31	31	29	-	
25	Arith. Reasoning	20	15	27	17	12	19	36	22	37[b]	31	24	23	39	46	-

Note. Decimals omitted. Entries in parentheses are average within group correlations.

[a]Test identification numbers from Lord (1956). Dashed entries are inclusive (e.g., 6-8 means tests 6,7, and 8 were averaged).

[b]Experimentally dependent correlation.

two level and three speed tests in that area. Thus, the level tests correlated as highly with the speed tests as with each other.

Further, the patterns of correlations between the experimental level and speed tests and other tests in the battery were virtually identical. Therefore, there is no evidence that the speed and level tests defined different factors. The few discrepant correlations may be explained by the extra length of the speed tests (5, 3, and 3times longer for Vocabulary, Intersections, and arithmetic Reasoning, respectively), and possible ceiling effects in the level tests. Precise estimates of these effects cannot be made since means and standard deviations were not reported.

While the experimental speed and level tests do not define different factors, the last item attempted scores appear to define separate factors. Further, these three scores correlated higher with each other than with other variables in the matrix.

This independence of the last item attempted scores clearly evident in the intercorrelations of the five spatial scores in Table 40. Here the two level scores were averaged as before, but this time only two speed scores were averaged. This was done to keep the last item attempted score independent of the speed score.

Insert Table 40 and Figure 24 about here

Principal components were extracted from the matrix in Table 40 and then plotted in Figure 24a. Tests are also plotted in the factor space defined by the two spatial factors in the Lord (1956) solution in Figure 24b. Both plots show that it is the last-item-attempted score that defines the speed factor.

But the last item attempted on a test is not a good estimate of speed. Both low and high scoring students may attempt many items, but for different reasons. Test taking strategy and perseverence also influence the number of items attempted. Thus, correlations between the three last-item-attempted scores reflect more about the consistency of test taking strategies than about intertask consistency in the speed of mental operations.

The slight link between the last-item-attempted score and the most speeded spatial test (see Figure 24) suggests that severely short time limits are necessary to alter the factor structure of a test when the dependent variable is correctness. Cronbach and Warrington (1951) reached the same conclusion in their reanalyses of the Tate (1948) data.

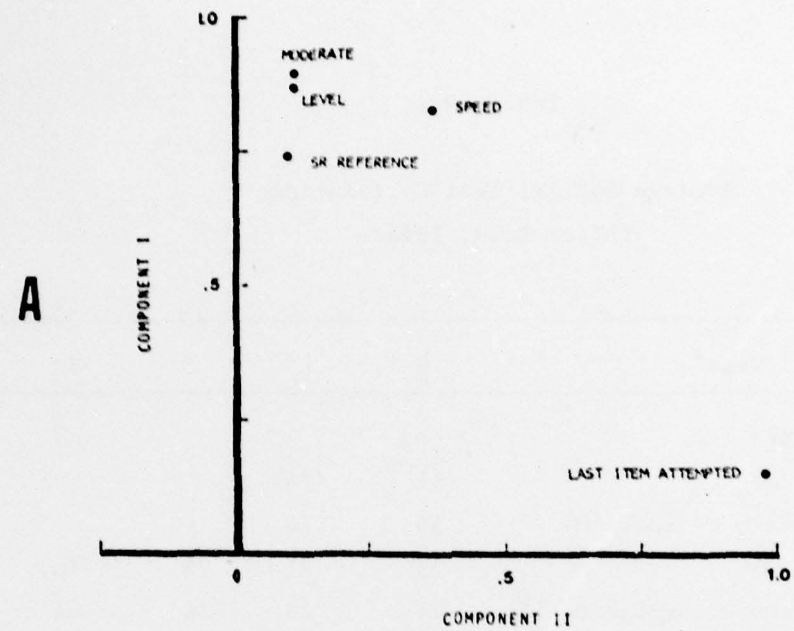
Table 40

Average Spatial Test Correlations
(After Lord, 1956)

	Spatial Test[a]	R	L	M	S	LIA
10	Reference (R)	-				
11, 12	Level (L)	51	(74)			
13	Moderately Speeded (M)	55	74	-		
14, 16	Speed (S)	55	71	76	(76)	
17	Last Item Attempted on 15	24	26	26	44	-

Note. Decimals omitted. Entries in parentheses are correlations between the two tests in the group.

[a]Test identification numbers from Lord (1956)



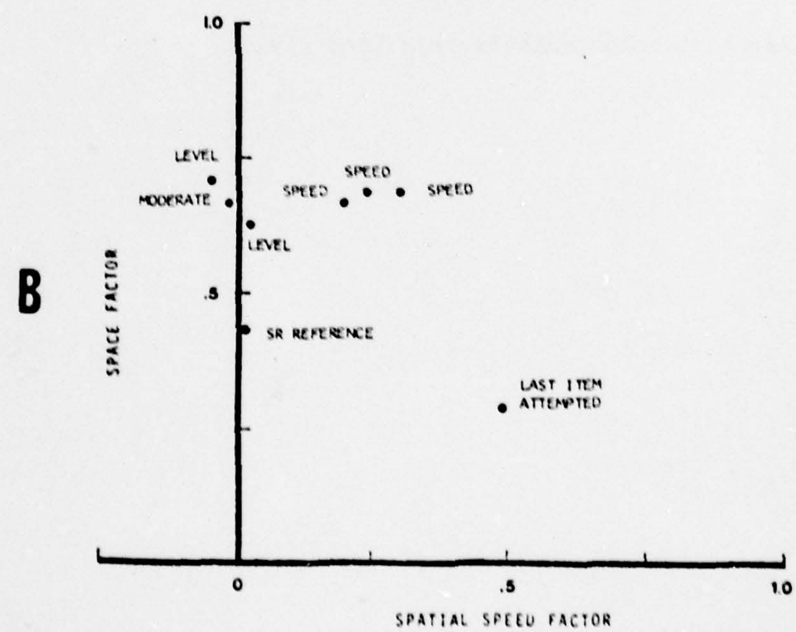


Figure 24. Plot of spatial tests for (A) the principal components reanalysis and (B) two spatial factors in Lord's (1956) solution.

Morrison (1960) obtained a similar result in his study. He examined the effects of altering the time limits on two types of reasoning tests. Five number series tests and five parallel letter matrices tests were administered to 81 undergraduate males. Two levels of item complexity (low, high), two methods of time limiting (paced, timed), three levels of time allottment (short, moderate, untimed), and two types of reasoning tests (number series, letter matrices) were varied over two testing sessions. Only nine of the 48 possible combinations of these factors were represented in the 10 tests actually administered. Each test contained 20 items.

Correctness and errors were obtained for each test, but solution time was recorded only for the three untimed tests. Speededness was estimated by computing τ for each time limit test (see Cronbach and Warrington, 1951). The factor structure of time-limit test scores was examined by a square root factor analysis of the correlation matrix for correctness and error scores for all tests, and time scores for the three untimed tests.

The major results of Morrison's study were:

1. Correctness and errors on untimed tests were largely independent of time taken to finish these tests.

2. Right and Wrong scores were differentially affected by variations in test speededness.

3. Practice increased the proportion of reliable variance in timed letter matrices tests attributable to the speed factor.

4. Pacing produced more reliable variance in correctness and error scores than either the time limit or no-time-limit conditions. Further, pacing produced higher speededness than an equivalent total time allowance.

5. Timed and untimed tests had different factor structures. In particular, only 34 percent of the variance in the paced tests and 47 percent of the variance in the time limit tests was accounted for by the power factor.

Tests administered under the same limiting condition were grouped for the square root factor analysis (or, more properly component analysis) making the analysis more confirmatory or procrustean than exploratory. This procedure overlooked several important results.

These are evident in multidimensional scalings of the correctness and error matrices reported in Figures 25a and 25b, respectively. Peak clusters from average method cluster analyses are superimposed on each plot. Principal components were also extracted from the correlation matrices, but the rotated components were essentially the same as the clusters.

Insert Figure 25 about here

Tests in Figure 25 are identified by the following mnemonic: score (R = right, W = wrong, T = time), content (M = matrices, S = series), complexity (H = high, L = low), and timing condition (P15 = paced 15 sec, P30 = paced 30 sec, T1 = time limit --first administration, T2 = time limit--second administration, U = untimed). Small circles identify the location of each test; numbers within the circles identify order of administration. Speededness estimates for the timed tests are shown in parentheses.

Scalings were produced by the KYST program (Kruskal et al., 1973). Initial configurations were generated by the metric Young-Torgersen procedure. Non-metric configurations were then iterated 13 times in two dimensions for both the correctness and error matrices. Final stress values (formula 1) were .0687 and .064 in three dimensions, and .119 and .111 in two dimensions for correctness and error, respectively. Error scores were reflected to obtain positive manifold, time scores were not reflected to preserve positive manifold. Thus, time means slowness and error means lack of errors.

The plot for the correctness scores (Figure 25a) shows that the four clusters separated timed matrices tests, timed series tests, untimed tests, and slowness scores. Slowness is better interpreted as carefulness, or willingness to cooperate and try hard in the experiment. It related highest to correctness on moderately timed tests, and lowest to correctness on highly speeded or untimed tests.

Within each of the two content clusters, more highly speeded tests fell near the periphery, while the less speeded tests were more centrally located. This concurs with previous observations that simple, speeded tests are more specific than complex, relatively unspeeded tests.

The location of the untimed tests suggests that they may not have been true power tests. Allowing unlimited time may permit inefficient but workable solution strategies that bypass the reasoning processes required when moderate time limits are imposed.

The locations of the two time limit matrices tests (points 1 and 10 in Figure 25a) suggests strong practice (or fatigue) effects, at least for matrices; the correlation between these two tests was only .39.

The scaling for the error scores (Figure 25b) revealed a markedly different structure. The content distinction was still strong, with series tests above matrices tests. With one exception, speeded tests were again more peripheral.

