Group Factors in Mental Tests Similar in Material or in Structure

BY

GEORGE MILTON SMITH JR., Ph.D.

ARCHIVES OF PSYCHOLOGY
R. S. WOODWORTH, Editor
No. 156

NEW YORK
July, 1933

ARCHIVES OF PSYCHOLOGY
COLUMBIA UNIVERSITY, NEW YORK CITY
The Subscription price is six dollars per volume of about 500 pages. Volume I comprises Nos. 2-10; Volume II, Nos. 11-18; Volume III, Nos. 19-25; Volume IV, Nos. 26-32; Volume V, Nos. 33-39; Volume VI, Nos. 40-46; Volume VII, Nos. 47-52;

and on General Efficiency: L. I. STECHER. 90c. (Cl., $1.15.)

78. Aggressive Behavior in a Small Social Group: E. M. RIDDLE. $1.75.

(Continued on inside back cover.)

Group Factors in Mental Tests Similar in Material or in Structure

BY

GEORGE MILTON SMITH JR., Ph.D.

ARCHIVES OF PSYCHOLOGY

R. S. WOODWORTH, Editor

No. 156

NEW YORK

July, 1933

TABLE OF CONTENTS

CHAPTER I

PURPOSE AND LITERATURE

1. Purpose[1]

Perhaps the most fruitful approach to an understanding of mental organization has been through an analysis of the intercorrelations of tests designed to measure various mental abilities. The question of mental organization is not merely one of theoretical interest but has important practical implications; for the practice of intelligence test construction depends, or should depend, upon the theory. Theories of mental organization, including Spearman's two-factor theory (20, Chap. 13), have had to reckon with so-called group factors present in tests which presumably evoke similar mental abilities. Furthermore, those who have striven to achieve internal consistency in test batteries designed to measure unitary traits have had to contend with group factors (4, p. 354). The all too common practice of using total scores for test batteries, without making allowance for possible group factors, has tended to lessen their significance and befuddle their interpretation.

Strictly speaking, group factors are purely statistical entities. They take on theoretical significance only in so far as they are psychologically plausible. In general, group factors, when present, have been attributed to similarity of content, or material; that is, attributed to the fact that the tests in a given group were all verbal, or all numerical, or all spatial, etc. With few exceptions, the form, or structure, of the tests has been held to be of little importance or has been wholly ignored.

It is the purpose of this investigation to determine whether the structure of a mental test, as well as the material, may not be important in the production of group factors; to determine, for example, whether three tests which deal with different material media (*e.g.*, verbal, spatial, numerical), but which all have the analogies structure, let us say, are likely to produce group factors. If structure is found to be important, an attempt will be made to get some notion of the relative potency of material and structural factors.[2]

---

[1] This is one of a series of studies under the general direction of Professor H. E. Garrett.

[2] I am most grateful to Dr. S. Asch for sugggestions which broadened the scope of the original problem and for generously relinquishing plans of his own for an investigation in this field.

Without some such knowledge as this, interpretations of test intercorrelations are likely to be confused and theories of mental organization will be based on inadequate data.

Before proceeding further it will be well to attempt a clarification of some of the concepts to be dealt with in this paper. Certain of the concepts in current use in this field have both objective and subjective connotations. In the interest of precision of meaning and of measurement, it is important to objectify concepts wherever possible. With this in view the term "material" has been substituted for the more familiar term "content," to denote the essential character of the test items: verbal, numerical, spatial, and the like; for "content" has commonly been associated, not only with test items, but also with "thought." In this study, verbal, numerical, and spatial materials only will be employed. The term "form" is equally ambiguous. It has been used by Hart and Spearman (12, p. 52) to denote the "kind of mental operation" involved in the solution of a given task, and by Davey (6) and others (3) to denote the form of the test itself (*e.g.*, analogies form, completion form, etc.). Hence, it has seemed desirable to substitute the term "structure," specifically restricting its meaning to the tests themselves. Items of the analogies, generalization,[3] and "construction"[4] structure only will be studied in the present investigation.

The use of objective terms is not intended to imply a behavioristic slighting of "mental processes," but is intended merely to lessen the confusion which inevitably results from the use of subjective terms in a field of such complexity, and to furnish a tangible basis for classifying the tests. One cannot ignore the fact that objective concepts have subjective implications. It is, for example, a reasonable hypothesis that test items of similar structure will to some extent evoke similar mental processes. (One of the theoretical purposes of this study will be to throw light on this hypothesis.) But the structure cannot, of course, be identified with the process. Even similarity of material in test items implies a degree of similarity in the mental operations required for their solution; for there are certain operations which can be carried out only with words and others which can be carried out only with numbers, etc. For example, it would be difficult to work out the $\sqrt{\text{procrastination}}$ or to name the opposite of e^{-2x}.

[3] Sometimes called "classifications"; see Section 2 of Chapter II.

[4] This refers to items which involve the assembling of elements into a whole. It includes "completions" items.

Thus, not only is it impossible wholly to divorce the objective from its subjective implications, but it is impossible to base a subjective distinction between the concepts "material" and "structure" on the presence or the absence of implied mental operations *per se*. Rather, the distinction must be based upon the *nature of* the operations, which are, alike, demanded by the material and by the structural aspects of test items. Furthermore, since every test item has both a material and a structural aspect, we cannot isolate structure and study it in complete independence of material, and *vice versa*, any more than we can study heredity completely divorced from environment. We can, however, "control" one of these aspects and vary the other; and this is what has been attempted, however imperfectly, in the present research.

2. Literature

Consideration of group factors associated with the form, or structure, of mental tests has been rare and usually incidental to other purposes in correlation analyses. Though there are numerous studies in which tests of similar structure or tests which conceivably deal with similar relations have been used, such similarity has generally been fortuitous and the number of tests showing such similarity has been small. Though Spearman (21) and workers in his laboratory (23) have recognized "similarity of form" and "similarity of relations" as potential "disturbers of tetrads" (*i.e.*, factors which may produce tetrad values of disturbing magnitude), significant systematic attempts to study test structure in its own right are almost non-existent.

An early study by Bailor, "Content and Form in Tests of General Intelligence" (3), is apparently a direct attack upon the problem. The study is impressive because of the 1039 subjects employed, 489 boys appearing in an age and sex control group alone, and because of the prodigious number of correlations calculated, raw, corrected, zero order, first order, and second order. But the author worked without benefit of tetrad analysis, the sub-tests had low reliabilities (.25 to .66), the form and content groups were not comparable, and the results were inconclusive. The main conclusion is a negative one: "No adequate evidence has been found to indicate that the results of tests differing in type of content, such as words, numbers, space, and differing in form, such as analogies, completion, and generalization, are duplicate and independent measures of a common mental function." (P. 62.)

Studies which have employed tetrad analysis deserve more serious consideration, though the results are somewhat conflicting. Results which are suggestive for our purposes, though far from conclusive, were obtained by Davey (6), who used a battery of eight "oral"[5] and six pictorial tests.[6] These groups each contained tests of the analogies, the classification,[7] and the completion forms. The subjects were 243 school children of both sexes from seven different schools, with ages ranging from eight to fourteen years. Using the "proportionality formula" (which makes use, apparently, of one tetrad only from each triplet set), and selecting for consideration a limited number of test combinations, she found no evidence of group factors due to similarity of form. However, she concluded that a group factor was present in four of the oral tests, opposites, synonyms, classification, and questions; and that "it must be due to their content, that is, either to the fundaments or to the similarity of the relations educed—." (P. 41.) Note that the "similarity of relations" here refers, *not* to the similarity between the oral and the pictorial classifications tests, but to a possible similarity between the oral opposites and the oral synonyms, between the oral synonyms and the oral classifications, etc. This kind of "similarity of relations" in tests *within the same material field* as a potential cause of large tetrad differences has also been postulated in Stephenson's study (23) and might well have been postulated to account for the group factors found in the study reported by Brigham (4). Before discussing these two studies, however, an earlier study by Spearman may be mentioned.

Using the scores of 2599 adults on both "inventive" and "selective" varieties of each of three different types of test,[6] analogies, opposites, and "passages," Spearman (20, pp. 153*ff*) found marked evidence of overlapping abilities, as he had anticipated. However, when the r's between the two different varieties of the same tests were eliminated from the calculations, the observed distribution of the tetrads was in close agreement with that to be expected from the two-factor theory. The remaining similarity with respect to the "inventive" and "selective" aspects of the tests seemed to give rise to no group factors. The major cause of the overlap was apparently the similarity between the two different forms of the same tests.[8] Any

[5] Apparently all of these tests dealt with verbal material.
[6] No reliabilities were reported.
[7] Similar in type to the "generalizations" tests used in the present study.
[8] Tryon in discussing this work of Spearman (30, pp. 421*f*) observes that many of the tetrads eliminated by Spearman do not fall into the simple patterns

other result would have been surprising; for there was here a dual similarity, a similarity both of structure and of material.

In Stephenson's recent research (22, 23, 24) eight verbal and eight non-verbal (mostly spatial) tests[9] were given to 1037 girls, between 8½ and 14 years of age, from 11 different schools. A test of the analogies structure was present in each battery. The "specificality" within the non-verbal group (indicated by the excess magnitude of the tetrads, beyond that to be expected from sampling errors) was attributed to a "speed preference" factor present in two of the variables. Within the verbal group "similarity of relations" in the opposites, analogies, and classification[10] tests (relations of "likeness" or of "unlikeness") was postulated as a cause of large tetrads. Elimination of the relevant r's produced a very slight reduction in the P.E. of the observed tetrads, .0194 to .0181; and yet Stephenson concludes: "Of the various suggested disturbers of tetrads, that of 'similarity of relations' for the Analogies, Opposites and Classification subtests appears to be the most discernible." (P. 266). When four of the non-verbal tests were studied in conjunction with seven of the verbal tests, somewhat tenuous evidence of "specificality" between the two analogies tests was educed. The elimination of the tetrads involving the r between these two tests reduced the value of the observed P.E. for cross-tetrads of a certain type from .0194 to .0191, and for another type from .0144 to .0138. This reduction is characterized as a "noticeable diminution." Here, then, we have seen for the first time evidence (albeit somewhat meager) for a "structure factor" between two *tests which differ in material.*[11] However, judging by the effect on the tetrads, this factor was less potent than the factor due to "similarity of relations" in tests couched *in the same material idiom.*

---

that we should expect if the group factors postulated (those due to similarity of form) were the only ones present. In fact, he states that group factors, by Spearman's rule (tetrads > 5 $P.E._t$), are not even suggested. However, if we apply a less rigid criterion than Spearman's, and if we apply Tryon's own principles of tetrad analysis to the tetrads that Tryon has calculated from Spearman's data, we find very strong evidence to support Spearman's postulation of group factors between the two varieties of the analogies test and the two varieties of the "passages" test. Tryon's data are not sufficient to warrant a judgment on the completion test. But, as Tryon insisted, other factors, in addition to the ones postulated by Spearman, are also strongly suggested.

[9] Reliabilities not reported; the number of items ranged from 10 to 26; the testing times, from 2½ to 5 minutes.

[10] Similar in type to the "generalizations" tests of this study.

[11] See also the tetrad analysis of E. E. Cureton (14, pp. 203–205) based on the data collected by Rogers (18) from 61 school girls. He found rather sketchy evidence for a factor involving "thinking in terms of antitheses" in a verbal opposites and a spatial symmetry test.

More convincing evidence of group factors in tests involving similar relations appears in an impressive study reported by Brigham (4, esp. pp. 365, 366, from the Fourth Annual Report of the Chairman of the Commission on Scholastic Aptitude Tests, 1929). The findings are based on the scores of 4354 boy and 3061 girl candidates for college on seven verbal tests of high reliability (.85 to .91). Group factors were located by the use of mean tetrads.[12] Large mean tetrads, derived in every case but one from the maximum number of positive tetrads, were found for the following test pairs: synonyms and antonyms, easy paragraph reading and difficult paragraph reading, classifications[13] and analogies. The author concludes that there is here "community of function over and above that ascribable to a general factor which did appear—." To account for this excess "community of function" two hypotheses are ventured: (1) a factor which "might possibly be of the nature of rushing through the tests" (*Cf.* Stephenson's "speed preference."); and (2) a factor "due to the kind of material entering into the tests." With regard to the synonym-antonym factor the author further remarks: "The conclusion that synonyms and antonyms are really psychologically the same function has been drawn by various people."[14] The easy-difficult paragraph reading factor demands no explanation. But Brigham does not feel that the analogies-classifications factor has been adequately explained by either of his hypotheses; he says: "Both of these explanations leave vague the cause of the large mean tetrads for the analogies and classifications." The "similarity of relations" of Stephenson, Davey, *et al.,* is strongly suggested. Though the statistical adequacy of this research cannot well be questioned, it fails to throw much light on the present problem; for it deals only with verbal material and with but single representatives of various structural types.

Further evidence for group factors based on "community of function" in verbal tests appears in the recently published work of Findley (8). Using later forms of three of the Scholastic Aptitude Tests dealt with in Brigham's report, antonyms, definitions, and "paragraphs," in conjunction with three "Scientific Reading Tests," antonyms and two different "paragraphs" tests, with 369 applicants for The Cooper Union Institute of Technology as subjects, he found "evidence that similarity of form brings a group factor"

[12] See discussion of mean tetrads in Section 4 of Chapter III below.
[13] Similar to the "generalizations" tests of this study.
[14] *Cf.* Schneck (19).

(p. 22), between the three "paragraphs" tests and between the two antonyms tests. (The evidence was better for the paragraphs factor than for the antonyms factor, though the author refused to draw this conclusion.) He also found evidence for group factors due to similarity of "content" in both "scientific" and "literary" vocabularies. On the basis of two tetrads,[15] Findley ventures a judgment as to the relative strength of the form and content factors in his tests. He concludes that, in addition to the community of function evidenced throughout the entire battery of verbal tests, "specialization of content in respect to 'scientific' and 'literary' vocabulary to the extent employed here is probably of less significance than the form of the verbal function tested." (P. 23.)

In connection with attempts to contrast the importance of form and content, or material and structure, a study by Dockerill *et al.*, briefly reported by Spearman (20, p. 240*f*), may be mentioned. Two tests "perfectly similar in substance or the essential nature of the tasks to be executed" were constructed, one in the completions form and the other in the questions form. No group factor was introduced by the similarity of substance. Hence, it was concluded that form was more effective than substance. This jibes with Findley's conclusion; but it conflicts with Stephenson's findings. On the basis of work in the field of memory, Anastasi has also ventured the opposite conclusion: "material is more potent that method[16] in determining inter-test correlation." (2, pp. 45, 54*ff*.)

Summary

From the point of view of the present problem, the studies reviewed have suffered from one or more of the following defects:

(1) Test reliabilities were either low or not reported, or both. (2) Subjects were either too few or too heterogeneous. (3) There were not enough variables in the entire battery, or there were not enough representatives of either the material or the structural types (often only one of most types). (4) When more than one variable of two different types was present, no attempt to counterbalance the competing influences of group factors was made.

In spite of the prevalence of some of these defects, the evidence has rather consistently suggested that "similarity of relations" *within the same material* (*or content*) *field* may be a basis for group

[15] Throughout the study Findley has made an admirable attempt to "counterbalance" the influences of competing group factors in selecting crucial tetrads.

[16] Both recognition and recall methods were used.

factors. (See studies of Davey, Stephenson, Spearman, and Brigham.) It remains to investigate the potency of similarity of relations with diversified material media. Rather weak evidence has been cited which suggests that, in isolated cases, "similarity of structure," as defined in this study, may produce group factors which "cut across" two material fields. (See studies of Stephenson and Cureton.) It remains to investigate this with more reliable tests and with a greater diversity of material and structural types. The evidence concerning the relative significance of material and structural similarity for the production of group factors is conflicting and much of it based on insufficient data. (See the studies of Findley, Dockerill, Stephenson, and Anastasi.) More light may well be thrown on this question.

CHAPTER II

PROCEDURE

1. The Subjects

The importance of a certain degree of homogeneity (with respect to age, sex, race, social and educational background, etc.) in the subjects of any correlational study has been stressed by Kelley (14, pp. 24*ff*), Garrett and Anastasi (11, pp. 248*ff*), and others. The subjects of this investigation, 186 students in the introductory course in psychology at The College of the City of New York, met most of the requirements of homogeneity in an unusually satisfactory manner. This group of subjects had the further advantage of being closely similar to the groups used by Schneck in his analysis of verbal and numerical abilities (19) and by Anastasi in her analysis of certain mnemonic abilities (1). The degree of homogeneity is indicated by the following summary of characteristics:

1. *Age.* The average age was 19 years and 5 months, with a standard deviation of 16 months.

2. *Sex.* All the subjects were male.

3. *Race.* 94% of the subjects were Jewish.[17]

4. *Nationality.* 88%[18] of the subjects were born in the United States. 92% of their fathers and 89% of their mothers were born abroad; 45% of them in Russia, 24% in Austria, and 18% in Poland.

5. *Language spoken in the home.* In 84% of the cases some language other than English was spoken in the home; in 35% of these cases "a little," in 34% "about half the time," and in 15% "most of the time." The other language was Yiddish in 86% of the cases.

6. *Educational status and background.* (a) *Class in college.* 80% of the subjects were in either the Upper Sophomore or the Junior Class. (b) *Pre-college training.* 94% of the subjects had spent at least 6 years, and 75% at least 9 years, in the schools of New York City.

---

[17] The 11 subjects that were not Jewish obtained scores that were fairly typical for the group: even the *least* typical of these received scores within 1.3 sigma of the average on *all but two* of the 14 tests.

[18] This percentage and those that follow, with the exception of those dealing with the fathers' occupations, were estimated from a sample of 100 questionnaires.

7. *Fathers' education and occupation.* (a) *Education.* 25% had had the equivalent of high school training and 5% a college training. (b) *Occupation.* About 5% were professional men; about 75% were merchants, shopkeepers, clerks, tailors, or salesmen; and about 20% were laborers, mostly skilled.

2. The Tests

Ideally, in a study of this sort, a wide range of material media and structural patterns should be employed; each type should be represented by a number of tests; and each test should fit neatly into just one material group and also into just one clear-cut structural pattern. There are, however, a limited number of material media and structural patterns which are sufficiently familiar to be used in general testing and which will meet the practical requirements of test construction and administration. Hence, the battery of fourteen tests constructed for this study is at best a rough approximation to the ideal. It does, however, contain at least four representatives of each of three material media, namely, four numerical, four spatial, and five verbal tests; and at least four representatives of each of three structural patterns, namely, four analogies, four generalizations,[19] and five "construction"[20] tests. With two exceptions, each test falls naturally into one of the three material groups and also into one of the three structure groups. The interrelationships of the tests are indicated in the schema below:

N_a	N_g	N_c	N_x	
S_a	S_g	S_c	S'_c	
V_a	V_g	V_c	V'_a	V'_g
(V'_a)	(V'_g)	(S'_c)		
		X_c		

The capital letters, N, S, V, denote the material groups to which the tests belong, *N*umerical, *S*patial, and *V*erbal, respectively; the subscripts, a, g, c, denote the structure groups, *a*nalogies, *g*eneralizations, "*c*onstruction." The parentheses indicate that the test has been listed elsewhere with another group; a prime indicates that there is some other test which is similar both with respect to material and with respect to structure. N_x denotes the Arithmetic Reasoning Test, which has no clear-cut structure; and

---

[19] Sometimes called "classifications." See examples of tests below.

[20] This implies the putting together of given elements into a whole. It includes tests of the "completion" type.

X_c denotes the Anagrams Test, which is of "completion" structure but does not readily fit into any of the three material groups.

The nine tests in the upper left-hand part of the figure are of especial interest because of the neat balance which exists between the various groups: there are just three representatives of each of the three material groups and just three representatives of each of the three structure groups; and each test appears once and only once in a material rôle and in a structure rôle. This special group of nine furnishes a rough sort of "control" for contrasting the material groups with each other and with the structure groups, etc. (See Section 6 of Chapter III.) The other tests have been added to this ideally balanced group of nine in order to facilitate tetrad analysis. It is unfortunate that a battery of sixteen variables could not have been selected which would have maintained this balance on the basis of four groups of each type and four variables per group; but this was not feasible.

In selecting the tests, it was relatively easy to find suitable tests for the three material groups, but to find tests which would *at the same time* meet the structural requirements was not a simple task. Hence, in order to have tests which met the dual requirements of material and structure, and in order to insure structural similarity, ten of the fourteen tests were constructed by the experimenter, in some cases *de novo*. The greatest care was taken with the analogies and the generalizations groups; for no very sanguine expectations of positive results from the relatively amorphous "construction" group were entertained. With the exception of the Verbal and Grammatical Analogies Tests and the Verbal Generalizations Test (for which about half the items were selected from outside sources), almost all of the analogies and generalizations items were new,[21] and in the case of the Numerical Analogies Test the entire plan was new, it is believed. This large number of untried items and new test forms demanded considerable preliminary experimentation. For this work some 250 subjects[22] were used, about 40–60 on each of the tests. The results were used to weed out ambiguous

---

[21] To my wife, Frances Tappan Brown, I am grateful for her assistance in inventing some of the verbal items, in correcting tests, in tallying scores, in reading proof, and in countless other matters, ponderable and imponderable, throughout this study.

[22] Students from the course in general psychology in Brooklyn College, through the kindness of Mr. Austin Wood; students from the course in educational psychology at the College of the City of New York, through the kindness of Dr. H. H. Abelson and Mr. B. Epstein; and students from The School of Civic and Business Administration, through the kindness of Dr. A. Mintz and Mr. W. Vogt.

items, to determine the order of difficulty, and to clarify the instructions. Samples of the tests are given below. With the exception of the Minnesota Paper Form Board, the tests were power tests of some difficulty.

Analogies Tests

For all four analogies tests the instructions and the method of indicating the answers were identical, except for the variations in the name of the material dealt with. This is indicated in the skeleton set of instructions which follows:—

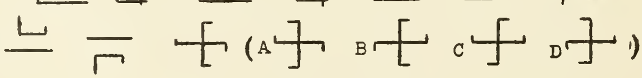
INSTRUCTIONS:—(1) In each line below the first two words (numbers, figures) are related to each other in some way. Pick out the ONE word (number, figure) in parentheses that is related to the THIRD word (number, figure) in the SAME WAY, or MOST NEARLY THE SAME WAY, that the SECOND is related to the FIRST.

(2) Put the capital letter which precedes this word (number, figure) into the blank parentheses at the end of the line, as in the samples.

The sample problems, of course, differed. Typical items from each test are presented below:—

	Answer
Verbal:	
No. 1. catcher team bird (A fly B sky C flock D wing)	(C)
No. 30. plumber pipes hammer (A carpenter B saw C nails D house)	(C)
Numerical:	
No. 1. 64 8 169 (A 11 B 13 C 26 D 96)	(B)
No. 30. 3 1/27 2 (A 1/4 B 0.125 C 1/16 D 0.25)	(B)
Grammatical:	
No. 1. nativity natively diversity (A diversiveness B diversely C diverse D diversily)	(B)
No. 31. I us he (A his B they C them D we)	(C)
Spatial:	

No. 9.  (D)

No. 34.  (B)

Generalizations Tests

As with the analogies tests, the instructions, except for the variations in the name of the material used, were identical. A skeleton set of instructions follows:—

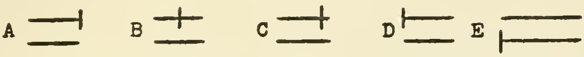
INSTRUCTIONS:—(1) In each line below, pick out the THREE words (numbers, figures) which have SOME SPECIFIC QUALITY IN COMMON, or the MOST SPECIFIC QUALITIES IN COMMON.