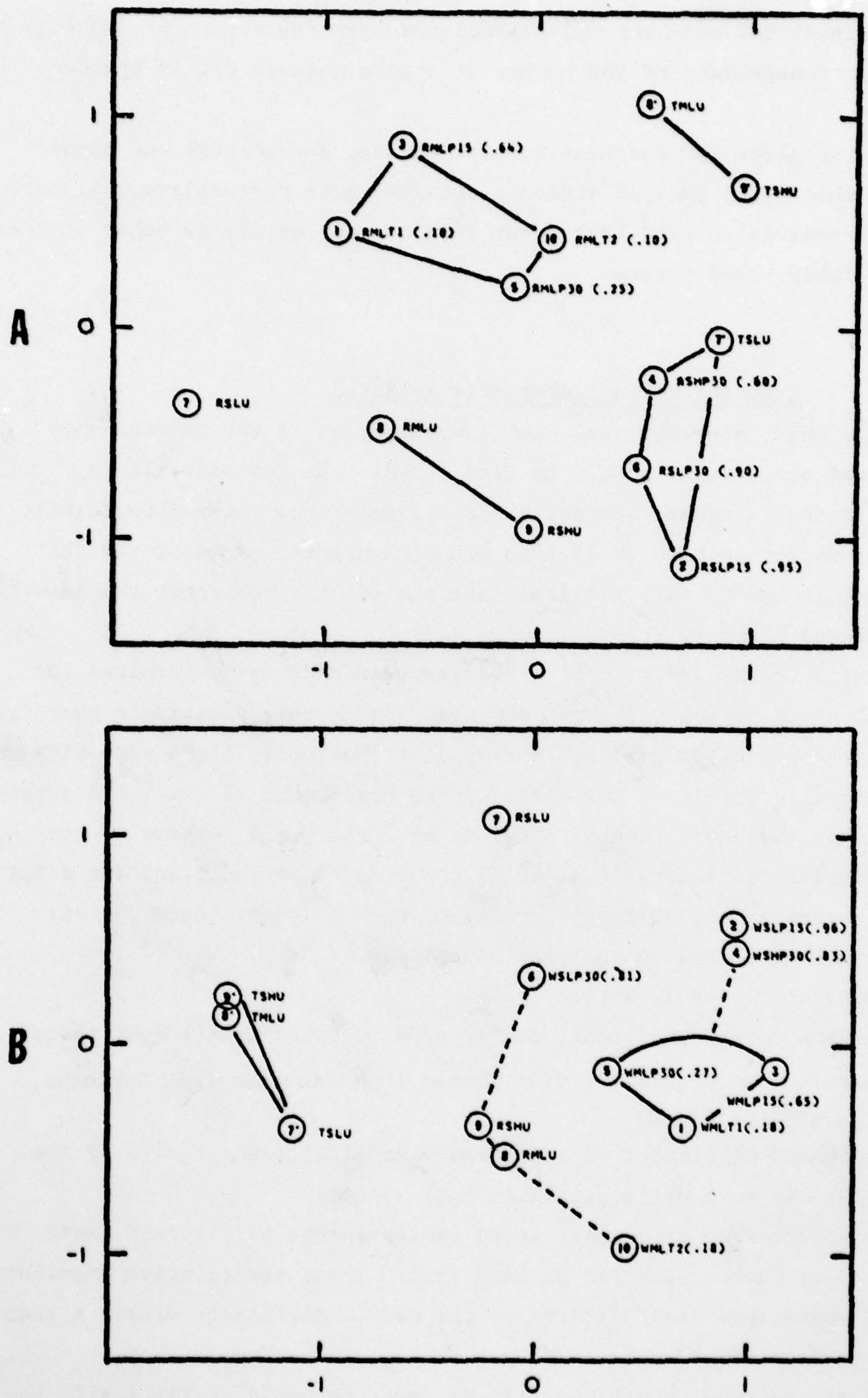


Figure 25. Two dimension scalings of (A) rights and (B) wrongs with average method clusters superimposed (After Morrison, 1960).

Further, carefulness was most strongly associated with few errors on low time-press tests, but independent of the number of errors made on highly speeded tests.

Together, the plots and correlation matrices for correctness and error scores suggest that slow, careful students obtained more correct answers, particularly on moderately speeded tests, but made as many errors as other students, especially on highly speeded tests.

Time for Correct Responses as Speed

Perhaps the most carefully conducted investigation of the relationship between speed and power was reported by Tate (1948). He administered an arithmetic reasoning, a number series, a spatial relations (easy Form Board), and a sentence completion test to 36 high school students. None of the test items were multiple choice; all required that the student construct the answer. Each test contained about 60 items.

Students were tested individually, and response time was determined for each item. The distributions of these raw time scores were positively skewed, however a log transformation produced normal distributions. Items were divided into three difficulty levels on the basis of the percentage of the total sample failing the item. The easy items were failed by three to 11 percent of the students, the medium difficulty items by 17 to 39 or 42 percent, and the difficult items by approximately 42 to 61 percent of the students. Data for correct and incorrect answers were analyzed separately.
The major results of the study were:

1. There were marked individual differences in speed in all four tests.

2. Log time was more normally distributed than raw time (see Furneaux, 1961, for a similar conclusion).

3. Accuracy and difficulty of the items were significant facets of the design; incorrect and more difficult items took longer.

4. There appeared to be a small speed factor common to all four tests, but a much larger speed factor specific to each test. Thus, the relative standing of subjects in speed were less affected by changes in difficulty within a test than by a change from one test to another.

5. Speed of response on difficult items, when adjusted statistically for accuracy, appeared to be independent of altitude or power in all four tests.

New Directions

This brief review of the speed-power problem has raised more questions than it has solved. The studies are unanimous on one point, however, and that is that speed is at least partially independent of power. Careful study of the relationship between the two is difficult, particularly in the light of the moderating effects of difficulty and correctness.

Group statistics such as percentage failing an item provide only a rough index of difficulty at the individual level. If accuracy, motivation and other extraneous factors are controlled, then time to solution should be positively related to difficulty. However, the reasoning is obviously circular, for if difficulty is defined in terms of time, then speed, power, and difficulty can not be disentangled.

For spatial tests, substituting complexity for difficulty may provide at least a partial solution to this dilemma. Complexity could be defined both in terms of the stimulus characteristics (i.e., two vs. three dimensions) and the number and type of mental operations required for solution.

Furneaux (1961) suggests an alternate solution for the construction of difficulty states. However, the method involves questionable assumptions and requires numerous transformations and modifications of the original data. It is certainly the most sophisticated mathematical model of speed, accuracy, and continuance (i.e., persistance) yet devised. However, the model has not been applied to tasks other than the set of Letter Series items used by Furneaux (1961).

The key to the speed-power problem is the construction of useful, psychologically meaningful difficulty scales. Furneaux has recognized this. However, his method of constructing difficulty scales yields indices that may be mathematically useful, but have no compelling psychological foundation.

The Egan Study

The first tentative steps in this direction are contained in a recent investigation by Egan (1976). He reported two experiments in which spatial tests were administered to naval officer candidates. The tests were administered in a group setting using paper-and-pencil multiple choice tests, and then individually with item exposure controlled and only one response alternative. Response latencies and correctness were obtained in the individual condition.

In the first experiment, two tests thought to measure Spatial Orientation and one Visualization test were administered. The choice of tests was unfortunate on two accounts. First, any study hoping to distinguish two constructs should have at least two measures of each construct (Camppbell and Fiske, 1959). Second, the tests chosen to represent the factors do not measure two factors, but only one factor. The Spatial Orientation factor was represented by the Guilford-Zimmerman Spatial Orientation test and the Navy's Spatial Apperception test. Both tests derive from the AAF Aerial Orientation test, the former showing shoreline pictures from a boat and the latter from an airplane. The Visualization factor was represented by the Guilford-Zimmerman Spatial Visualization test.

It is clear that the Guilford-Zimmerman Spatial Orientation and Spatial Visualization tests do not measure different factors not only from the re-analyses reported here (see p. 87), but also from two studies of the convergent and discriminant validity of these tests (Borich & Bauman, 1972; Price and Eliot, 1975). Even the manual for the Guilford-Zimmerman Aptitude Survey (Guilford and Zimmerman, 1948) shows that the correlation between these two tests is about as high as their respective reliabilities.

The second limitation of the study was a severe restriction of range in spatial ability in both experiments. The officer candidates were selected on a battery of tests, "a major component" of which was the Spatial Apperception Test (p. 4).

Although all subjects did not take all tests, pairwise correlations were based on the maximum number of cases. Thus, sample size ranged from 31 to 61 in the correlations for the first experiments, 48 to 72 for the second experiment, and 31 to 127 for a combined analysis.

In both experiments, speed was defined as the mean response latency for correct responses. Power (or level) was defined as the total number correct. The speed estimate is inadequate, as it is based on different items for different subjects. Thus, the speed estimate for the subject who responded correctly to only a few easy items was an estimate of speed of performing simple items. On the other hand, the speed estimate for the subject who answered almost all the items correctly was an estimate of speed of responding to moderately complex items.

There were nine variables in the correlation matrix for the first experiment: number correct and mean latency for correct responses for each of the three individually administered tests and number correct in the group administered version of each test. Average within and between group correlations for each of these three types of variables are shown in Table 41. The correlations were averaged here because there was no support for the hypothesis that the two SO tests measured something different than the Vz test.

Insert Table 41 about here

The average within group correlation for number correct in the individually administered tests (.36) was lower than the corresponding average within group correlation for the group administered tests ($\underline{r}$ = .50). This probably reflects the increased influence of guessing on the individual session scores. The correlation between correctness on the group and individual tests (.44) excludes the diagonal elements of this submatrix. These values are really reliability coefficients.

The average within group correlation for the mean latency scores ($\underline{r}$ = .54) was the highest in Table 41. On the other hand, correlations between these speed estimates and the level score were low and negative: -.15, -.30, and -.27 for the Spatial Apprehension, Guilford-Zimmerman Spatial Visualization, and Spatial Orientation tests, respectively. These correlations indicate a slight negative relationship between response latency on easy to medium difficulty items and total number correct.

The second experiment employed different subjects (72 in all, 48 with complete data). An adaptation of Shepard and Metzler's (1971) block rotation task was used instead of the Guilford-Zimmerman Spatial Orientation test. No group tests were administered this time.

Average within and between group correlations for the mean correct latencies and total number correct for the three tasks are shown in Table 42. Again, mean correct latencies intercorrelated higher than the correctness indices. Mean correct latencies on the block rotation task and the Spatial Visualization test were highly correlated. Both tests require mental rotation of an object. In the former, items differ in the angle of rotation and in the latter they differ both in the number of rotations and the angle of each rotation.

Insert Table 42 about here

The mean correlation between average latency and number correst was smaller than in the first experiment but still negative. The correlations between mean correct latency and total number correct were -.12, -.18, and -.26 for the Spatial Apprehension, Spatial Visualization, and Block Rotation tests, respectively. As in the first experiment, then, the implication is a small

Table 41

Average Within and Between Session Correlations for Latency and Correctness for Three Experiment 1 Spatial Tests (After Egan, 1976)

Score	1	2	3
Individual Session			
1. Total Correct	.36		
2. Mean Correct Latency	-.25	.54	
Group Session			
3. Total Correct	.44	-.30	-.50

Note. N varies from 31 to 61.

Table 42

Average Correlations for Latency and Correctness on Three Experiment 2 Spatial Tests (After Egan, 1976)

Score	1	2
1. Total Correct	.45	-.14
2. Mean Correct Latency		.53

Note. N varies from 48 to 72.

negative relationship between latency for correct responses on easy to medium difficulty items and total number of items answered correctly.

To this point, the Egan study is similar to the Tate (1948) study. However, Egan went beyond the usual analysis. He proposed three information processing models, one for the Spatial Visualization test, one for the block rotation task, and one for the Spatial Orientation tests.