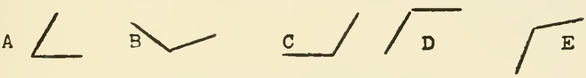
(2) The "qualities in common" which you select must apply to THREE AND ONLY THREE of the words (numbers, figures).

(3) Put the capital letters which precede these three words (numbers, figures) into the parentheses at the end of the line in alphabetical order, as in the samples.

Typical items from each test are given below:—

						Answer
Verbal:						
No. 2.	A mastiff	B bite	C kennel	D poodle	E bloodhound	(ADE)
No. 28.	A youth	B reporter	C sapling	D cub	E bear	(ACD)
Numerical:						
No. 1.	A 3	B 15	C 16	D 21	E 8	(ABD)
No. 25.	A 9	B 27	C 45	D 81	E 24	(ABD)
Grammatical:						
No. 1.	A open	B door	C broken	D buildings	E enter	(ACE)
No. 31	A damn	B sky	C heavens	D blue	E oh	(ACE)
Spatial:						

No. 3.  (ADE)

No. 30.  (ACD)

These few samples from the analogies and generalizations groups obviously can give no adequate notion of the breadth of the tests. Suffice it to say, that in constructing these tests an effort was made to use items which dealt with a wide range of relationships. And, likewise, with the exception of the Anagrams Test and possibly the Minnesota Paper Form Board Test, the tests described below show a satisfactory breadth.

"Construction" Tests

As has been suggested, the tests in this group do not have the precision of structural pattern that characterizes the analogies and generalizations tests. They are, however, roughly similar in structure in that each involves the putting together of elements into a whole; hence, the generic name "construction." The group includes:—

(1) *A Sentence Completion Test.*[23] This was of the selective type, which made more objective scoring possible. (2) *A Number Series Completion Test*[23] of more than the usual breadth. This was of the inventive type. (3) *The Minnesota Paper Form Board Test,*[24] which requires the subject to indicate with pencil lines how several small two-dimensional figures may be fitted into certain larger areas. This was the only one of the fourteen tests in which the speed element was prominent. (See Anastasi's analysis of the speed element in this test given to a closely similar group: 1, pp. 37*f.*) (4) *The Modified Kelley Spatial Test.* The spatial power

---

23 These tests were composed of items from a number of sources assembled by Dr. S. Asch for the Columbia University Psychology Department.

24 Form A only was used. This test is reproduced in 16, pp. 96–101.

test described in Kelley's "Crossroads" (14, pp. 178*f*, 190) was modified[25] so as to be very much more difficult and so that the scoring could be entirely objective. The task involved is similar to that in The Minnesota Paper Form Board; but the ability to determine (in imagination) which of certain elementary shapes need to be turned over (not merely rotated in the same plane) in order to fit the given larger figures was also required. This was the most difficult of the tests. (5) An *Anagrams Test.* This was a supernumerary, intended as a substitute in case any of the other "construction" tests turned out to be unreliable. The words employed were all familiar; difficulty was attained by increasing the number of letters.[26]

The Numerical Tests

In addition to the Numerical Analogies, the Numerical Generalizations, and the Number Series Completion tests already mentioned, this group included an Arithmetic Reasoning Test[27] of some difficulty. Though most of the problems could have been solved by arithmetical means, simple algebraical techniques were required for some.

The Spatial Tests

This group comprised the Spatial Analogies, the Spatial Generalizations, the Minnesota Paper Form Board, and the Modified Kelley Spatial tests, all of which have been described above.

The Verbal Tests

Included in this group were the Verbal Analogies, the Grammatical Analogies, the Verbal and the Grammatical Generalizations, and the Sentence Completion tests, all noted above.

The number of items in each test is given in the Testing Schedule in the next section.

3. The Testing Procedure

The fourteen tests in their final form were administered to the entire class in general psychology at The College of the City of New York, about 350 students, either by the experimenter (one of

[25] With the kind permission of Professor Kelley.

[26] About half of the items were taken from Foster and Tinker's "Experiments in Psychology"; the rest appear in Webster.

[27] Assembled by Dr. S. Asch for the Columbia University Psychology Department.

the instructors) or by the other regular instructors.[28] The tests were introduced as "intelligence tests," and the students were promised their scores. At the same time, they were assured that their scores would in no way affect their grades in the course; so the temptation to seek inspiration from without was minimized. There was every indication that the students, with very few exceptions,[29] accepted the tests as a challenge to their intellectual powers and put forth their best efforts. The four analogies tests were administered by the experimenter (and an assistant[30]) in lecture sections of approximately 80, 120, and 150 students. All other tests were given in the 20 laboratory sections of about 20 students each. In most cases the experimenter gave the initial general instructions and got the tests under way. Where this was impossible, carefully prepared duplicate instructions were used by experienced administrators.[28] Specific instructions appeared in print on each test paper. Altogether eight classroom hours were required. These were spread over a period of about eight weeks, in order to eliminate ennui effects. The schedule follows:

Testing Schedule

Period	*Test*	*Time* (in minutes)	*Number of Items*
I	Sentence Completion	42	36 (40)*
II	Number Series Completion	25	40 (52)
	Minnesota Form Board	15	56
III	Verbal Analogies	16	50
	Numerical Analogies	26	38 (50)
IV	Grammatical Analogies	16	50
	Spatial Analogies	24	48
V	Arithmetic Reasoning	25	30 (40)
	Anagrams	18	36 (50)
VI	Verbal Generalizations	19	50
	Numerical Generalizations	27	38 (46)
VII	Grammatical Generalizations	19	48
	Spatial Generalizations	24	36 (48)
VIII	Modified Kelley Spatial	40	46

Total time: 5 hours, 36 minutes.

* Figures in parentheses indicate the number of items which appeared on the test; but the other figures are the number of items actually used in scoring. See Section 4.

[28] Professor H. D. Marsh, Dr. J. G. Peatman, and Mr. W. Vogt, to whom I am much indebted for their friendly co-operation.

[29] In general, the few students who did not co-operate were among those whose records were incomplete; hence, their scores did not enter into the final data.

[30] Mr. I. Glouberman, for whose kindness in reading a portion of the MS. I am most grateful.

4. The Scoring

The scoring was done by three responsible persons; much of it was checked. There is reason to believe that a high degree of accuracy was achieved. Because the crowded testing schedule did not in every case leave time for all the subjects to complete the tests, seven of the tests were scored on the basis of less than the full number of items.[31] This was done in order to eliminate, as much as possible, the speed element in the tests, and, thus, to make them more nearly pure power tests. Before adopting this procedure with any of the tests, however, a careful check was made[32] to make sure that the distributions would not be seriously skewed or the reliabilities seriously lowered. As a result of this procedure, approximately 80%[33] of the subjects tackled all but two of the items from which the scores were derived on every test, with the exception of the Minnesota Form Board, the Modified Kelley Spatial, and the Arithmetic Reasoning tests. The Minnesota test is admittedly largely a speed test. (See p. 17.) It was not deemed necessary to shorten the Modified Kelley Test, because the speed and power scores were found to correlate .96 with each other. (See note[36], p. 22.) The Arithmetic Reasoning Test could not have been further shortened without seriously skewing the distribution of scores. As scored, 56% of the subjects finished all but two of the items. Since this was a power test of some difficulty, it is probable that an opportunity to finish all the items would not have affected the relative standing of the subjects appreciably.

---

31 The tests shortened are indicated in the Testing Schedule on p. 19 above.

32 On the basis of a sample of 80 tests.

33 The % of subjects finishing the specified number of items on the various tests ranged between 74 and 88. Median % was 80.

CHAPTER III

RESULTS

1. Evaluation of the Tests

The reliability coefficients, means, standard deviations, and measures of skewness and kurtosis are given in Table I. All values are based on the scores from the entire 186 subjects. The reliability coefficients were obtained by the usual odd-even item correlation,[34] corrected by the Spearman-Brown "prophecy" formula (Garrett, 10, p. 269), except in the case of the Minnesota Form Board Test. Since this test was essentially a speed test (See 1, p. 37.), and since the crowded testing schedule did not permit the use of both forms of the test, the reliability was estimated[35] from

TABLE I
Evaluation of the Tests*

Test	*Reliability Coefficient*	*Mean*	*Standard Deviation*	Sk/σ_{Sk}	*Ku.*
1. Sentence Completion	.7887	22.0054	4.6955	−0.89	.269
2. Number Series Completion	.8886	30.1935	6.1284	−0.01	.306
3. Minnesota Form Board	.8109†	29.8710	7.8728	−1.63	.250
4. Verbal Analogies	.8515	36.2258	6.5847	1.29	.283
5. Numerical Analogies	.8842	26.8710	5.2068	1.66	.280
6. Grammatical Analogies	.7974	36.7204	6.2930	2.84	.260
7. Spatial Analogies	.7534	35.8495	5.2082	4.21	.262
8. Arithmetic Reasoning	.8648	21.9785	5.1336	2.63	.267
9. Anagrams	.8911	22.3656	7.0349	0.43	.253
10. Verbal Generalizations	.7806	35.1774	5.5868	1.62	.265
11. Numerical Generalizations	.8298	26.2581	5.4838	2.57	.300
12. Grammatical Generalizations	.6887	33.5161	5.2593	0.42	.284
13. Spatial Generalizations	.7448	21.0376	4.6623	0.40	.270
14. Modified Kelley Spatial	.9623	22.3065	9.5279	0.05	.271

* The reliability coefficients, means, and standard deviations were computed by The Columbia University Statistical Bureau.
† Estimated from Anastasi's finding. See page 17 above.

[34] In the Sentence Completion Test the order of difficulty of the items for the S's of this study differed considerably from that obtained with the original group on which the test was standardized. It seemed, therefore, not only permissible but requisite to *renumber* the items before correlating the odds and evens. With a sample of 80 papers, this had the effect of changing the reliability coefficient from .68 to .82. A similar procedure was adopted with the Numerical Generalizations Test.

[35] By using in the Spearman-Brown "prophecy" formula N equal to 1/2.

Anastasi's finding for a closely similar population (1, pp. 26*ff*). Nine of the fourteen tests have reliabilities of .80 or better; eleven, of .78 or better. The range of reliabilities is .69 to .96.[36] The standard deviations in all but three of the tests are remarkably uniform in magnitude, falling between 4.7 and 6.6. The range is 4.7 to 9.5. This indicates, on the whole, a desirable uniformity of difficulty. The measure of kurtosis used was $Q/(P_{90}-P_{10})$. For the normal curve (which is described as "mesokurtic") this formula gives the value .263. The standard error of this value is $.2778/\sqrt{N}$ or, for 186 cases, .020. A kurtosis greater than .263 indicates a "platykurtic" curve, one which tends to be rectangular in shape; that is, flattened on top and truncated at the ends. A kurtosis less than .263 indicates a "leptokurtic" curve, one which tends to have a high, thin central peak. (See 15, p. 45 and pp. 75–77.) The most extreme deviations from normal kurtosis appear in Tests 2 and 11, which show a tendency toward platykurtosis; but even in these cases the deviations are well within 3 σ_{Ku}.

The skewness of the tests was determined by the formula $P_{50}-\frac{1}{2}(P_{90}+P_{10})$; the standard error of the skewness by the formula $.519\ (P_{90}-P_{10})/\sqrt{N}$.* (15, pp. 75–77; and 7, p. 112.) It will be seen that only in the case of Test 7 does the Sk/σ_{Sk} exceed 3.0. This is supposed to indicate a significant skewness. However, the exact meaning of "significant skewness" is somewhat obscure. For our purposes the crucial question is: "Does the skewness produce a significant change in the intercorrelations of the tests?" A rough empirical answer to this question may be obtained by working with the most extreme cases of skewness. If the extreme *positive* skewness shown by Test 7 has had any effect on the correlations we should expect it to appear most markedly in connection with Test

[36] The value .9623 obtained for the Modified Kelley Spatial Test may be suspected of being partly spurious; for it will be remembered that a considerable proportion of the S's did not finish all 46 items. However, two checks were made to ascertain the influence of a speed element, and it was found to be negligible. Firstly, the reliability was recalculated from the papers of the 112 S's who finished 30 or more items. This turned out to be almost identical with the value obtained from the entire 186 S's: .9618 as compared with .9623. Secondly, using the papers of 80 S's as a sample, the tests were scored in two ways: (a) on the basis of 34 items, which at least half of the S's had finished, and (b) on the basis of the entire 46 items. These two methods of scoring correlated with each other .968. It seems, then, that finishing the test made little difference in the relative standing of the subjects. That it was a genuine power test for the group involved is further indicated by the fact that on the average about 3/10 of the items were missed, even by the better S's (*i.e.*, by those that finished at least two thirds of the test).