The model for the visualization task hypothesized that items differ primarily in the number of times the clock must be rotated. Separate plots for (group) mean correct latencies, incorrect latencies, and proportion failing versus the number of rotations (0 - 4) were made, and all increased monotonically. Slopes and intercepts for the regression of mean correct latency on number of rotations were then calculated for each subject. Of the 106 slopes and intercepts calculated, only one of each was negative.

Following Shepard and Metzler (1971), the model proposed for the block rotation task hypothesized that items differed in the angle through which the stimulus figure had to be rotated. Similar plots were made for the block rotation task, this time with angle of required rotation on the abscissa. Average latency increased monotonically, although a logarithmic transformation of these mean latencies would have produced an almost perfectly linear plot. Again, slopes and intercepts were calculated for the 60 subjects. All of the slopes were positive, and only one intercept was negative.

Finally, the model proposed for the orientation tasks hypothesized that items differed primarily in the number of discrepant dimensions between the subject's concept of how the visual pattern should appear and the response alternative. Discrepancies could occur in any of three dimensions: heading, pitch, or bank.

The model predicted moderately well for the Spatial Apprehension test but poorly for the Guilford-Zimmerman Spatial Orientation test. Regression lines of latency for correct "no" responses on the number of discrepant dimensions were calculated for each subject on both tests. While no negative intercepts were obtained, 28 of the 127 of the individual slopes were positive on the Spatial Apprehension Test and 8 of 32 were positive on the Spatial Orientation test. The model predicts that all slopes should be negative.

Egan then computed the correlations between the various slope and intercept parameters, number correct, and mean correct latency. For each task, the highest correlations were obtained between slope and intercept parameters for the same task. In every case, steep slopes were associated with small intercepts ($\underline{r}$ = .70 for both SO tests, -.79 for the visualization test and -.63 for the block rotation test).

One set of results that were overlooked were the correlations between the intercept and total correct on each task. These correlations represent the correlation between speed of responding correctly to the "simplest" items and the power estimate. Here, simplicity is defined in terms of the unidimensional three or four point scales on which times were located.

Except for the Spatial Orientation test, the correlations were all in the -.20 to -.22 range. Once again, this indicates a mild negative relationship between latency for correct responses on simple items and power. Of course, it also indicates a substantial independence between the two measures. The corresponding correlation for the Spatial Orientation test was -.07, but the intercept parameter for this task is probably not meaningful.

While the results are interesting, the study ignores the possible confounding effects of item difficulty on the relationship between speed and power. The development of specific information processing models for the various tasks is indeed a step forward. However, the modeling and derivation of slope and intercept parameters was not directed toward the speed-power problem.

However, it is precisely this sort of analysis that can clarify the relationships between speed, power, and complexity, and provide an important link between "reaction time" information processing psychology and "number correct" differential psychology.

<u>Factors Affecting the Relationship between Speed and Level</u>

The relationship between speed and power is moderated by the following factors:

<u>Difficulty</u>. Speed and power probably correlate differently when speed is measured over simple tasks than when it is measured over complex tasks. However, since latencies for incorrect responses are uninterpretable, the correlation between speed and level can only be computed for error free items. A method for estimating the correlation at different levels of complexity is presented below.

<u>Content</u>. The relationship is probably different for verbal tests than for spatial tests. Speed and level appear to be more independent in spatial tests than in verbal tests.

Accuracy. Correct responses are generally faster than incorrect responses (Tate, 1948), although on extremely simple items incorrect responses are usually faster (Pachella, 1974).

Correct "yes" versus correct "no". Correct "yes" responses are usually faster than correct "no" respsonses, although some subjects show the opposite pattern on some spatial tasks (Cooper, 1976). Further, latency for these two types of correct responses may relate differently to difficulty, and thus complexity.

Guessing. The error variance introduced by this factor can seriously cloud the relationship between speed and power. Further, there are differences in the willingness to guess both between subjects and within subjects across tasks and situations. The effect of guessing is most pronounced in experiments where a yes/no response is required.

Alternative solution strategies. Many tasks can be solved in different ways. Some evidence suggests that particular tests (such as the Guilford-Zimmerman Spatial Orientation test) are particularly vulnerable. Items where several alternatives are provided are also suspect. The use of open-ended items (as in Tate, 1948) would eliminate alternative solution strategies in which the subject "works backwards" from the alternatives to the problem, or uses cues in the alternatives to help solve the problem. Some tasks may not be amenable to this procedure, particularly when the construction of the response is difficult or when processing time is short. However, many tasks can be administered in this way, such as the original Binet Paper Folding task or Thurstone's (1938) Punched Holes. The subject must draw how the holes will appear (Binet) and where they will be located (Binet and Thurstone) when the paper is unfolded.

Major alternative solution strategies that are not controlled can be included in the experiment. Here, best solution is to design the task so that different solution strategies produce qualitatively different patterns of performance over specific design facets.

Motivation. Thorndike (1926), Thurstone (1937) and Furneaux (1961) all agree that motivation (or persistence) can influence both level and speed. Further, the relationships between motivation, speed, and power are probably not linear. One can literally "try too hard." Thurstone's (1937) conclusion that increased motivation has no effect upon power but may increase speed undoubtedly oversimplifies matters. Nevertheless, it suggests that speed may be more dramatically affected by changes in motivation than level, particularly when motivation is increased from low to medium levels of arousal. The effects of motivation would probably best be assessed within a signal detection paradigm. The problem is so complex, however, that it may be better to attempt to control for motivational differences experimentally, at least until one can reasonably account for the effects of other variables such as difficulty.

Speed-accuracy tradeoff. Small changes in speed-accuracy tradeoff can produce large changes in response latency. Further, subjects interpret instructions to respond as fast and accurately as possible in different ways (Lohman, 1979). Even on extremely simple tasks, changes in instructions can produce large changes in performance (Howell and Kreider, 1963, 1964). The problem is even more vexing when complex items are presented. Instructions that assure a good power estimate vitiate the speed scores, while those that enhance the validity of the speed score may invalidate the power score.

Latency and Correctness as Dependent Variables

Latency and error rate are complimentary aspects of performance. Latency is most interpretable when there are no errors, while error rate becomes most meaningful when latency is uninterpretable. Keeping latency interpretable by studying only simple items or high ability subjects is unacceptable, as these models may not generalize to complex tasks or low ability subjects.

This dilemma is exemplified in an early investigation by Peak and Boring (1926). They individually administered two forms of the Otis and two forms of the Army Alpha to five subjects. Solution time was recorded for each item. Only those items answered correctly by every subject were included in the analysis "since differences in speed are significant only when accuracy is kept constant" (p. 80). After thus eliminating most of the difficult items, Peak and Boring (1926) concluded that "speed of reaction is an important, and probably the most important factor in individual differences in the intelligent act" (p. 92).

But, as Brigham (1932) notes, the procedure guarantees the result. With a larger, more representative sample, fewer and simpler items would be available for the analysis. While this procedure is obviously flawed, many investigations still routinely discard error trials, and yet hope their experiments explain individual differences in aptitude.

Experiments that seek generalizable results must also include a broad range of both task complexity and subject ability. Figure 26 shows how ability and task complexity are necessarily confounded in this type of experiment.

Each combination of ability and complexity represents a possible information processing model. There are potentially three different models for simple items, two for moderately complex items, and one for complex items. The model for complex items is necessarily a model for high ability subjects.

Insert Figure 26 about here

Contrasts between various cells indicate how performance of low and high ability subjects differ on simple items, or how the performance of high ability subjects is affected by shifts in item complexity.

However, even this type of analysis on correct response items overlooks some important problems. Figure 27 shows a hypothetical plot of the relationship between item complexity and latency for one subject, assuming motivation and solution strategy are held constant. Solution latency increases exponentially and approaches infinity as complexity approaches the level where the subject can no longer solve the items.

Insert Figure 27 about here

Latencies in the nonlinear portion of the curve are not as interpretable as those for the lower levels of complexity. As items become increasingly difficult, subjects may cycle through the item several times, try different solution strategies, or the like. While understanding such processes is important for a full understanding of aptitude processes, leaving them uncontrolled enormously complicates the task of modeling the data. Further, time taken for such processes may be erroneously attributed to the complexity facet, since the two are confounded. Thus, while investigations of aptitude processes demand that level scores be determined, latency data become increasingly uninterpretable as task complexity is increased.

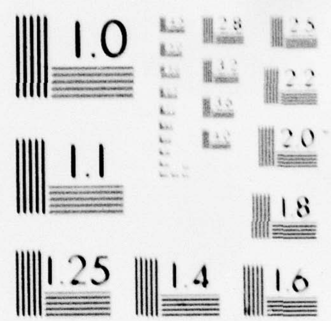
Error scores pose the reverse problem. Errors on simple problems may reflect different processes than errors on complex problems. Errors on simple problems may be caused by carelessness, fatigue, or inattention, while on complex problems they may indicate the breakdown of one or more component processes. Further, analysis of errors presents important scaling problems, since the variance of these scores is essentially zero for the simplest and most complex items, and maximum at the point where fifty percent of the items are failed.

Further, correctness and error rate sometimes relate differently to

AD-A075 972 STANFORD UNIV CALIF SCHOOL OF EDUCATION F/G 5/10
SPATIAL ABILITY: A REVIEW AND REANALYSIS OF THE CORRELATIONAL L--ETC(U)
OCT 79 D F LOHMAN N00014-75-C-0882
UNCLASSIFIED TR-8 NL

3 OF 3
AD
A075972

END
DATE
FILMED
12-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

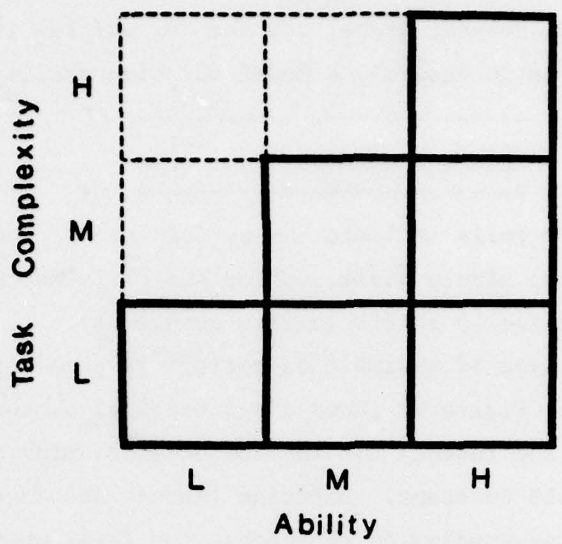


Figure 26. Task complexity by ability. Meaningful latency data in solid boxes.