* In this formula Kelley (*15*) gives as a constant .59914; this has been corrected by Dunlap and Kurtz (*7*), who give .51850.

3, which has the greatest *negative* skewness. The distributions of these two tests were artificially made symmetrical (the kurtosis being held constant).[37] This was done with negligible changes in the rank order of the scores and with minimal changes in the absolute values.[38] With the new scores the correlation between these two tests became .470; with the original scores it was .476, a difference of only .006. Though this procedure may not be wholly free from theoretical flaws, the results are certainly suggestive. It seems highly improbably that the degrees of skewness indicated in Table I can have an appreciable effect on the intercorrelations of these tests.

2. Intercorrelations

In Table II are given the intercorrelations of the fourteen tests and also the correlation of each test with age. *Raw* values appear above the diagonal; the r's below have been corrected for attenuation. For most of the analysis that follows it has seemed desirable to use raw r's, which are a measure of actualities, rather than corrected coefficients, which approximate theoretical ideals. Correction for attenuation when reliabilities are large and uniform makes little change in the tetrad differences. When reliabilities are small or of unequal magnitude, the corrections are large or of unequal amount and may distort the original pattern considerably. Though only five of the fourteen reliabilities in this study are below .80, and only three below .78, the range (.69 to .96) is sufficiently large to suggest that raw coefficients are preferable.

It will be seen that the correlations with age are in every case negative, with the single exception of the Anagrams Test, which shows a negligible positive correlation (.069). The median of these correlations is −.13, which is not large enough to be significant. However, the largest negative value is −.272, which is significant, though still small, being 5.7 times its P.E. (P.E. equals .048.) Positive correlations with age would be expected from mental tests given to a younger group (with an age range in which the curve of

[37] For Test 7 the kurtosis was changed from .262 to .264 only; for Test 3, from .250 to .251.

[38] In the case of Test 3, 46% of the scores were unchanged, 77% were either unchanged or changed one point only, and 96% were changed by not more than two points; the maximum change of four points was applied to one score only. In the case of Test 7, 52% of the scores were unchanged, 76% were either unchanged or changed by one point only, 80% were changed by not more than two points; and the maximum change of seven points was made with two scores only.

TABLE II
CORRELATION COEFFICIENTS (PEARSON), RAW AND CORRECTED*

Variable:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Age
1. Sentence Completion	X_1	.4280	.1758	.5334	.1046	.4485	.3462	.4313	.3542	.3309	.3699	.4098	.2574	.3041	−.1669
2. Number Series Completion	.5113	X_2	.2540	.4874	.4416	.3079	.5334	.5773	.3130	.3684	.4367	.4744	.4636	.5050	−.1356
3. Minnesota Form Board	.2198	.2992	X_3	.2625	.2371	.1145	.4760	.3051	.2656	.3061	.2508	.1292	.4809	.6625	−.0180
4. Verbal Analogies	.6509	.5603	....	X_4	.3348	.5308	.5048	.5624	.2719	.5103	.4060	.5195	.3885	.4133	−.2536
5. Numerical Analogies	.1253	.4982	....	.3858	X_5	.3152	.4889	.4279	.3535	.2505	.4754	.3256	.3386	.3921	−.1597
6. Grammatical Analogies	.5655	....	....	.6442	.3754	X_6	.3763	.4210	.3681	.3310	.4143	.5897	.3123	.2104	−.2721
7. Spatial Analogies	.4491	.6519	.6090	.6303	.5990	.4855	X_7	.6017	.3205	.3473	.4986	.3475	.4929	.5846	−.2711
8. Arithmetic Reasoning	....	.6586	....	....	.4893	....	....	X_8	.4072	.3726	.4893	.4568	.4316	.4620	−.2539
9. Anagrams	.4225	.3517	.3125	....	....	....	....	....	X_9	.2665	.3979	.3441	.3476	.3954	.0691
10. Verbal Generalizations	.4217	.4423	....	.6259	.3015	.4195	.4529	....	....	X_{10}	.4103	.3772	.4719	.4434	−.0411
11. Numerical Generalizations	.4573	.5086	....	.4830	.5550	....	.6306	.5776	....	.5098	X_{11}	.4631	.4900	.3965	−.1259
12. Grammatical Generalizations	.5560	....	....	.6784	....	.7958	....	....	....	.5144	.6126	X_{12}	.3611	.2818	−.1886
13. Spatial Generalizations	.3358	.5699	.6188	.4878	.4172	....	.6580	....	....	.6169	.6233	.5042	X_{13}	.4868	−.0071
14. Modified Kelley Spatial	.3491	.5461	.7500	.4566	.4251	....	.6866	....	.4270	.5116	.4437	....	.5750	X_{14}	−.0701

* Raw coefficients above the diagonal, corrected coefficients below. The raw r's were calculated by The Columbia University Statistical Bureau.

mental growth has a positive slope) and a group less homogeneous with respect to class in school. Such correlations would indicate that mental ability tends to increase with age, and partialling out age would be legitimate procedure; for in general the younger members of the group would have had higher scores if they had arrived at the average age for the group and the older members would have had lower scores if they had not advanced beyond the average age. In the present group the average age was 19 years and 5 months, with an S.D. of 16 months; that is, all but about 16% of the subjects were at least 18 years old. In this age range the slope of the curve of mental growth has reached, or is rapidly approaching, zero. To interpret the negative correlations here found as indicating that mental ability tends to decrease with age is to suggest an appalling advance in the age of senility.

A more plausible explanation is that any group so largely from the same rung of the academic ladder will tend to include a number of relatively dull older students and a number of relatively bright younger students. To partial out age in such a group would be tantamount to raising the scores of the old-dull group, implying that they had seen better days in their youth (in particular, when they were at the average age of the group), and lowering the scores of the young-bright group, implying that they would never exhibit the same ability again. In light of our knowledge of the curve of mental growth and decline, this is absurd. The differences in scores in such a group as this seem to be due to intrinsic differences in ability, the apparent dependence on age being spurious. Hence, it has been thought unwise to partial out age.

Returning to Table II, the most striking fact revealed by a cursory examination of the raw intercorrelations is their uniform magnitude: 69% of these coefficients fall between .300 and .499, and 93% between .200 and .599; their range is .105 to .663. The possibility of a general factor of some consequence running through the entire battery is suggested. This will be considered later.

In order to strengthen any hypotheses as to the probable loci of independent group factors a rough preliminary analysis of the correlation coefficients is usually made. This consists in comparing the average r's for all possible *intra-group* pairs of variables (*i.e.,* pairs *within* the groups of variables which *seem* to have something in common) with the average r's for the *cross-group* pairs (*i.e.,* pairs consisting of one variable from each of the two groups to be contrasted). This exploratory procedure applied to the raw r's from Table II yields little to encourage further analysis. However, the results are given in order to emphasize the value of the more refined methods used below which have given good results:

Group	*Av. of Intra-Group r's*†	*Range**	*Groups Contrasted*	*Av. of Cross-Group r's*†	*Range**
Material Groups					
Numerical	.48	.15	Num.-Spat.	.42	.30
Spatial	.53	.19	Num.-Verb.	.39	.46
Verbal	.46	.26	Spat.-Verb.	.32	.39
Structure Groups					
Analogies	.43	.22	Anals.-Gens.	.41	.34
Generalizations	.43	.13	Anals.-''Cons.''	.36	.48
''Construction''	.37‡	.49‡	Gens.-''Cons.''	.36	.36

* When the difference between the averages is not large, the range may be the most significant indicator; for, the smaller the range the more uniform the r's and the better the chances of satisfying the tetrad criterion.

† Since this whole preliminary procedure is rough at best, converting the r's into Fisher's z-scale units (9, pp. 175–189) has been thought unnecessary.

‡The dubious nature of the ''Construction'' group is emphasized by these values.

In the above figures there is a bare suggestion of a ''cleavage'' between the verbal and the spatial groups; for the average of the intra-group r's is somewhat higher than the average of the cross-group r's, and the ranges are narrower for the intra-group r's. It so happens that the more searching methods which have been employed in connection with what follows show a marked cleavage between these two groups. But predictions claimed on the basis of such meager evidence as appears in this table usually turn out to be *ex post facto* predictions.

In spite of this unpromising beginning let us see what tetrad analysis will reveal.

3. Simple Tetrad Analysis

The tetrad differences for each of the three material groups (numerical, spatial, verbal) and for each of the three structure groups (analogies, generalizations, construction) are given in Table III. It appears that the ordinary tetrad criterion for the presence of a general factor and uncorrelated specific factors is not satisfied for any of these six groups *as a whole* (unless a very lax interpretation is adopted); though for three of the verbal combinations of four (2, 3, and 4) and for one, or possibly two, of the ''construc-

TABLE III

TETRAD DIFFERENCES FOR THE VARIOUS MATERIAL AND STRUCTURE GROUPS

Nature of Groups	*Variables*	t_{1234}*	t_{1243}	t_{1342}	*Tryon's Type*†	*Probable Group Factors*†
			Material Groups			
Numerical	2 5 8 11	−.0584	.0292	.0876	– – – – *or* B	Gen. "N" factor (?); *or* 2–8 or 5–11 (?)
Spatial	3 7 13 14	−.0494	−.0948 (±.0249)‡	−.0454	– – – – *or* CCB	Gen. "S" factor (?); *or* 3–14, 7–13, 3–13; or 3–14, 7–13, 7–14 (?)
Verbal (1)	1 4 6 12	.0816	.0970 (±.0263)‡	.0155	A	1–4 or 6–12
Verbal (2)	1 4 6 10	−.0523	.0009	.0532	B *or* – – – –	1–6 or 4–10; *or* Gen. "V" factor (?)
Verbal (3)	1 4 10 12	.0293	−.0079	−.0372	– – – –	Gen. "V" factor
Verbal (4)	1 6 10 12	−.0260	.0335	.0595	– – – – (*or* B)	Gen. "V" factor; (*or* 1–10 or 6–12 ??)
Verbal (5)	4 6 10 12	−.1007	.0283	.1290	B	4–10 or 6–12
			Structure Groups			
Analogies	4 5 6 7	−.1335	−.0331	.1004	B	4–6 or 5–7
Generalizations	10 11 12 13	−.0367	−.0704	−.0337	– – – – *or* CCB	Gen. "G" factor (?); *or* 10–13, 11–12, 10–12; or 10–13, 11–12, 11–13 (?)
"Construction" (1)	1 2 9 14	−.0096	.0740	.0837	AB (*or* – – –)	1–2, 1–9 or 1–2, 2–14 or 9–14, 1–9 or 9–14, 2–14 (*or* Gen. "C" factor ??)
"Construction" (2)	2 3 9 14	−.1069	−.0337	.0732	B	2–9 or 3–14
"Construction" (3)	1 3 9 14	−.1651	−.0113	.1539	BB *or* B	1–9 and/*or* 3–14
"Construction" (4)	1 2 3 14	.1948	.2063	.0115	AA *or* A	1–2 and/*or* 3–14
"Construction" (5)	1 2 3 9	.0587	.0237	−.0349	– – – – (*or* AC)	Gen. "C" factor; (*or* 1–2, 1–9 or 1–2, 2–3 or 3–9, 1–9 or 3–9, 2–3 ??)

Mean $P.E._t$ is of the order .025.§

* Kelley's notation (14, p. 47).

† For interpretation of Tryon's types see p. 31, below.

‡ Spearman's formula (20, appendix, x).