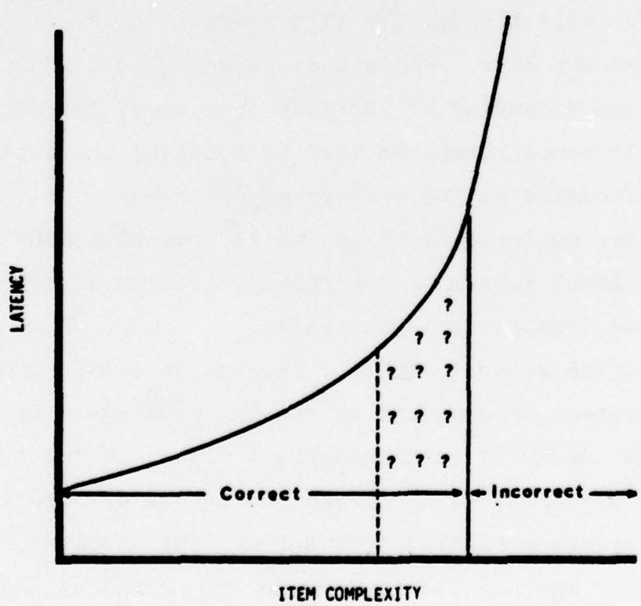


Figure 27. Hypothetical plot of response latency versus item complexity for one subject.

external variables (Davis, 1947; Fruchter, 1950; Morrison, 1960). This occurs when subjects attempt different items and all the reliable variance in the less reliable score (correctness or error) is not reflected in the correlation between the indices.

Thus, there are serious problems in the analysis and interpretation of latency and error data. Routine statistical analyses of such data may be misleading. Plots of raw data or simple descriptive statistics such as means, medians, and standard deviations would be more meaningful in most experiments.

Estimating the Correlation between Speed and Level

The only unambiguous speed-level correlation is between speed of solving error-free items and level. The correlation between speed and level can only be estimated when some subjects miss some items. This is shown graphically in Figure 28. The plot shows hypothetical regressions of latency for correct responses on item difficulty. Linear relationships are assumed for clarity. The solid portion of each regression line indicates the range of correct responses for each individual. Thus, the length of the solid portion of the regression line is the subject's level estimate.

Insert Figure 28 about here

The correlation between the intercepts and the lengths of the regression lines yields the correlation between speed of performing easiest item types and level. The individual regression lines can then be projected so that they all extend to the point of maximum complexity. At this point the correlation between these (for the most part, predicted) latencies and the range of the solid regression lines yields the best estimate of the relationship between speed of performing complex tasks and level. Projected or known latencies at intermediate points on the scale can also be correlated with range to yield intermediate values of the speed-level correlation.

When formulated in this manner, it is obvious that the correlation between speed and level will remain constant over the range of item complexity only if there are no individual differences in the regression slopes. This has important implications both for the speed-power problem and the generalizability of information processing parameters derived from simple tasks. If individual regression slopes are parallel, then the relationship between speed and level is constant throughout the range of complexity represented in the analysis.

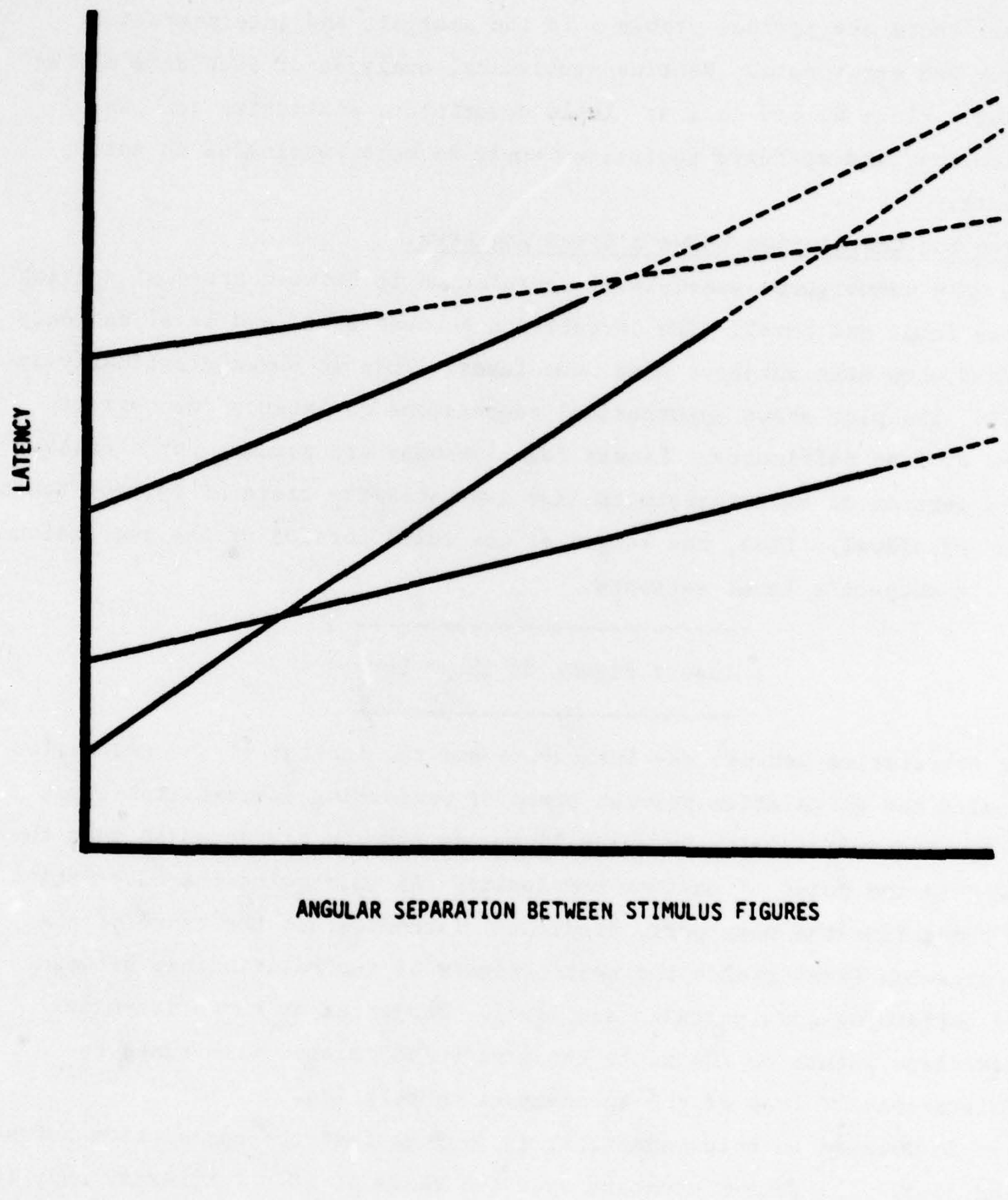


Figure 28. Hypothetical regressions of response latency on item complexity.

Parallel regression slopes also imply that individual differences in speed of solving simple tasks generalizes to the speed of solving complex tasks of the same type.

If there are individual differences in slopes, then speed of solving simple tasks does not generalize to speed of solving complex tasks. In either case, the relationship between speed and level, whether constant (i.e., slopes constant) or variable (i.e., slopes differ), is the crucial issue. A reasonable prediction in the area of spatial tasks would be that there are individual differences in regression slopes (see Cooper and Shepard, 1976).

Implications for Speed Factors

It is impossible to obtain meaningful speed and level scores from total time (or average latency) and total number correct on the same test. Studies that attempt to determine the relationship between speed and level by correlating speed and level indices derived from the same test assume that speed of solving easy items is perfectly correlated with speed of solving complex items, and that speed of correct responses is perfectly correlated with speed of incorrect responses. But these are unlikely assumptions.

Even if these assumptions were true, they would not erase the psychological ambiguity of latency for incorrect responses. Therefore, the only meaningful speed factors are those based on error-free performance. There can be no "speed of reasoning" factor in the traditional psychometric sense, for reasoning is, by definition, a construct based on level scores. This holds for all aptitude constructs defined by level scores.

These limitations do not apply to the more limited psychometric problem of determining the effects of altering the time limit of a test on the factor structure or predictive validity of a test (e.g., Yates, 1966a; Morrison, 1960). However, severely short time limits may allow the solution of only the easier items, especially under paced administrations. Changes in the correlations with other tests could then reflect changes in test content rather than the influence of a speed factor. Also, fewer items are solved under shorter time limits, making the total score on the test less reliable and producing lower correlations with other variables (e.g., see p.166).

Implications for Research on Aptitude

The most important implication of this literature for research on aptitude processes is that individual differences in latency on simple tasks may be largely independent of level scores on more complex tasks. This conclusion clearly warrants further study, for it questions much of the current research on aptitudes.

However , if (as advocated) a broad range of both task complexity and subject ability is represented in an experiment, then there need not be high correlations between latency based process parameters and reference constructs. This is because these parameters can be computed for all subjects only over the easiest items; parameters for the more difficult items can be computed only for the more able subjects (see Figure 26). Restriction of range would then limit the correlations between process parameters and reference constructs for all item types not solved by some subjects. Error scores, however, should show convergent validity with reference constructs.

Summary and Conclusions

1. It appears that severe changes in test speededness are required to alter the factor structure of a test. This suggests that changes in complexity may be the important dimension, since fewer difficult items would be solved with extremely short time limits. Morrison (1960) found that pacing produced higher speededness than an equivalent total time allowance. But, again, fewer difficult items would be solved in the paced condition than in time limit condition. Similarly, reanalyses of Lord's (1954) data showed that, within each content area, level tests correlated as highly with parallel speed tests as with other level tests. Thus, changes in speededness may be less important than changes in item complexity, and test length in producing changes in the factor structure of a test. At the other extreme, allowing unlimited time may alter test factor structure by permitting inefficient but workable solution strategies that bypass the aptitude processes required under moderate time limits. The generally lower predictive validity of untimed level scores than time limit scores may reflect these strategic shifts.

2. Speed factors for level constructs such as reasoning are impossible, since individual differences in speed can be measured only over error free items. While individual differences in speed of reasoning undoubtedly exist, they cannot be represented using conventional correlational methods. Therefore, studies in which total time and number correct were obtained for each test, and then used to define level and speed factors are flawed. Either the level scores are invalid because items are too simple (e.g., the verbal analogies test

in Davidson & Carroll, 1945), or the speed scores reflect time for guesses, abandonments, and incorrect responses as well as time for correct responses. Speed factors can be defined only by simple, error free tests. No evidence for a general speed factor was found, even in factor analyses of contaminated speed scores.

3. Speed of solving simple spatial items appears to be largely independent of level scores over more complex items of the same type (Egan, 1976). Speed of solving simple items appears to be more highly correlated with level on verbal and reasoning tests (Davidson & Carroll, 1945; Lord, 1956), but methodologically sound studies of the relationship are lacking.