§ Spearman's formula (20, appendix, xi).

tion'' combinations (5 and possibly 1) it is roughly satisfied. This means that either the similarity which superficially unites the members of these six groups is not an intrinsic similarity (a similarity in the abilities evoked), or there are narrower group factors, tending to unite the variables in groups of two, three, or four, which interfere with the tetrad machinery and make it powerless to reveal such intrinsic similarity running through the larger groups.

A hint as to the plausibility of this second interpretation is given by an examination of the possible first order group factors (that is, factors linking the variables in pairs) listed in the column at the extreme right of the table. For example, in the verbal group a possible bond between variables 6 and 12 is indicated whenever the two appear together in the same combination of four. (From the psychological standpoint this is not a surprising result; for they possess a dual similarity of material: each one is not only a verbal test but it is also a grammatical test). And it will be observed that in combination 3, in which this pair does not occur, the terad differences are most nearly equal to zero. Likewise, in the other combination from which this pair is absent (number 2) the tetrad differences are small. Thus it seems probable that with the complete elimination of petty group factors the tetrad differences for the entire verbal group would approximate zero. A similar situation obtains in the ''construction'' group. A possible bond between variables 3 and 14 (which are outwardly similar not only in structure but in material) is indicated by the tetrad pattern[39] on every possible occasion. And in the two combinations of four from which this pair is absent (1 and 5) the tetrad differences are the smallest.

More convincing evidence for the presence of first order group factors is obtained by calculating the so-called ''mean tetrads,'' which are treated below.

4. Mean Tetrads[40]

The principal function of mean tetrads is to resolve the ambiguity that arises in the interpretation of tetrad patterns by definitely locating first order bonds between certain pairs of variables. They also give suggestive, if not final, evidence as to the strength of such bonds.[41] The explanation of the significance of the mean

---

[39] See Table IV.

[40] Sometimes called ''directed means tetrads.''

[41] T. L. Kelley (14, pp. 80–83*ff*) has made extensive use of mean tetrads in working out factor patterns for batteries of as many as thirteen variables. Preliminary factor values are chosen so as to make the bonds largest for those pairs of variables which give the largest mean tetrads. But see note 51, p. 39, below.

tetrad given by Kelley (14, pp. 66–85*ff*) is here modified and extended to cover other combinations of group factors, the tetrad patterns for which have been suggested by R. C. Tryon (30, pp. 408–416).

Kelley's Proposition 16 (14, p. 69) deals with a common tetrad pattern of ambiguous significance. Let us examine one phase of this pattern. Suppose the six intercorrelations from four variables x_1, x_2, x_3, x_4 yield tetrad values of the type $t_{1234} = t_{1243} > 0$, and $t_{1342} = 0$. (Following Tryon, we shall symbolize this pattern thus: $++0$.) The simplest interpretation of this pattern is in terms of a general factor through all four variables, and in addition a second (positive) factor common either to x_1 and x_2 *or* to x_3 and x_4 (or it may be that there are *two* secondary factors, one in x_1 and x_2 and the other in x_3 and x_4). To determine which of these possible conditions is the most probable we may make use of other variables, x_5, x_6, $--x_n$. If we should combine x_1 and x_2, not merely with x_3 and x_4, but successively with all possible combinations of the remaining $n-2$ variables taken two at a time (so that we got combinations of four of the type x_1, x_2, x_i, x_j, where i and j take all values from 3 to n, but $i \neq j$); and if such combinations of four should invariably yield tetrad patterns of the type $++0$; then we should be justified in concluding that a special positive bond was present between x_1 and x_2.[42] (If, also, all combinations of the type x_3, x_4, x_i, x_j gave the same pattern, there would be an x_3x_4 bond as well.)

In practice it is seldom necessary to calculate all three tetrads from a combination of four variables, say x_1, x_2, x_3, x_4; for $t_{1243} - t_{1234}$ will always equal t_{1342}. So, in any particular case, if $t_{12ji} = t_{12ij}$, the corresponding t_{1ij2} will invariably equal 0. In general, then, we need concern ourselves with the first two tetrads only in any pattern of three. The mean tetrad is a device for getting the average value of these first two tetrads from all combinations of the type x_1, x_2, x_i, x_j. In general terms, the mean tetrad t_{ab} is the average of all tetrads (derived from n variables) of the types t_{abij} and t_{abji}.[43]

---

[42] It will be shown presently that it is not necessary for t_{12ij} always to equal t_{12ji} in order to establish the probable presense of an x_1x_2 bond; but that it is merely necessary to show that the values of t_{12ij} and t_{12ji} are in general > 0.

[43] The full expressions for the tetrads derived from combinations of the x_a, x_b, x_i, x_j type are: $t_{abij} = r_{ab}r_{ij} - r_{ai}r_{bj}$; $t_{abji} = r_{ab}r_{ji} - r_{aj}r_{bi}$; and $t_{aijb} = r_{ai}r_{jb} - r_{aj}r_{ib}$. Here r_{ab} occurs *only* in t_{abij} and t_{abji}, and in each case on the *left* of the minus sign. The mean tetrad t_{ab} may, therefore, also be defined

For the sake of simplifying the exposition above, the assumption was made that in a particular case all x_1, x_2, x_i, x_j combinations produced tetrad patterns of the $++0$ type. Such uniformity, however, is seldom experienced in practice; nor is it necessary that this particular pattern appear uniformly in order to demonstrate the presence of an x_1 x_2 bond. It can be shown that it is merely necessary for the mean tetrad t_{12} to be > 0. Tryon has shown that there are a number of other relatively simple patterns which indicate a possible x_1 x_2 bond; and it may be inferred from a study of Table IV, which lists these patterns, that they too would tend to increase the value of the mean tetrad.

It will be seen that whenever the letter "A" appears once in the type name, the chances are 1 in 2 that an x_1 x_2 bond is present in any particular x_1 x_2 x_i x_j combination. (The "A's" also indicate possible x_3 x_4 bonds; but we recall that x_3 and x_4 are only transient cases of x_i and x_j in the mean tetrad t_{12}, while x_1 and x_2 occur in every x_1 x_2 x_i x_j combination.) If "AA" appears, the chances are exceedingly strong for the presence of an x_1 x_2 bond.[44] On the other hand, whenever "B's" and "C's" appear, bonds other than x_1 x_2 are indicated (though an x_1 x_2 bond is not thereby ruled out, unless the "A's" are wholly absent).

Under the caption "Approximate Patterns," the *first two columns* are the only ones which will affect the value of the mean tetrad, as we have seen above (p. 29). It will be observed that for all patterns involving a possible *x_1 x_2 bond* the average value of the first two tetrads is *positive* (except for Types BBA, CCA, and ABC, where the "A's" have a minority influence only and the average is 0); and for all patterns *not involving an x_1 x_2 bond* the average

---

as the mean of all tetrads in which r_{ab} occurs on the *left* of the minus sign. There will be $(n-2)$ $(n-3)$ such tetrads.

Kelley (14, p. 87) gives a skeleton formula for calculating t_{ab} from the previously calculated individual tetrads. C. Brolyer (5, pp. 212*f*) has worked out a formula for calculating t_{ab} without the necessity of calculating a single individual tetrad:

$$t_{ab} = \frac{\sum_{i=1}^{i=n} r_{ai}r_{bi} + r_{ab}\sum_{i,j=1}^{i,j=n} r_{ij} + r_{ab}^2 - \sum_{i=1}^{i=n} r_{ai}\sum_{i=1}^{i=n} r_{bi} - r_{ab}\sum_{i=1}^{i=n} r_{ai} - r_{ab}\sum_{i=1}^{i=n} r_{bi}}{(n-2)\ (n-3)}$$

A rule of thumb method for applying this formula with the aid of a calculating machine is given in the above reference. t_{ab} for as many as 14 variables, where $(n-2)$ $(n-3)$, the number of tetrads to be averaged, is 132, may be calculated in from four to eight minutes. This is a labor-saving device of the first importance and has been used extensively in the present investigation.

[44] See note ‡ below Table IV.

TABLE IV
TRYON'S TETRAD PATTERNS*

Type	*Approximate Patterns* t_{1234}	t_{1243}	t_{1342}	*Probable Bonds*	*Remarks Suggesting Method of Proof, etc.* (In part reduced, with modifications, from Tryon, Kelley, *et al.*)
A (Basic Types)	+	+	0	*12 or* 34	*Type A:* same as Kelley's Prop. 16: $t_{1234} = \beta_1 \beta_2 \alpha_3 \alpha_4$† (*or* $\alpha_1 \alpha_2 \beta_3 \beta_4$) $= t_{1243}$; $t_{1342} = 0$. *Types B and C:* essentially corollaries of Prop. 16, the only difference in the proof being the location of the bonds assumed. *For B:* $t_{1234} = -t_{1342} = -\beta_1 \alpha_2 \beta_3 \alpha_4$ (*or* $-\alpha_1 \beta_2 \alpha_3 \beta_4$); $t_{1243} = 0$. *For C:* $t_{1234} = 0$; $t_{1243} = t_{1342} = -\beta_1 \alpha_2 \alpha_3 \beta_4$ (*or* $-\alpha_1 \beta_2 \beta_3 \alpha_4$). *Type AA* is implicit in Prop. 16, but is rejected by Kelley for more parsimonious interpretations: the term $\beta_1 \beta_2 \alpha_3 \alpha_4$ of Type A (first case) is supplemented, becoming: $\beta_1 \beta_2 \alpha_3 \alpha_4 + \alpha_1 \alpha_2 \gamma_3 \gamma_4 + \beta_1 \beta_2 \gamma_3 \gamma_4$. Similar transformations occur in *Types BB and CC.*‡ The *AB pattern* is obtained by combining, for each tetrad, the corresponding terms of the basic types, A and B, thus: t_{1234} (first case) becomes $\beta_1 \beta_2 \alpha_3 \alpha_4 - \beta_1 \alpha_2 \beta_3 \alpha_4$, or approximately 0. (If $\beta_2 = \alpha_2$ and $\beta_3 = \alpha_3$, exactly 0.); t_{1243} (first case) remains $\beta_1 \beta_2 \alpha_3 \alpha_4$; and t_{1342} (first case) remains $\beta_1 \alpha_2 \beta_3 \alpha_4$. The *AC and BC patterns,* and *all remaining patterns,* are obtained by similar combining processes. *Empirical Illustrations from Data of Present Study* The combination of variables 1 6 5 11 gives the tetrads .1699, .0966, -.0733. This pattern is clearly of the AAC type, suggesting a 1–6 bond (both are verbal), a 5–11 bond (both are numerical), and a possible 6–5 bond (both are analogies). See also Table III, p. 27, above.
B	–	0	+	13 *or* 24	
C	0	–	–	14 *or* 23	
AA	++	++	00	*12 and* 34	
BB	––	00	++	13 *and* 24	
CC	00	––	––	14 *and* 23	
*A*B	?0	+	+	*12*, 13 *or 12*, 24 *or* 34, 13 *or* 34, 24	
*A*C	+	?0	–	*12*, 14 *or 12*, 23 *or* 34, 14 *or* 34, 23	
BC	–	–	?0	13, 14 *or* 13, 23 *or* 24, 14 *or* 24, 23	
*AA*B	+	++	+	*12*, 34, 13 *or* *12*, 34, 24	
*AA*C	++	+	–	*12*, 34, 14 *or* *12*, 34, 23	
BB*A*	–	+	++	13, 24, *12 or* 13, 24, 34	
BBC	––	–	+	13, 24, 14 *or* 13, 24, 23	
CC*A*	+	–	––	14, 23, *12 or* 14, 23, 34	
CCB	–	––	–	14, 23, 13 *or* 14, 23, 24	
*A*BC	?0	?0	?0	*12*, 13, 14; *or 12*, 13, 23; 34, 13, 14; 34, 13, 23; *12*, 24, 14; *12*, 24, 23; 34, 24, 14; 34, 24, 23	

* Only those patterns arising from one, two, or three first order bonds, in all possible positions, are given here. The more complex patterns are difficult to interpret with certainty and are less likely to fit the variables of the present study.