4. Latency and error are complementary aspects of performance. Latency is most interpretable when there are no errors, while error rate becomes most meaningful when latency is uninterpretable. Further, it is extremely difficult to gather clean latency data. Small changes in speed-accuracy tradeoff or item difficulty can produce large changes in response latency. Even latency for correct responses may be uninterpretable (see Figure 27). But this sensitivity makes latency a powerful variable in detecting individual differences in cognitive processes.

5. Experiments that hope to explain general (i.e., level) aptitude constructs must represent a wide range of both aptitude and item complexity. Total errors on the experimental task should show convergent validity with reference tasks. Latency based process parameters may be independent of these reference tasks since process parameters can be computed for all subjects only on the easier items. Process models for complex items are necessarily models for high ability subjects.

6. The relative independence of individual differences in speed of solving simple items and level challenges much of the recent work on the nature of aptitude processes. Many of these studies have avoided the problem of latency for incorrect responses by keeping items simple or studying only high ability subjects. But such process models may not generalize to complex items or low ability subjects. Therefore, investigators must pay more attention to the speed-level problem. Failure to do so has caused considerable confusion in differential psychology. Ignoring level has produced constructs in cognitive psychology that are of questionable generalizability. Resolution of the relationships between speed and level is important not only for the separate understandings of differential psychology and information processing psychology, but is at the heart of any attempt to forge a rapprochement between them.

GENERAL CONCLUSIONS

Spatial Ability

1. Definition. Spatial ability may be defined as the ability to generate, retain, and manipulate abstract visual images.

2. Major spatial factors. Three major spatial factors were identified in this review. All three require mental transformation. They are:

Spatial Relations (SR). This factor is defined by tests such as Cards, Flags, and Figures (Thurstone, 1938). It emerges only if these or highly similar tests are included in the battery. Although mental rotation is the common element, the factor probably does not represent speed of mental rotation. Rather, it represents the ability to solve such problems quickly, by whatever means.

Spatial Orientation (SO). This factor appears to involve the ability to imagine how a stimulus array will appear from another perspective. In the true spatial orientation test, the subjects must imagine they are reoriented in space, and then make some judgment about the situation. There is often a left-right discrimination component in these tasks, but this discrimination must be made from the imagined perspective. However, the factor is difficult to measure since tests designed to tap it are often solved by mentally rotating the stimulus rather by reorienting an imagined self.

Visualization (Vz). The factor is represented by a wide variety of tests such as Paper Folding, Form Board, WAIS Block Design, Hidden Figures, Copying, and Surface Development. The tests that load on this factor, in addition to their spatial-figural content, share two important features: (a) all are administered under relatively unspeeded conditions, and (b) most are much more complex than corresponding tests that load on the more peripheral factors. Tests designed to measure this factor usually fall near the center of a two dimensional scaling representation, and are often quite close to tests of Spearman's "g" (such as Raven Matrices or Figure Classification) or Cattell's (1963) Gf.

3. <u>Minor spatial factors</u>. At the most basic level, spatial thinking requires the ability to encode, remember, transform, and match spatial stimuli. Factors such as Closure Speed (i.e., speed of matching incomplete visual stimuli with their long term memory representations), Perceptual Speed (speed of matching visual stimuli), Visual Memory (short term memory for visual stimuli) and Kinesthetic (speed of making left-right discriminations) may represent individual differences in the speed or efficiency of these basic cognitive processes. These factors surface only when extremely similar tests are included in a test battery. Such tests and their factors consistently fall near the periphery of scaling representations, or at the bottom of a hierarchical model.

4. <u>Types of spatial transformation</u>. Two types of spatial transformation are required by tests that define the three major spatial factors (SR, SO, and Vz). The first is mental movement. Reflecting, rotating, folding, or simply imagining that a stimulus is moved from one position in an array to another position, are all varieties of mental movement.

The second type of mental transformation may be called construction. There are two types of construction: reproduction (i.e., physical construction) and combination (i.e., mental construction). At the simplest level, reproduction is represented in tests like Thurstone's (1938) Copying, where the subject must correctly copy a stimulus design. At the next level, it is represented by tests like Graham and Kendall's (1948) Memory for Designs, where the design must be reproduced, not just recognized, and the reproduced design must be a veridical representation of the stimulus. Retaining a veridical mental image of a design may be an important component of other complex spatial tasks, such as Hidden Figures (French et al., 1963).

In the mental construction tasks, on the other hand, the subject must actually construct a mental image, usually by reorganizing the stimulus in a new way. The clearest examples of this sort of process are tests like Form Equations (El Koussy, 1935) and Paper Form Board (e.g., Thurstone, 1938; French, Ekstrom and Price, 1963). Mental construction is an important component of many complex spatial tests. For example, in Surface Development (French et al., 1963), the examinee must construct new holes as he mentally unfolds the stimulus. Finally, mental construction may take the form of mentally deleting parts of a stimulus, as in Match Problems (Guilford and Hoepfner, 1971). This may also be an important component of tests such as Embedded Figures (Witkin, Oltman, Raskin and Karp, 1971) or Hidden Figures (French et al., 1963).

5. <u>The analytic nature of spatial tests</u>. Tests that consistently define the major spatial factors represent only a limited portion of the visual thinking domain. All require analytic problem solving skills. Spatial tests may become analytic because subjects use all the resources at their disposal when placed in problem solving situations. Thus, they use both verbal-analytic processes and analog spatial processes to solve spatial test items. In other words, spatial tests measure spatial problem solving skills, not necessarily analog spatial ability. Individual differences in spatial ability may be more independent of verbal ability than the correlational literature suggests.

6. <u>The non-hierarchical nature of ability factors</u>. Level factors that define general abilities cannot be subdivided into speed primaries. Conversely, second order factors of speed primaries do not coincide with the level factors. Level tests and their factors are highly intercorrelated, while speed tests and their factors are largely independent of one another and level tests. This non-hierarchical nature of ability factors reflects the imperfect relationship between individual differences in speed and level.

<u>Solution Strategies</u>

1. There are important differences in solution strategy both between subjects and within subjects over items. Tests often measure different abilities for different students, depending on how problems are solved.

2. Complex, power tests elicit a wider range of alternative solution strategies than simple, highly speeded tests. Vz tests are often solved in more ways than SR tests.

3. Within a test, the more difficult items elicit a wider range of solution strategies than easy items.

4. High ability students report studying the problem stem and constructing an answer before examining the alternatives. They are usually able to give a coherent verbal report of how they solved the item, and they express confidence in their answers. Low ability students, on the other hand, frequently report that they attempt to solve the item by analyzing the alternatives. Further, they report more internal verbalization, more guessing, and less confidence in their answers than do high ability students.

5. Certain tests are particularly susceptible to alternative solution strategies. For example, many Spatial Orientation tests can be solved by a Visualization strategy. On a more general level, multiple choice paper and pencil tests permit a number of alternative solution strategies that are not

possible when the student must construct rather than select an answer. Students can also draw or mark on the test, thereby reducing the need to remember more than a single step in the solution of the problem. They can attempt to solve the problem by "working backwards" from the alternatives to the stem, or look for clues in the alternatives that may reveal the correct answer, or simply narrow the field. A range of alternative solution strategies could be eliminated by using free response rather than multiple choice items.

6. Introspective reports are of limited value. Whenever possible, such reports should be validated against external information. Many processes, especially those that are extremely rapid, cannot be accessed through introspection (see Nisbett and Wilson, 1977). Retrospective reports are even less trustworthy. Such reports are best used as a rough index of strategy rather than as a guide to mental processes. Detailed retrospections are probably quite unreliable. Thus, subjects could be expected to indicate whether they mentally rotated an object or, instead, mentally projected themselves into the picture. It is unlikely, however, that they would be able to accurately decompose these global behaviors into component processes.

7. Perhaps the most promising technique for obtaining valid introspective evidence is to ask subjects to report <u>specific</u> strategy information immediately before (Karpf and Levine, 1971), during (Kroll and Kellicutt, 1972), or after (Paivio and Yuille, 1969) they solve an item, usually by anonymously pressing a button. The validity of the self report rises dramatically, although reactive effects might present problems.

8. Individual differences in solution strategies challenge a basic assumption of factor analysis. Factor structures obtained from analyses of such tests may be severely distorted. The most likely outcome is an overestimation of the factorial complexity of a test. Thus, that some SO tests load on both Vz and SO factors may only mean that students solve the tests differently: some use a predominately SO strategy, while others rely on a Vz strategy. Alternately, students may switch between these two strategies while solving different items. However, even in this straightforward example, it is impossible to know whether the test measures two different aptitudes in any one individual, or whether it measures different aptitudes in different individuals. On a more general level, the presence of several tests in a battery that are amenable to alternate solution strategies seriously distorts the factor structure, so that the obtained factor structure may not apply to anyone in the sample. Factoring within strategy groups would undoubtedly produce cleaner factor patterns.

9. Individual differences in solution strategy present a major stumbling block for both correlational and experimental investigations of spatial ability. The challenge for future research is to devise experiments that reveal solution strategies for each subject on each item or on each item-type. Only by knowing how subjects solve items can the investigator know what the task measures, or evaluate the generalizability of the processing models that are proposed to describe overall task performance.

Speed-Level

1. It appears that severe changes in test speededness are required to alter the factor structure of a test. This suggests that changes in complexity may be the important dimension, since fewer difficult items would be solved with extremely short time limits. Morrison (1960) found that pacing produced higher speededness than an equivalent total time allowance. But, again, fewer difficult items would be solved in the paced condition than in time limit condition. Similarly, reanalyses of Lord's (1954) data showed that within each content area, level tests correlated as highly with parallel speed tests as with other level tests. Thus, changes in speededness may be less important than changes in item complexity and test length in producing changes in the factor structure of a test. At the other extreme, allowing unlimited time may alter test factor structure by permitting inefficient but workable solution strategies that bypass the aptitude processes required under moderate time limits. The generally lower predictive validity of untimed level scores than time limit scores may reflect these strategic shifts.

2. Speed factors for level constructs such as reasoning are impossible, since individual differences in speed can be measured only over error-free items. While individual differences in speed of reasoning undoubtedly exist, they cannot be represented using conventional correlational methods. Therefore, studies in which total time and number correct were obtained for each test, and then used to define level and speed factors are flawed. Either the level scores are invalid because items are too simple or the speed scores reflect time for guesses, abandonments, and incorrect responses as well as time for correct responses. Speed factors can be defined only by simple, error free tests. No evidence for a general speed factor was found, even in factor analyses of contaminated speed scores.

3. Speed of solving simple spatial items appears to be largely independent of level scores over more complex items of the same type (Egan, 1976). Speed

of solving simple items appears to be more highly correlated with level on verbal and reasoning tests (Davidson & Carroll, 1945 ; Lord, 1956), but methodologically sound studies of the relationship are lacking.

4. Latency and error are complementary aspects of performance. Latency is most interpretable when there are no errors, while error rate becomes most meaningful when latency is uninterpretable. Further, it is extremely difficult to gather clean latency data. Small changes in speed-accuracy tradeoff or item difficulty can produce large changes in response latency. Even latency for correct responses may be uninterpretable (see Figure 27). But this sensitivity makes latency a powerful variable in detecting individual differences in cognitive processes.

5. Experiments that hope to explain general (i.e., level) aptitude constructs must represent a wide range of both aptitude and item complexity. Total errors on the experimental task should show convergent validity with reference tasks. Latency based process parameters may be independent of these reference tasks since process parameters can be computed for all subjects only on the easier items. Process models for complex items are necessarily models for high ability subjects.

6. The relative independence of individual differences in speed of solving simple items and level challenges much of the recent work on the nature of aptitude processes. Many of these studies have avoided the problem of latency for incorrect responses by keeping items simple or studying only high ability subjects. But such process models may not generalize to complex items or low ability subjects. Therefore, investigators must pay more attention to the speed-level problem. Failure to do so has caused considerable confusion in differential psychology. Ignoring level has produced constructs in cognitive psychology that are of questionable generalizability. Resolution of the relationships between speed and level is important not only for the separate understandings of differential psychology and information processing psychology, but is at the heart of any attempt to forge a rapprochement between them.

General Comments

1. The purpose of this review was to explore the implications of correlational research on spatial ability for experimental research on individual differences in spatial ability. The review does not defend factor analysis or advocate this type of research. In fact, little was added to our understanding of spatial ability by the hundreds of investigations that followed Thurstone's (1938) Primary Mental

Abilities study. Factor analysis, or better, multidimensional scaling of test correlations generates a rough map of the individual differences terrain. This map provides a fertile ground for hypothesis generation, but a weak foundation for psychological theory. One of the major problems is that tests are solved in different ways by different subjects. Subjects change their solution strategies with practice or when item difficulty increases. Further, most factors represent individual differences in speed of solving particular types of problems, not general problem solving skills or abilities.

But in spite of these limitations, there are important lessons for experimental research. Factors are defined by the common covariation in tests; the idiosyncratic variance in each test is discarded. But the unique variance in each test is frequently as large as the portion of the tests' variance "explained" by the factor on which the test has its primary loading. The major strength of the correlational method is that it immediately captures this common, generalizable variance in each task. But in the experimental analysis of one test, there is no easy way to separate the generalizable from the specific. A complete accounting of individual differences in Paper Folding is not an explanation of spatial ability, for many of the processes required by Paper Folding are task specific. Only a small subset generalize to mental rotation tasks such as Cards or Figures.

2. The process of adapting a test to an experimental task may drastically alter the nature of the test. If a general ability test is made simpler to eliminate the problem of latency for incorrect responses, then the experimental task will most likely tap some specific problem solving skill, not general ability. Making the task more speeded by controlling item presentation or altering speed-accuracy instructions may also make the task more specific. Solution strategies change with practice, and so including more experimental trials than test items may make the task more specific than the test, or vise versa. Finally, mode of item presentation may eliminate or favor particular solution strategies. Some of these changes may enhance the construct validity of the task, but improvements must be verified, not assumed. An experimental task will rarely tap exactly the same mental processes as the source test.

References

Anzai, Y., & Simon, H.A. The theory of learning by doing. Psychological Review, 1979, 86, 124-140.

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.

Baxter, B. An experimental analysis of the contributions of speed and level in an intelligence test. Journal of Educational Psychology, 1941, 32, 285-296.

Bechtoldt, H.P. Factorial investigation of the perceptual speed factor. American Psychologist, 1947, 2, 304-305.

Bloom, B.S., & Broder, L.J. Problem-solving processes of college students. Supplementary Educational Monographs, 1950, 73.

Bock, R.D. Word and image: Sources of the verbal and spatial factors in mental test scores. Psychometrika, 1973, 38, 437-457.

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.

Botzum, W.A. A factorial study of reasoning and closure factors. Psychometrika, 1951, 16, 361-386.

Brigham, C.C. A study of error. New York: College Entrance Examination Board, 1932.

Burt, C. The structure of the mind: A review of the results of factor analysis. British Journal of Educational Psychology, 1949, 19, 100-114.

Campbell, D.T., & Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Cattell, R.B. Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 1963, 54, 1-22.

Cattell, R.B. Abilities: Their structure, growth, and action. Boston: Houghton Mifflin, 1971.

Christal, R.E. Factor analytic study of visual memory. Psychological Monographs, 1958, 72(13, Whole No. 466).

Clark, H., & Chase, W. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

Cliff, N. Orthogonal rotation to congruence. Psychometrika, 1966, 31, 33-42.

Conry, R., & Lohman, D.F. Characteristics of a facet-designed measure of memory for pictures. Paper presented at the annual meeting of the American Educational Research Association, New York, August 1976.

Cooper, L.A. Individual differences in visual comparison processes. Perception & Psychophysics, 1976, 19, 433-444.

Cooper, L.A., & Shepard, R.N. Transformations on representations of objects in space. Tech. Rep. 59, Center for Information Processing, University of California, San Diego, California, 1976.

Cronbach, L.J. Essentials of psychological testing (3rd ed.). New York: Harper & Row, 1970.

Cronbach, L.J., & Snow, R.E. Aptitudes and instructional methods: A handbook for research on interactions. New York: Irvington, 1977.

Cronbach, L.J., & Warrington, W.G. Time-limit tests: Estimating their reliability and degree of speeding. Psychometrika, 1951, 16, 167-188.

Davidson, W.M., & Carroll, J.B. Speed and level components in time-limit scores: A factor analysis. Educational and Psychological Measurement, 1945, 5, 411-427.

Davis, P.C. Measures of carefulness. In J.P. Guilford & J.I. Lacey (Eds.), Printed classification tests, AAF aviation psychology research program reports (No. 5). Washington, D.C.: GPO, 1947.

Dewey, J. How we think: A restatement of the relation of reflective thinking to the educative process. New York: D.C. Heath & Co., 1933.

Dixon, W.J. (Ed.). Biomedical computer programs. Berkeley: University of California Press, 1975.

Egan, D.E. Accuracy and latency scores as measures of spatial information processing (NAMRL Rep. No. 1224). Naval Aerospace Medical Research Laboratory, Pensacola, Florida, February 1976.

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

Eysenck, H.J. Intelligence assessment: A theoretical and experimental approach. British Journal of Educational Psychology, 1967, 37, 81-98.

Frandsen, A.N., & Holder, J.R., Spatial visualization in solving complex verbal problems. Journal of Psychology, 1969, 73, 229-233.

Freeman, F.N. Note on the relation between speed and accuracy or quality of product. Journal of Educational Research, 1923, 7, 87-89.

French, J.W. The description of aptitude and achievement tests in terms of rotated factors. Psychometric Monograph, 1951, 5.

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, N.J.: Educational Testing Service, 1963.

Fruchter, B. Error scores as a measure of carefulness. Journal of Educational Psychology, 1950, 41, 279-291.

Furneaux, W.D. Intellectual abilities and problem-solving behavior. In H.J. Eysenck (Ed.), Handbook of abnormal psychology. London: Pitman, 1960.

Gavurin, E.I. Anagram solving and spatial aptitude. Journal of Psychology, 1967, 65, 65-68.

Guilford, J.P. The structure of the intellect. Psychological Bulletin, 1956, 53, 276-293.

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

Guilford, J.P., Fruchter, B., & Zimmerman, W.S. Factor analysis of the Army Air Forces Shepard Field battery of experimental aptitude tests. Psychometrika, 1952, 17(1), 45-68.

Guilford, J.P., & Hoepfner, R. The analysis of intelligence. New York: McGraw-Hill, 1971.

Guilford, J.P., & Lacey, J.I. (Eds.). Printed classification tests. AAF aviation psychology research program reports (No. 5). Washington, D.C.: GPO, 1947.

Guilford, J.P., & Zimmerman, W.S. The Guilford-Zimmerman aptitude survey. Beverly Hills, California: Sheridan Psychological Services, 1948.

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

Holzinger, K.J., & Harman, H.H. Comparison of two factorial analyses. Psychometrika, 1938, 3, 45-60.

Holzinger, K.J., & Swineford, F. A study in factor analysis: The stability of a bifactor solution. Supplementary Educational Monographs, 1939, 48.

Horn, J.L. Human abilities: A review of research and theory in the early 1970's. Annual Review of Psychology, 1976, 27, 437-485.

Horn, J.L., & Bramble, W.J. Second-order ability structure revealed in right and wrongs scores. Journal of Educational Psychology, 1967, 58, 115-122.

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

Howell, W.C., & Kreidler, D.L. Information processing under contradictory instructional sets. Journal of Experimental Psychology, 1963, 65, 39-46.

Howell, W.C., & Kreidler, D.L. Instructional sets and subjective criterion levels in a complex information processing task. Journal of Experimental Psychology, 1964, 68, 612-614.

Humphreys, L.G. The organization of human abilities. American Psychologist, 1962, 17, 475-483.

Hunsicker, L.M. A study of the relationship between rate and ability. In Contributions to Education, No. 185. New York: Columbia University, Teachers College, 1925.

Hunt, E., Frost, N., & Lunneborg, C.E. Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), Advances in learning and memory (Vol. 7). New York: Academic Press, 1973.

Hunt, E., Lunneborg, C.E., & Lewis, J. What does it mean to be high verbal? Cognitive Psychology, 1976, 8, 1-31.

Johnson, S.C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 241-254.

Kelley, T.L. Crossroads in the mind of man. Stanford, California: Stanford University Press, 1928.

Kohs, S.C. Intelligence measurement. New York: Macmillan, 1923.

Koussy, A.A.H. El The visual perception of space. British Journal of Psychology, 1935, 20. (Monograph supplement)

Kroll, N.E.A., & Kellicutt, M.H. Short-term recall as a function of covert rehearsal and of intervening task. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 196-204.

Kruskal, J.B., Young, F.W., & Seery, J.B. How to use KYST. Murray Hill, N.J.: Bell Laboratories, 1973.

Lieberman, D.A. Behaviorism and the mind: A (limited) call for a return to introspection. American Psychologist, 1979, 34, 319-333.

Lohman, D.F. The relationship between hypnotizability and speed of closure (Tech. Rep. No. 6). Aptitude Research Project, School of Education, Stanford University, Stanford, California, 1977.

Lohman, D.F. Eye movement differences reflecting aptitude processes. Paper presented at the meeting of the American Psychological Association, San Francisco, August 1977. (b)

Lohman, D.F. Spatial ability: Individual differences in speed and level (Tech. Rep. No. 9). Aptitude Research Project, School of Education, Stanford University, Stanford, California, 1979.

Lord, F.M. A study of speed factors in tests and academic grades. Psychometrika, 1956, 21, 31-50.