† For a general discussion of these coefficients see Kelley (14, pp. 34*ff*), Tryon (29, pp. 325*ff*), Holzinger (13, pp. 4*ff*), etc. The α coefficients are due to a general factor through all four variables; the β's are due to a second factor common to two variables only; and the γ's (when present) are due to a third factor common to two other variables.

‡ Without knowing the values of these coefficients, there is no sure way of distinguishing (from single examples) Types AA, BB, and CC from the corresponding basic types, since the patterns are the same except for difference in size. The large size of the tetrads might be due either to one strong bond or to two weak ones. However, these types are important for building up the subsequent types, which actually appear in practice. See "Empirical Illustrations" above.

value of the first two tetrads is *negative*. The full significance of the mean tetrad at once becomes clear. If the value of t_{12} is reliably > 0, there is well-nigh overwhelming evidence for the presence of an x_1 x_2 bond; for it is the "A" types *only* which produce net positive values in the first two columns. The concept of the mean tetrad t_{12} is now no longer restricted to the original $++0$ pattern of Kelley's Proposition 16, but it has been extended to cover a large number of other patterns which also indicate a possible x_1x_2 bond, all the "A" types. If t_{12} is negative, the linkages of x_1 and x_2 *with other variables* are more potent (or numerous) than any linkage which may exist *between* x_1 and x_2.[45] If the value of t_{12} is approximately 0, either x_1 and x_2 have no *special* linkages (either with each other or with other variables), *or* the bonds tending to hold them together are neutralized by bonds attaching them to other variables.[46]

We are now prepared to evaluate the mean tetrads calculated for the variables of the present study.

5. Importance of Material and Structural Similarity

(a) *Evidence from Analysis of Mean Tetrads*

The values of the mean tetrads for all possible pairs of variables formed from the entire fourteen are given in Table V. Except in V–D, the mean tetrads were not derived from the entire fourteen *taken together,* but were derived from combinations of eight or ten only, combinations of numerical with spatial, spatial with verbal, and numerical with verbal variables. These smaller combinations were used partly to bring out more clearly any "cleavage" which might exist between the various material groups, and partly to lessen the danger of obliterating some of the weaker bonds. The mean tetrad values in V–D were obtained from all fourteen variables taken together, for the purpose of checking some of the results from the smaller combinations.

In the first three sections of this table the values which appear in the "triangles" are in every case derived from pairs of variables which are similar in material; *i.e.,* they are both numerical, both spatial, or both verbal. They may be referred to as "intra-group pairs." The values which appear in the rectangles are in every

---

45 For it is the "B's" and "C's" *only* which produce negative values in the first two columns.

46 Findley (8, pp. 73*ff*) makes it very clear that a zero mean tetrad value by no means indicates the complete absence of group factors.

TABLE V

MEAN TETRADS

A. Derived from Numerical and Spatial Variables Combined (N = 8)

	Variable	*Numerical*			*Spatial*			
		5	*8*	*11*	*3*	*7*	*13*	*14*
Numerical	No. Series 2	.0291	.0624	.0011	(−.0783)	−.0081	.0008	(−.0070)
	Num. Anal. 5		.0111	.0675	−.0560	.0137	−.0359	−.0309
	Arith. Reas. 8			.0257	−.0534	.0235	−.0279	−.0428
	Num. Gens. 11				−.0666	−.0089	.0360	−.0548
Spatial	Minn. F. B. 3					.0221	.0711	[.1611]
	Spat. Anal. 7						−.0305	−.0119
	Spat. Gens. 13							−.0137
	M. K. Spat. 14							

Mean P.E.$_{t}$ = .0250* (Spearman's formula) (n−2) (n−3) = 30†

B. Derived from Verbal, Spatial, and Anagrams Variables Combined (N = 10)

	Variable	*Spatial*			*Verbal*					*Anagrams*
		7	*13*	*14*	*1*	*4*	*6*	*10*	*12*	*9*
Spatial	Minn. F. B. 3	.0467	.0750	[.1560]	(−.0540)	−.0521	−.0938	−.0012	−.0898	(.0007)
	Spat. Anal. 7		0190	.0503	−.0229	.0003	−.0170	−.0369	−.0357	−.0200
	Spat. Gens. 13			.0177	−.0552	−.0408	−.0366	.0355	−.0179	.0033
	M. K. Spat. 14				(−.0436)	−.0427	−.0967	.0047	−.0678	(.0167)
Verbal	Sent. Comp. 1					.0611	.0589	−.0066	.0351	(.0300)
	Verb. Anal. 4						[.0515]	.0331	.0404	−.0490
	Gram. Anal. 6							−.0137	.1169	.0306
	Verb. Gens. 10								[.0042]	−.0431
	Gram. Gens. 12									.0146

Mean P.E.$_{t}$ = .0245* (n−2) (n−3) = 56†

* *Not* P.E.$_{\text{Mean } t}$. The calculation of the P.E. for even a single mean t is extremely involved (14, p. 87*ff*); however, these values are not required for the analysis made in this study. The values of mean P.E.$_{t}$ are given merely as a *rough* measure of the size of individual mean t's; for, in general, they will be considerably *larger* than P.E.$_{\text{Mean } t}$, and hence doubly safe.

† *Each* value in these tables is based on the average of (n−2) (n−3) individual t's, 30, 56, or 132, as the case may be.

TABLE V—(*Continued*)

MEAN TETRADS

C. Derived from Numerical, Verbal, and Anagrams Variables Combined (N = 10)

	Variable	*Numerical*			*Verbal*					*Anagrams*
		5	8	11	1	4	6	10	12	9
Numerical	No. Ser. 2	.0491	.0415	–.0108	(.0156)	–.0051	–.0675	–.0022	.0015	(–.0221)
	Num. Anal. 5		.0246	.0654	–.1067	–.0255	–.0124	–.0193	–.0190	.0437
	Arith. Reas. 8			–.0059	–.0016	.0092	–.0304	–.0179	–.0292	.0097
	Num. Gens. 11				–.0168	–.0489	–.0152	.0176	–.0064	.0209
Verbal	Sent. Comp. 1					.0509	.0337	.0028	–.0007	(.0227)
	Verb. Anal. 4						[.0253]	.0530	.0021	–.0610
	Gram. Anal. 6							–.0155	.0693	.0127
	Verb. Gens. 10								[–.0048]	–.0138
	Gram. Gens. 12									–.0128

Mean P.E.$_{\cdot t}$ = .0246* (n–2) (n–3) = 56†

D. Derived from All *14* Variables Combined

Pair	*Mean t*	*Pair*	*Mean t*	*Pair*	*Mean t*	*Pair*	*Mean t*	*Pair*	*Mean t*
1– 2	.0235	2– 3	–.0335	3– 7	.0607	4–10	.0442	7–13	.0049
1– 3	–.0398	2– 5	.0340	3– 8	–.0197	4–11	–.0346	7–14	.0371
1– 4	.0683	2– 6	–.0422	3– 9	.0063	4–12	.0361	8–11	–.0059
1– 5	–.0908	2– 8	.0302	3–11	–.0327	5– 6	–.0001	9–14	.0134
1– 6	.0634	2– 9	–.0263	3–12	–.0785	5– 7	.0410	10–11	.0062
1– 9	.0294	2–11	–.0149	3–13	.0815	5– 8	.0118	10–13	.0414
1–10	.0058	2–14	.0143	3–14	.1609	5–11	.0493	11–12	.0186
1–12	.0334	3– 4	–.0343	4– 5	–.0242	6– 7	–.0232	11–13	.0204
1–13	–.0471	3– 5	–.0092	4– 6	.0562	6–10	–.0031	12–13	–.0224
1–14	–.0330	3– 6	–.0760	4– 7	–.0083	6–12	.1082	13–14	.0180

(n–2) (n–3) = 132†

* See note * beneath Table V–B.
† See note † beneath Table V–B.

case derived from pairs of variables which are dissimilar with respect to material. They may be referred to as "cross-group pairs." The underlined values are from pairs of variables similar in structure; and the values in brackets are from variables similar in *both* material and structure. The values in parentheses are from pairs of variables in the relatively amorphous "construction" group.

Before subjecting these data to detailed analysis we may note certain features of interest with the naked eye. Judged by the size and the positive sign of the mean tetrad, the most potent bond appears between variables 3 and 14 (Tables V–A and V–B), the Minnesota Form Board Test and the M. K. Spatial Test, which are both spatial tests and have in addition a marked similarity in structure. The next most important bond is indicated between variables 6 and 12 (Tables V–B and V–C), which are both not only verbal but are also grammatical tests. It is interesting to recall that these same two bonds were strongly suggested by the earlier simple tetrad analysis (p. 28). This is a rough dual check on the validity of the two techniques. A more important check is the fact that these statistical bonds make psychological sense. Other bonds of considerable strength, all showing a similar psychological consistency, appear between variables 5 and 11 (both numerical), 3 and 13 (both spatial), 1 and 4 (both verbal), 2 and 8 (both numerical), etc. Less potent bonds appear between some pairs of variables which are similar in structure only: *e.g.,* 10 and 13 (both generalizations), and 11 and 13 (both generalizations), etc.

Looking more closely at the first three sections of Table V, another feature stands out. The mean tetrads which appear in the "triangles" are quite commonly of positive sign (though there are exceptions), in marked contrast to the values in the rectangles, which are in general negative. Since the values in the triangles are in every case derived from intra-group pairs (in these cases pairs from the same *material* groups), and those in the rectangles from the cross-group pairs (with respect to material groups), the hypothesis that similarity of material is influential in producing first order bonds is considerably strengthened. By the same token, "cleavages" are suggested between the material groups, taken as a whole: that is, between the numerical and the spatial groups, the spatial and the verbal, the numerical and the verbal. On still closer scrutiny, it may be surmised that the mean tetrad values for the variables which are similar in structure (those that are underlined) are larger, algebraically, than the values for the remaining cross-pairs, which represent neither similarity of material nor of structure.

These bird's-eye impressions when subjected to the usual statistical tests are found to be essentially sound. The importance of material and structural similarity for producing first order bonds may be clearly demonstrated by pooling the data from Tables V–A, V–B, and V–C and comparing the averages for the different categories.

Table VI–A summarizes these comparisons. The average (M) of the mean tetrads for the 19 pairs of variables which are *similar in material* but *not* in structure[47] is .0297. The average (N) of the mean tetrads for the 50 pairs of variables which are similar *neither* in material nor in structure[48] is –.0280. The difference between

TABLE VI

SUMMARY OF MEAN TETRADS

A. Absolute and Relative Importance of Structure and Material

Mean t's Averaged	*Average*
(M) Average for 19 pairs* of variables *similar in material* but *not* in structure	.0297
(S) Average for 19 pairs of variables *similar in structure* but *not* in material	–.0064
(S′) Average for the 10 *analogies and generalizations pairs only, similar in structure* but *not* in material†	.0024
(N) Average for 50 pairs‡ of variables similar *neither* in material nor in structure	–.0280

Averages Compared	*Difference*	*Diff./P.E.*$_{Diff.}$
Average M-Average N	.0577	10.55
Average S-Average N	.0216	3.915
Average S′-Average N	.0304	5.56
Average M-Average S	.0361	5.641
Average M-Average S′	.0273	4.29

* *Each* item based on an average of two mean t's for the same pair of variables from two different tables (*e.g.*, the pair 2, 8 appears both in Table V–A and V–C).

† The pairs from the relatively amorphous "construction" group have been omitted from the structure group average.

‡ Four of these items (9, 4; 9, 6; 9, 10; and 9, 12) are based on an average of two mean t's from two different tables.

[47] The values which appear in the "triangles" in Tables V–A, B, and C with the exception of those in brackets. Since each of the 19 pairs is represented twice, the average value of the mean tetrads is used in this final summary.

[48] The values which appear in the rectangles, with the exception of those underlined, in Tables V–A, B, and C.