Marshalek, B. The complexity dimension in the radex and hierarchical models. Paper presented at the meeting of the American Psychological Association, San Francisco, August 1977.

May, M.A. Psychological examining in the U.S. Army. Memoirs of the National Academy of Sciences, 1921, 15, 416.

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

Merrifield, P.R. Structuring mental acts. In W.B. Dockrell (Ed.), On Intelligence. Toronto: Ontario Institute for Studies in Education, 1970.

Metzler, J., & Shepard, R.N. Transformational studies of the internal representation of three-dimensional objects. In R. Solso (Ed.), Theories in cognitive psychology: The Loyola Symposium. Potomac, Maryland: Lawrence Erlbaum, 1974.

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 the spatial relations and visualization factors. Educational and Psychological Measurement, 1950, 10, 187-243.

Michael, W.B., Zimmerman, W.S., & Guilford, J.P. An investigation of the nature of the spatial relations and visualization factors in two high school samples. Educational and Psychological Measurement, 1951, 11, 561-577.

Miller, G.A., Galanter, E., & Pribram, K.H. Plans and the structure of behavior. New York: Holt, Rinehart & Winston, 1960.

Mooney, C.M. A factorial study of closure. Canadian Journal of Psychology, 1954, 8, 51-60.

Morrison, E.J. Effects of time limits on the efficience and factorial compositions of reasoning measures. Unpublished doctoral dissertation, University of Illinois, 1960.

Myers, C.T, The factorial compostion of differently speeded tests. Psychometrika, 1952, 17, 347-352.

Myers, C.T. Some observations and opinions concerning spatial relations tests (ETS RM 57-7). Princeton, N.J.: Educational Testing Service, 1957.

Myers, C.T. Some observations of problem solving in spatial relations tests (ETS RB 58-16). Princeton, N.J.: Educational Testing Service, 1958.

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

O'Sullivan, M., Guilford, J.P., & deMille, R. The measurement of social intelligence (Rep. No. 34). Psychological Laboratory, University of Southern Californa, Los Angeles, 1965.

Pachella, R.G. The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition. Hillsdale, N.J.: Lawrence Erlbaum, 1974.

Paivio, A., & Yuille, J.C. Changes in associative strategies and paired-associate learning over trials as a function of word imagery and type of learning set. Journal of Experimental Psychology, 1969, 79, 458-463.

Peak, H., & Boring, E.G. The factor of speed in intelligence. Journal of Experimental Psychology, 1926, 9, 71-94.

Pemberton, C. The closure factors related to other cognitive processes. Psychometrika, 1952, 17, 267-288.

Price, L., & Eliot, J. Convergent and discriminant validities of two sets of measures of spatial orientation and visualization. Educational and Psychological Measurement, 1975, 35, 975-977.

Raven, T.C. Progressive matrices. London: Crichton, 1947.

Ruch, G.M. The speed factor in mental measurements. Journal of Educational Research, 1924, 9, 39-45.

Ruch, G.M., & Koerth, W. Power vs. speed in Army Alpha. Journal of Educational Psychology, 1923, 14, 193-208.

Seibert, W.F., & Snow, R.E. Studies in cine-psychometry I: Preliminary factor analysis of visual cognition and memory. Audio Visual Center, Purdue University, Lafayette, Indiana, 1965.

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

Smith, I.M. Spatial ability. San Diego: Knapp, 1964.

Snow, R.E. Research on aptitudes: A progress report (Tech. Rep. No. 1). Aptitude Research Project, School of Education, Stanford University, Stanford, California, 1976. (a)

Snow, R.E. Theory and method for research on aptitude processes: A prospectus (Tech. Rep. No. 2). Aptitude Research Project, School of Education, Stanford University, Stanford, California, 1976. (b)

Snow, R.E. Toward a theory of aptitude. Invited address at the American Educational Research Association, Division C, March 1978.

Snow, R.E., Lohman, D.F., Marshalek, B., Yalow, E., & Webb, N.M. Correlational analyses of reference aptitude constructs (Tech. Rep. No. 5). Aptitude Research Project, School of Education, Stanford University, Stanford, CA, 1977.

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

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

Sternberg, R.J. Intelligence, information processing and analogical reasoning: The componential analysis of human abilities. Hillsdale, N.J.: Lawrence Erlbaum, 1977.

Swineford, F., & Holzinger, K.J. A study in factor analysis: The reliability of bi-factors and their relation to other measures. Supplementary Educational Monographs, 1942, 53.

Tate, M.W. Individual differences in speed of response in mental test materials of varying degrees of difficulty. Educational and Psychological Measurement, 1948, 8, 353-374.

Thorndike, E.L. The measurement of intelligence. New York: Columbia University Press, 1926.

Thorndike, R.L. Factor analysis of social and abstract intelligence. Journal of Educational Psychology, 1936, 27, 231-233.

Thurstone, L.L. Ability, motivation, and speed. Psychometrika, 1937, 2, 249-254.

Thurstone, L.L. Primary mental abilities. Psychometric Monographs, 1938, 1.

Thurstone, L.L. A factorial study of perception. Chicago: University of Chicago Press. 1944.

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

Thurstone, L.L. An analysis of mechanical aptitude (Rep. No. 62). Psychometric Laboratory, University of Chicago, Chicago, 1951.

Thurstone, L.L., & Thurstone, T.G. Factorial studies of intelligence. Psychometric Monographs, 1941, 2.

Thurstone, L.L., & Thurstone, T.G. SRA primary mental abilities. Chicago: Science Research Association, 1947.

Vernon, P.E. The structure of human abilities. London: Methuen, 1950.

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

Walters, F.C. A statistical study of certain aspects of the time factor in intelligence. In Contributions to Education, No. 248. New York: Columbia University, Teachers College, 1927.

Wechsler, D. The measurement of adult intelligence. Baltimore: Williams & Wilkins, 1939.

Wertheimer, M. Productive thinking. New York: Harper, 1945.

Wherry, R.J. Hierarchical factor solutions without rotation. Psychometrika, 1959, 24, 45-51.

Witkin, H.A., Oltman, P.K., Raskin, E., & Karp, S.A. Embedded figures test. Palo Alto, California: Consulting Psychologists Press, 1971.

Woodrow, H. Factors in improvement with practice. Journal of Psychology, 1939, 7, 55-70. (a)

Woodrow, H. The common factors in fifty-two mental tests. Psychometrika, 1939, 4, 99-107. (b)

Wolfe, D. Factor analysis to 1940. Psychometric Monographs, 1940, 3.

Wrightstone, J.W., & O'Toole, C.E. Prognostic test of mechanical abilities. Monterey, California: California Test Bureau, 1947.

Wrigley, C., Saunders, D.R., & Neuhaus, J.O. Application of the quartimax method of rotation to Thurstone's primary mental abilities study. Psychometrika, 1958, 23, 151-169.

Yalow, E., & Webb, N. Introspective strategy differences reflecting aptitude processes. Paper presented at the meeting of the American Psychological Association, San Francisco, August 1977.

Yates, A.J. A further study of progressive matrices (1947). British Journal of Educational Psychology, 1963, 33, 307-311.

Yates, A.J. The relationship between level and speed on two intelligence tests. British Journal of Educational Psychology, 1966, 36, 166-170. (a)

Yates, A.J. Level, speed and personality factors in the intellectual performance of young children. British Journal of Educational Psychology, 1966, 36, 312-316. (b)

Zaidel, D., & Sperry, R. Performance on the Raven's colored progressive matrices by subjects with cerebral commissurotomy. Cortex, 1973, 9, 34-39.

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

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

Footnotes

1. Pat Kyllonen of our project recently performed another reanalysis of the PMA data using nonmetric multidimensional scaling and hierarchical clustering. These analyses will be summarized in a future technical report.

2. This was evident in one analysis performed on the Aptitude Project reference battery (see Snow et al., 1977). The high school sample (N=243) was divided into two groups on the basis of factor scores on the general factor, estimated by the first unrotated centroid. Within group correlation matrices were then separately factored. There were four important differences between the high and low ability groups. (a) The general factor was larger in the low ability group. (b) Uses for Things loaded strongly on a verbal factor in the high ability group but loaded strongly on visual memory and spatial factors in the low ability group. (c) One spatial and three verbal factors were obtained for the high ability students, while one verbal and three spatial factors were obtained for the lows. (d) Factors were generally more interpretable and congruent with other factor analytic work for the high ability sample than for the low ability sample.

DISTRIBUTION LIST

Navy

1 Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 DR. JOHN F. BROCK
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 MR. MAURICE CALLAHAN
Pers 23a
Bureau of Naval Personnel
Washington, DC 20370

1 CAPT. H. J. CONNERY
NAVY MEDICAL R&D COMMAND
NNMC
BETHESDA, MD 20014

1 Dr. Richard Elster
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 CDR John Ferguson, MSC, USN
Naval Medical R&D Command (Code 44)
National Naval Medical Center
Bethesda, MD 20014

1 Dr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 CAPT. D.M. GRAGG, MC, USN
HEAD, SECTION ON MEDICAL EDUCATION
UNIFORMED SERVICES UNIV. OF THE
HEALTH SCIENCES
6917 ARLINGTON ROAD
BETHESDA, MD 20014

1 MR. GEORGE N. GRAINE
NAVAL SEA SYSTEMS COMMAND
SEA 05L1C2
WASHINGTON, DC 20362

1 CDR Robert S. Kennedy
Naval Aerospace Medical and
Research Lab
Box 29407
New Orleans, LA 70189

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

1 CHAIRMAN, LEADERSHIP & LAW DEPT.
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin
USS Francis Marion (LPA-Z49)
FPO New York, NY 09501

1 Dr. James McBride
Code 301
Navy Personnel R&D Center
San Diego, CA 92152

1 DR. WILLIAM MONTAGUE
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Commanding Officer
U.S. Naval Amphibious School
Coronado, CA 92155

1 Commanding Officer
Naval Health Research
Center
Attn: Library
San Diego, CA 92152

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

1 Library
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 JOHN OLSEN
CHIEF OF NAVAL EDUCATION &
TRAINING SUPPORT
PENSACOLA, FL 32509

1 Psychologist
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 200
Arlington, VA 22217

1 Office of Naval Research
Code 441
800 N. Quincy Street
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
OFFICE OF NAVAL RESEARCH BRANCH
223 OLD MARYLEBONE ROAD
LONDON, NW, 15TH ENGLAND

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Scientific Advisor to the Chief of
Naval Personnel (Pers-Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 LT Frank C. Petho, MSC, USNR (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 DR. RICHARD A. POLLAK
ACADEMIC COMPUTING CENTER
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Roger W. Remington, Ph.D
Code L52
NAMRL
Pensacola, FL 32508

1 Dr. Bernard Rimland
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 A. A. SJOHOLM
TECH. SUPPORT, CODE 201
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr. Robert Smith
Office of Chief of Naval Operations
OP-987E
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis & Evaluation Group (TAEG)
Dept. of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 CDR Charles J. Theisen, JR. MSC, USN
Head Human Factors Engineering Div.
Naval Air Development Center
Warminster, PA 18974