TABLE VI—*Continued*

B. "Cleavage" between Material Groups

Groups Contrasted	*Average*	*Difference: Intra-group minus Cross-group*	*Diff./P.E.*$_{Diff.}$
Numerical vs. Spatial: average of mean t's for 11 *intra-group* pairs (pairs of similar structure having been eliminated)	.0213	.0514	6.13
Average for 12 *cross-group* pairs*	–.0301		
Verbal vs. Spatial: average of mean t's for 13 *intra-group* pairs (pairs of similar structure having been eliminated)	.0411	.0881	10.7
Average for 14 *cross-group* pairs*	–.0470		
Numerical vs. Verbal: average of mean t's for 14 *intra-group* pairs (pairs of similar structure having been eliminated)	.0257	.0503	6.95
Average for 15 *cross-group* pairs*	–.0246		

C. "Cleavage" between Structure Groups

Groups Contrasted	*Average*	*Difference: Intra-group minus Cross-group*	*Diff./P.E.*$_{Diff.}$
Analogies vs. Generalizations: average of mean t's for 10 *intra-group* pairs (pairs of similar material having been eliminated)	.0024	.0321	6.15
Average for 10 *cross-group* pairs†	–.0297		
Analogies vs. "Construction": average of mean t's for 14 *intra-group* pairs (pairs of similar material having been eliminated)	–.0133	.0262	2.95
Average for 15 *cross-group* pairs†	–.0395		
Generalizations vs. "Construction": average of mean t's for 14 *intra-group* pairs (pairs of similar material having been eliminated)	–.0058	.0176	2.11
Average for 15 *cross-group* pairs†	–.0234		

* Pairs containing one variable from each of the two material groups, pairs of similar structure not included.

† Pairs containing one variable from each of the two structure groups, pairs of similar material not included.

these averages is highly reliable, Diff./P.E.$_{Diff.}$ being 10.5. The average (S) for the 19 pairs of variables which are *similar in structure* but *not* in material[49] is – .0064. There is good reason to interpret this neutral value, *not* as indicating the general absence of structural bonds, but as indicating the presence of relatively weak, or infrequent, bonds which are in general not permitted to assert themselves positively because of the restraining influence of the stronger, or more numerous, material bonds attached to the same variables, but pulling in other directions. In addition to the evidence of individual structural bonds already cited (p. 35), the evidence obtained from a comparison of mean tetrad averages supports this interpretation: the chances are very good that average S is greater (algebraically) than average N; for Diff./P.E.$_{Diff.}$ = 3.9.[50] In these comparisons we have included in our structure group the somewhat dubious "construction" tests, dubious because of their lack of precisely similar form. If we eliminate these relatively amorphous "construction" variables from the structure group, and average the mean tetrads for the 10 pairs of analogies and generalizations tests only (average S′), the difference between averages is still more marked: average S′ is reliably greater than average N, Diff./P.E.$_{Diff.}$ becoming 5.6. Thus, not only is structural similarity important for the production of bonds, but the greater the similarity the stronger the bonds.

The *relative* importance of material and structural similarity (as represented by the variables of this study) for the production of first order bonds is brought out by comparing average M with average S. The average of the mean tetrads for the material pairs is reliably greater than the average for the structural pairs: Diff./P.E.$_{Diff.}$ = 5.6. If the 10 pairs from the analogies and generalizations groups only are used, the difference is less marked but still reliable: Diff./P.E.$_{Diff.}$ = 4.3.

As a check on the evidence from Table V–A, V–B, and V–C, the 50 mean tetrad values in V–D have been introduced. The mean terads in sections A, B, and C were calculated from combinations of 8, 10, and 10 variables, respectively, for the reasons given above (p. 32); the values in section D are derived from the entire 14 variables. *Each* value in A, B, and C sums up the evidence from 30, 56, and 56 individual tetrads, respectively; whereas, each value

---

49 The values which are underlined in the rectangles in Tables V–A, B, and C.

50 See also note 46, p. 32.

in D is based on 132 tetrads. In other words, the values in D have survived "open market competition," and should be even more dependable than those in A, B, and C. It has been thought unnecessary to calculate the remaining 41 of the 91 mean tetrads which are possible from 14 variables; since the sample of 50 gives results which are so consistent with the results obtained by pooling the data from the smaller combinations. For example, the three most important bonds appear between variables 3 and 14, 6 and 12, and 3 and 13 in Table V–D, just as they did in sections A, B, and C. Though in individual cases there are minor variations in the absolute values from the different tables,[51] quite generally the relative sizes are closely similar. The most rigorous check is afforded by calculating, for the data of Table V–D, average M and average S, using as before the 19 pairs of variables which are similar in material and the 19 which are similar in structure, and comparing the results from the two different sources. This comparison is made below:

Source of Data	*Average M*	*Average S*	*Av. M–Av. S*	*Diff./P.E.$_{Diff.}$*
Table V–A, B, C:—	.0297	–.0064	.0361	5.64
Table V–D:—	.0349	.0002	.0347	5.57

Though there is a slight shift in the values of the averages, the M–S differences and the values of Diff./P.E.$_{Diff.}$ are almost identical.

If it is true, then, that simliarity either of material or of structure tends to "tie" variables together, it follows (other things being equal) that variables which are dissimilar in these respects will tend to be "pulled away" from each other. That is, we might reasonably expect a "cleavage" to appear between the various material and structure groups. Tables VI–B and VI–C show that this is the case, by comparing the averages of the mean tetrads from

[51] These differences are due to the fact that the various ij pairs associated with any particular ab pair in determining the mean tetrad t_{ab} were not all the same for the different tables. In this connection, Findley's excellent discussion of the limitations of the mean tetrad should be read (8, Appendix D). On the basis of a detailed analysis of the tetrads, both individual and mean, from seven variables, he shows that a *single* mean tetrad may be misleading. However, the major conclusions in the present study are based, not on single mean tetrads, but on the averages of from 10 to 50 mean tetrads. Furthermore, as the number of variables employed in calculating the mean tetrads increases, the chances of a distorted value are in general lessened. In the present study n was 8, 10, or 14, as compared with 7, which Findley used to illustrate his contentions. Thus, though Findley's warnings are well-founded, the validity of the results above is not impaired.

the intra-group pairs of variables with the averages from the cross-group pairs. VI–B summarizes the material group comparisons. The average of the 11 intra-group pairs[52] from the numerical and spatial groups is .0213, as compared with the average of the 12 corresponding cross-group pairs, which is −.0301. The differences between these two averages is highly reliable, Diff./$P.E._{Diff.}$ being 6.1. This means that the numerical tests have much more in common with each other than they have in common with the spatial tests, and *vice versa*. The cleavage between the numerical and verbal groups, as indicated by a similar comparison of the average mean tetrads for the intra-group and cross-group pairs, is just as evident: Diff./$P.E._{Diff.}$ equals 6.9. The most marked cleavage is between the verbal and the spatial groups, Diff./$P.E._{Diff.}$ being 10.7. The order of magnitude of these three cleavages seems to indicate that the numerical group stands somewhere between the other two groups; that is, it is less far removed from the verbal and spatial groups than they are from each other. It should be stressed that the existence of a "cleavage" between two groups does not imply that the groups have nothing in common. That there is an important factor common to all fourteen variables will be shown in Section 7.

For the structure groups the cleavages, in general, are less pronounced, as we should expect from the relative strength of the material and structure bonds. However, the evidence (summarized in Table VI–C) indicates a reliable cleavage between the analogies group and the generalizations group. Here the difference between the averages for the intra-group mean tetrads and the cross-group values[53] is 6.2 times the P.E. of the difference. Thus this cleavage is of the same order as the cleavages between the numerical and spatial groups, and between the numerical and verbal groups. But between the "construction" and the analogies groups, and between the "construction" and generalizations groups, cleavages are not reliably established, though they are strongly suggested; for the "construction" and analogies groups Diff./$P.E._{Diff.} = 3.0$; for the "construction" and generalizations groups Diff./$P.E._{Diff.} = 2.1$. That the most marked cleavage should appear between the gen-

[52] The value of $t_{3,14}$ has been omitted; for 3 and 14 are similar both in structure and in material. For similar reasons, $t_{4,6}$ and $t_{10,12}$ have been omitted in subsequent comparisons.

[53] Pairs of similar material are not included in these averages. Here cross-group pairs are pairs containing one variable from each of the two *structure groups*.

eralizations and analogies groups is not surprising in view of the more clear-cut structural similarity of the variables in these groups, as contrasted with the relatively amorphous variables in the "construction" group.

Taking all of the evidence of this section into consideration, it seems clear that either material or structural similarity may be important in the production of *first order* group factors; and that (so far as represented by the variables of this study) material similarity is more important than structural similarity. In Section 3 of this chapter the simple tetrad technique failed to reveal convincing evidence of *third or fourth order* group factors running through groups of four or five variables of similar material or structure. It remains now to investigate the possibility of *second order* group factors due to either type of similarity. In the following section this is attempted.

6. Importance of Material and Structural Similarity

(b) *Evidence from the Application of Hotelling's Method of Principal Components to Groups of Three*

Professor H. Hotelling has recently developed a technique for analyzing the factors present in n variables,[54] which is much easier to apply than that of Kelley (14, pp. 80*ff*). By this device one is able to account for the intercorrelations of n variables in terms of n "principal components"[55] (or factors); to get the correlation[56] of each variable with each of the principal components; and, hence, to determine the percentage of the variance in any particular variable due to each principal component,[57] or to determine the percentage of the sum of the variances of the n variables due to each principal component.[58]

---

[54] This technique determines factor weights from a matrix of intercorrelations (or covariances). A discussion of the method is to be published shortly by Professor Hotelling.

[55] The method assumes n principal components for n variables. It gives the heaviest weight to the "first principal component" and decreasing weights to those which follow.

[56] Prof. Hotelling prefers to express this in terms of "covariance," for the units are chosen so that all σ's are equal to unity: *i.e.*, reduced scores are assumed. This is a legitimate assumption, for all calculations are based on the intercorrelations of the variables, and r's from raw and from reduced scores are the same.

[57] By simply squaring the correlation (or "covariance") of the variable with the principal component.

[58] Since the σ's are taken as equal to 1, the variances (σ^2's) will also equal 1. The sum of the variances, or the "variance sum," for n variables will equal n; and the percentage of the variance sum due to any particular

Some aspects of this technique have been applied to a special set of nine variables in the hope of detecting the presence of possible second order group factors running through groups of three variables similar either in material or in structure. The nine tests selected for this analysis have certain important relationships which will be apparent in the scheme below:

Variables			*Nature of Variables*		
5	11	2	N_a	N_g	N_c
7	13	14	S_a	S_g	S_c
4	10	1	V_a	V_g	V_c

The capital letters indicate the material groups to which the variables belong (*N*umerical, *S*patial, *V*erbal); the subscripts indicate the structure groups to which they belong (*a*nalogies, *g*eneralizations, ''*c*onstruction''). From this set of nine it is possible to select three material groups, each composed of three variables which are similar with respect to material (*i.e.,* 3 N's, 3 S's, 3 V's); and three structure groups, each containing three variables of similar structure (*i.e.,* 3 a's, 3 g's, 3 c's); and each variable appears once in a material group and once in a structure group.

Now, if we apply Hotelling's method of principal components to one of these six groups, say the numerical group, and get a measure of ''what is common''[59] to the three variables, we find that the first principal component accounts for 59.1% of the variance in these three variables. If we apply the technique to the entire nine variables taken together, to get a measure of what is common to all nine, we find that the first principal component *of the nine* accounts for only 47.2% of the variance *in the same three variables.*[60] Thus, the three numerical variables have apparently something special in common with each other, in addition to whatever is common to all nine variables.[61] There is a temptation to explain this by assuming that to the general factor (common to all nine) has been added

principal component will equal the sum of the squares of the correlations (or covariances) of the n variables with the principal component divided by n (*i.e.,* the ''mean square correlation'' of the variables with the principal component).