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

1 Dr. Ronald Weitzman
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

1 DR. H.M. WEST III
DEPUTY ADCNO FOR CIVILIAN PLANNING AND PROGRAMMING
RM. 2625, ARLINGTON ANNEX
WASHINGTON, DC 20370

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

Army

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 LCOL Gary Bloedorn
Training Effectiveness Analysis Division
US Army TRADOC Systems Analysis Activity
White Sands Missile Range, NM 88002

1 DR. RALPH DUSEK
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Col. Frank Hart, Director
Training Development Institute
ATTNG-TDI
Ft. Eustis, VA 23604

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Milton S. Katz
Individual Training & Skill Evaluation Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Beatrice J. Farr
Army Research Institute (PERI-OK)
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milt Maier
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Harold F. O'Neil, Jr.
ATTN: PERI-OK
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 LTCOL Michael T. Plummer
Organizational Effectiveness Division
Office of the Deputy Chief of Staff for Personnel
Department of the Army
Washington, DC 20301

1 Dr. Robert Sasmor
U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air Force Human Resources Lab
AFHRL/PED
Brooks AFB, TX 78235

1 Air University Library
AUL/LSE 76/443
Maxwell AFB, AL 36112

1 DR. T. E. COTTERMAN
AFHRL/ASR
WRIGHT PATTERSON AFB
OHIO 45433

1 Dr. Philip De Leo
AFHRL/TT
Lowry AFB, CO 80230

1 DR. G. A. ECKSTRAND
AFHRL/AS
WRIGHT-PATTERSON AFB, OH 45433

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 CDR. MERCER
CNET LIAISON OFFICER
AFHRL/FLYING TRAINING DIV.
WILLIAMS AFB, AZ 85224

1 Dr. Donald E. Meyer
U.S. Air Force
ATC/XPTD
Randolph AFB, TX 78148

1 Dr. Ross L. Morgan (AFHRL/ASR)
Wright -Patterson AFB
Ohio 45433

1 Dr. Roger Pennell
AFHRL/TT
Lowry AFB, CO 80230

1 Personnel Analysis Division
HQ USAF/DPXXA
Washington, DC 20330

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

1 Dr. Malcolm Ree
AFHRL/PED
Brooks AFB, TX 78235

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Jack A. Thorpe, Capt, USAF
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Brian K. Waters, LCOL, USAF
Air University
Maxwell AFB
Montgomery, AL 36112

Marines

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Director, Office of Manpower Utilization
HQ, Marine Corps (MPU)
BCB, Bldg. 2009
Quantico, VA 22134

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

CoastGuard

1 Mr. Richard Lanterman
PSYCHOLOGICAL RESEARCH (G-P-1/62)
U.S. COAST GUARD HQ
WASHINGTON, DC 20590

1 Dr. Thomas Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Other DoD

1 Dr. Stephen Andriole
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

1 Military Assistant for Training and Personnel Technology
Office of the Under Secretary of Defense for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

1 Mr. Fredrick W. Suffa
MPP (A&R)
2B269
Pentagon
Washington, D.C. 20301

Civil Govt

1 Dr. Susan Chipman
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Lorraine D. Eyde
Personnel R&D Center
U.S. Civil Service Commission
1900 EStreet NW
Washington, D.C. 20415

11 Mr. James H. Ferstl
Bureau of Training
U.S. Civil Service Commission
Washington, D.C. 20415

1 Dr. Joseph I. Lipson
Division of Science Education
Room W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 William J. McLaurin
Rm. 301, Internal Revenue Service
2221 Jefferson Davis Highway
Arlington, VA 22202

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev. and Research
National Science Foundation
Washington, DC 20550

1 Dr. Jeffrey Schiller
National Institute of Education
1200 19th St. NW
Washington, DC 20208

Civil Govt

1 Dr. Thomas G. Sticht
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Frank Withrow
U. S. Office of Education
400 6th Street SW
Washington, DC 20202

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Non Govt

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Dr. John R. Anderson
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 DR. MICHAEL ATWOOD
SCIENCE APPLICATIONS INSTITUTE
40 DENVER TECH. CENTER WEST
7935 E. PRENTICE AVENUE
ENGLEWOOD, CO 80110

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. R. A. Avner
University of Illinois
Computer-Based Educational Research Lab
Urbana, IL 61801

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

1 Dr. Patricia Baggett
Department of Psychology
University of Denver
University Park
Denver, CO 80208

1 Ms. Carole A. Bagley
Minnesota Educational Computing
Consortium
2520 Broadway Drive
St. Paul, MN 55113

1 Dr. John Bergan
School of Education
University of Arizona
Tuscon AZ 85721

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. Lyle Bourne
Department of Psychology
University of Colorado
Boulder, CO 80302

1 Dr. Kenneth Bowles
Institute for Information Sciences
University of California at San Diego
La Jolla, CA 92037

1 Dr. John Brackett
SofTech
460 Totten Pond Road
Waltham, MA 02154

1 Dr. John S. Brown
XEROX Palo Alto Research Center
3333 Coyote Road
Palo Alto, CA 94304

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. Anthony Cancelli
School of Education
University of Arizona
Tuscon, AZ 85721

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. William Chase
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Micheline Chi
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. John Chiorini
Litton-Mellonics
Box 1286
Springfield, VA 22151

1 Dr. William Clancey
Department of Computer Science
Stanford University
Stanford, CA 94305

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138

1 Dr. Meredith Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L Street N. W.
Washington, DC 20037

1 Dr. Ruth Day
Center for Advanced Study
in Behavioral Sciences
202 Junipero Serra Blvd.
Stanford, CA 94305

1 Dr. Hubert Dreyfus
Department of Philosophy
University of California
Berkely, CA 94720

1 Dr. Marvin D. Dunnette
N492 Elliott Hall
Dept. of Psychology
Univ. of Minnesota
Minneapolis, MN 55455

1 Dr. A. J. Eschenbrenner
Dept. E422, Bldg. 101
McDonnell Douglas Astronautics Co.
P.O.Box 516
St. Louis, MO 63166

1 MAJOR I. N. EVONIC
CANADIAN FORCES PERS. APPLIED RESEARCH
1107 AVENUE ROAD
TORONTO, ONTARIO, CANADA

1 Dr. Leonard Feldt
Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Mr. Wallace Feurzeig
Bolt Beranek & Newman, Inc.
50 Moulton St.
Cambridge, MA 02138

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Alinda Friedman
Department of Psychology
University of Alberta
Edmonton, Alberta
CANADA T6G 2J9

1 Dr. Vernon S. Gerlach
College of Education
146 Payne Bldg. B
Arizona State University
Tempe, AZ 85281

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ron Hambleton
School of Education
University of Massechusetts
Amherst, MA 01002

1 Dr. Harold Hawkins
Department of Psychology
University of Oregon
Eugene OR 97403

1 Dr. Dustin H. Heuston
Wicat, Inc.
Box 986
Orem, UT 84057

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 DR. LAWRENCE B. JOHNSON
LAWRENCE JOHNSON & ASSOC., INC.
SUITE 502
2001 S STREET NW
WASHINGTON, DC 20009

1 Dr. Wilson A. Judd
McDonnell-Douglas
Astronautics Co. East
Lowry AFB
Denver, CO 80230

1 Dr. Arnold F. Kanarick
Honeywell, Inc.
2600 Ridgeway Pkwy
Minneapolis, MN 55413

1 Dr. Steven W. Keele
Dept. of Psychology
University of Oregon
Eugene, OR 97403

1 Dr. Walter Kintsch
Department of Psychology
University of Colorado
Boulder, CO 80302

1 Dr. David Kieras
Department of Psychology
University of Arizona
Tuscon, AZ 85721

1 Dr. Stephen Kosslyn
Harvard University
Department of Psychology
33 Kirkland Street
Cambridge, MA 02138

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 LCOL. C.R.J. LAFLEUR
PERSONNEL APPLIED RESEARCH
NATIONAL DEFENSE HQS
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A 0K2

1 Dr. Jill Larkin
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Pk.
Goleta, CA 93017

1 Dr. Richard B. Millward
Dept. of Psychology
Hunter Lab.
Brown University
Providence, RI 82912

1 Dr. Stuart Milner
Department of Education
George Mason University
4400 Fairfax Drive
Fairfax, VA 22030

1 Dr. Allen Munro
Univ. of So. California
Behavioral Technology Labs
3717 South Hope Street
Los Angeles, CA 90007

1 Dr. Donald A Norman
Dept. of Psychology C-009
Univ. of California, San Diego
La Jolla, CA 92093

1 Dr. Melvin R. Novick
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analysis
400 Army Navy Drive
Arlington, VA 22202

1 Dr. Robert Pachella
Department of Psychology
Human Performance Center
330 Packard Road
Ann Arbor, MI 48104

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 Dr. Barbara Pflanz
Department of German
University of Redlands
Redlands, CA 92373

1 Dr. Martha Polson
Department of Psychology
University of Colorado
Boulder, CO 80302

1 DR. PETER POLSON
DEPT. OF PSYCHOLOGY
UNIVERSITY OF COLORADO
BOULDER, CO 80302

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MIN. RET. M. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 161
53 BONN 1, GERMANY

1 Dr. Peter B. Read
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkely, CA 94720

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. David Rumelhart
Center for Human Information Processing
Univ. of California, San Diego
La Jolla, CA 92093

1 PROF. FUMIKO SAMEJIMA
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

1 Dr. Benjamin Schneider
Dept. of Psychology
Univ. of Maryland
College Park, MD 20742

1 DR. WALTER SCHNEIDER
DEPT. OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS
CHAMPAIGN, IL 61820

1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138

1 Mr. William Stobie
McDonnell-Douglas
Astronautics Co.
P. O. Box 30204
Chico, CA 95926

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Brad Sympson
Office of Data Analysis Research
Educational Testing Service
Princeton, NJ 08541

1 Dr. Kikumi Tatsuoka
Computer Based Education Research
Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. Maurice Tatsuoka
Department of Educational Psychology
University of Illinois
Champaign, IL 61801

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 DR. PERRY THORNDYKE
THE RAND CORPORATION
1700 MAIN STREET
SANTA MONICA, CA 90406

1 Dr. Walt W. Tornow
Control Data Corporation
Corporate Personnel Research
P.O. Box 0 - HQN060
Minneapolis, MN 55440

1 Dr. Douglas Towne
Univ. of So. California
Behavioral Technology Labs
3717 South Hope Street
Los Angeles, CA 90007

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Benton J. Underwood
Dept. of Psychology
Northwestern University
Evanston, IL 60201

1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Dr. Karl Zinn
Center for research on Learning
and Teaching
University of Michigan
Ann Arbor, MI 48104