59 There are other possible ways of interpreting the first principal component, but this is the most parsimonious.

60 The 47.2% is *not* the variance *in the entire nine* due to the first principal component of the nine, but the variance *in the three numerical variables* due to the first principal component *of the nine.*

61 Part of this drop from 59.1% to 47.2% is due to the technique (which requires n principal components for n variables); but what follows makes it plain that some intrinsic factors are also involved.

merely a numerical factor. That this is too simple an explanation will be seen by applying the method to some *nondescript group of three;* that is, to a group in which no two variables are similar in either material or structure. For example, the combination N_g S_a V_c has a first principal component which accounts for 53.5% of the variance; whereas the first principal component *of the nine* accounts for only 45.3% of the variance *in the same three variables.* Here again there is apparently something common to the three in addition to the more general factor;[61a] but it cannot be attributed to any of the six kinds of similarity which characterize our material and structure groups. On reflection, this result is found to be entirely natural; for, in general, as we increase the number of items (not specifically selected for the identity of their elements) in any aggregate, we tend to decrease the number of elements common to the whole.[62]

If this is the case, in order to demonstrate the importance of material or structural similarity for the production of second order bonds, we must show that the first principal components in our six groups of three are in general *more* influential than the first principal components of nondescript groups *of the same size* and *from the same group of nine variables.* From our special nine it is possible to make up six nondescript groups of three, such that no two variables within any group will be similar either in material or structure (*e.g.*, N_g S_a V_c). These six groups together use each of the nine variables twice. The material groups together use each variable once; the structure groups together use each variable once; the material and structure groups together use each variable twice. Thus, the six nondescript groups taken together make an almost ideal "control group." If the material or structure groups give evidence of having larger first principal components, in general, than these nondescript control groups, we may fairly attribute this to their special material or structural similarity. The average of the squares of the correlations (or the % variances)[63] of the variables with their respective first principal components may be taken as a measure of the size of the components.

61a Part of this drop from 53.5% to 45.3% is due to the technique.

62 A homely example will make this clear. An orange and a lemon have many elements in common: they are both organic products, both fruits, both citrus fruits, etc. If we add a watermelon to this group, the element "citrus fruit" is no longer common to all. If we now add a mule to the group, the element "fruit" ceases to be common. And so on.

63 The square of the correlation of a variable with a principal component is the per cent variance in the variable due to the principal component.

TABLE VII

Correlations* with the Various First Principal Components of the Special Groups of Three

Material Groups

Variable	*Nature of Variable*	*r with Respective Prin. Comp.*
	Numerical	
5	Na	.785
11	Ng	.753
2	Nc	.767
	Spatial	
7	Sa	.772
13	Sg	.715
14	Sc	.876
	Verbal	
4	Va	.845
10	Vg	.705
1	Vc	.724

Structure Groups

Variable	*Nature of Variable*	*r with Respective Prin. Comp.*
	Analogies	
5	Na	.754
7	Sa	.772
4	Va	.744
	Generalizations	
11	Ng	.770
13	Sg	.756
10	Vg	.732
	"Construction"	
2	Nc	.817
14	Sc	.801
1	Vc	.639

Nondescript Groups

Variable	*Nature of Variable*	*r with Respective Prin. Comp.*
5	Na	.681
14	Sc	.851
10	Vg	.656
5	Na	.726
13	Sg	.701
1	Vc	.518
2	Nc	.829
7	Sa	.740
10	Vg	.644
2	Nc	.822
13	Sg	.689
4	Va	.762
11	Ng	.715
14	Sc	.802
4	Va	.737
11	Ng	.789
7	Sa	.733
1	Vc	.668

* Professor Hotelling prefers to express this in terms of "covariance"; for the units employed in his technique are so chosen that the σ's equal unity.

The method used in deriving the values in this table deviates slightly from the original method of Hotelling: for the reasons given in note 69, p. 47 below, raw r's have been used in the body of the matrices and reliability coefficients have been used in the diagonals.

In Table VII are given the r's of all the variables with their several first principal components. In every case the first principal component is for a group of three variables only. Each variable appears as a member of one material group, one structure group, and two nondescript groups. Casual inspection suggests that the r's for the material and structure groups are in general higher than those for the nondescript groups, though there is a considerable amount of overlap. The averages of the % variance of the variables in the various categories would be more illuminating.

In Table VIII a comparison of these averages is made. It will be seen that the average (M) of the 9 % variances due to their respective first principal components for the three material groups

TABLE VIII

AVERAGE % VARIANCE* DUE TO THE VARIOUS FIRST PRINCIPAL COMPONENTS OF THE SPECIAL GROUPS OF THREE

% Variances Averaged	*Average*	*P.E.$_{Av.}$*
(M) 9 % variances due to the first principal component of each of three groups of three variables *similar in material* but *not* in structure	.598	.019
(S) 9 % variances due to the first principal component of each of three groups of three variables *similar in structure* but *not* in material	.571	.016
(N) 18 % variances due to the first principal component of each of six groups of three variables similar *neither* in material nor in structure (the six *nondescript* groups†)	.533	.017
(M + S) 18 % variances due to the first principal component of each of six groups of three variables similar *either* in material *or* in structure	.584	.013

Averages Compared	*Diff.*	*Diff./P.E.$_D$.*	*Chances in 100 of true difference:* *Using Diff./P.E.$_D$.*	*Using t-function‡*
Average M–Average N	.065	2.50	95	92–93
Average S–Average N	.038	1.62	86	81–82
Average M–Average S	.027	1.09	77	75
Average (M + S)–Av. N	.051	2.37	95	92–93

* See note 63, p. 43.
† *E.g.*, the combination N_a S_c V_g.
‡ See note 64 below.

is the highest; the average (S) for the structure groups comes next (if we except the joint average (M + S) for material and structure groups combined); and the average (N) of the 18 % variances for the six nondescript groups is lowest. The differences between these averages, when judged by the most exacting criteria,[64] are

64 Fisher's t-function criterion (9, pp. 111–119, 151) has been used in conjunction with the familiar Diff./P.E.$_{Diff.}$ test. This sets up a slightly more rigorous standard by making allowance for the probable divergence of the distribution of a small number of measures from the normal curve.

not entirely reliable, but they are suggestive. The chances are between 92 and 95 in 100 (depending on which criterion is used) that material similarity in general introduces into the groups of three a common element in addition to the general factor common to all nine variables. The chances are less good that structural similarity introduces such an additional common element into the groups of three (only 81–86 in 100). The chances that material similarity has a greater potency than structural similarity in this respect are between 75 and 77 in 100. When the 9 % variances from the material groups and the 9 % variances from the structure groups are pooled and compared with the 18 % variances from the nondescript groups, the chances that similarity (either material or structural) introduces an additional common element into the groups of three are between 92 and 95.[65]

A significant feature of these results is that they give to material and structural similarity the same *relative* importance in the production of *second order* group factors that the mean tetrad results gave them in the production of *first order* group factors. There is a strong suggestion that either type of similarity may tend to produce second order group factors (in addition to the more general factor common to the larger battery), but material similarity is the more influential.

7. Tentative Analysis of the "Common Factor"[66] Present in the Variables of This Study

Although the main purpose of this research, an investigation of the influence of certain types of material and structural similarity, has now been carried out, the data provide other interesting possibilities. The fourteen variables on which the study is based, all but one of them difficult power tests, represent an unusually broad range of mental abilities. They represent, as well, five or six hours of intensive effort on the part of each of the 186 subjects. Taken as a battery, they should prove a searching test of general intelligence. The special group of nine tests used in the section above is perhaps an even better measure of general mental ability; for,

---

[65] It is a little surprising that the chances should be almost as good here as in the case of the material groups alone. This is due to the fact that the $P.E._{Av.}$ for the 9 material % variances was larger than the $P.E._{Av.}$ for the 18 material and structure % variances together.

[66] The most parsimonious interpretation of the first principal component of this group of variables is in terms of a common factor; but this is not identical with Spearman's "general factor."

with the exception of the Arithmetic Reasoning Test, the tests eliminated from the fourteen were either less reliable or less efficient measures of the factor common to all fourteen. (See Table I above and Table IX below.) On the other hand, one poor test from both of these points of view, the Sentence Completion Test, was included in the nine per force, because of the special requirements of that group.[67] It will be interesting to determine (a) how large the common factors in these two groups are, and (b) which individual tests and what groups of tests are the best measures of these common factors in the two large batteries. Hotelling's technique, applied independently to the entire fourteen variables and to the special nine, furnishes a ready means of doing this.

In Table IX is given the estimate of the percentage of the "variance sum"[68] due to each principal component and the correlations (or covariances) of each variable with the various principal components. "K_1," "K_2," etc., indicate "first principal component," "second principal component," etc. The values for the special nine tests have been determined from both raw and corrected r's. It is felt that the values derived from the raw r's are the more reliable,[69] and, accordingly, the determination of the influence of the second and the third principal components[70] has been based on raw r's only. The importance of the common factors is indicated by the percentage of the "variance sum" due to the various first principal components. For the entire fourteen tests, K_1 accounts for 42.9%; for the special nine, K_1 accounts for somewhat more, 46.8% (or 56.5% if the results from the corrected r's

---

[67] The special characteristic of this group, it will be recalled, is that it contains three representatives of each of the three material groups and three representatives of each of the three structure groups, *each* variable appearing both in a material group and in a structure group.

[68] See note 58, p. 41.

[69] In Section 2 some reasons for preferring raw r's were given. Another reason appears in this table. When the calculations are based on the raw r's the M. K. Spatial Test (which, it will be recalled, had the highest reliability, .96, and the largest S.D.) is tied with the Spatial Analogies Test and the Number Series Test for the place of first importance. When corrected r's are used the Spatial Analogies Test (which has a reliability of .75, next but one to the lowest) forges way ahead of all the other tests, and the M. K. Spatial Test drops into sixth rank. In other words, correction for attenuation puts a premium on tests of low reliability and handicaps tests of high reliability. It seems, therefore, that the findings based on raw r's should be given more serious consideration.

[70] The second principal component is obtained from the residual covariances after the influence of the first principal component has been "partialled out" from the original covariances (or intercorrelations). The third principal component is obtained in a similar manner from the second set of residual covariances.

TABLE IX

Analysis of All Fourteen Variables and of Special Nine by Hotelling's Method of Principal Components

	All 14 (Based on *raw* r's)*	*Special 9* (Based on *raw* r's)*			*Special 9* (Based on *corrected* r's)
Root:—	K_1	K_1	K_2	K_3‡	K_1
% of the "Variance Sum": †—	42.9	46.8	9.5	6.2±	56.5
Variable	An1§	An1§	An2§	An3§	An1§
Arithmetic Reasoning	.757	....	....	....	....
Spatial Analogies	.738	.749	−.149	−	.863
Verbal Analogies	.727	.720	+.354	−	.784
Number Series	.721	.748	−.007	−	.788
Modified Kelley Spatial	.713	.741	−.149	+	.746
Numerical Generaliz's	.696	.701	−.116	−	.774
Spatial Generaliz's	.665	.677	−.088	+	.793
Grammatical Generaliz's	.634	....	....	....	....
Verbal Generaliz's	.609	.634	+.218	+	.724
Grammatical Analogies	.603	....	....	....	....
Numerical Analogies	.590	.604	−.557	−	.634
Sentence Completion	.578	.555	+.557	−	.629
Anagrams	.569	....	....	....	....
Minnesota Form Board	.512	....	....	....	....

* The use of raw r's is a slight deviation from the original method of Hotelling. (Raw r's have been used in the body of the matrices, reliability coefficients in the diagonals.) Reasons for this are given on p. 47, note 69.

† See note 58, p. 41.

‡ The calculations on the third principal component were extremely protracted and have not been carried to the end. They were carried far enough, however, to determine the signs of the covariances and the approximate amount of the variance sum to be expected. Since no psychologically intelligible tendency was found in the signs and since the variance sum was reduced to about 6%, further analysis in this direction would have been fruitless.

§ Covariances of each variable with respective principal components. These may be thought of as correlations.

are considered). The % of the variance in any particular variable due to any particular principal component may be found by simply squaring the correlation (or covariance) of that variable with the component. The percentage of the variance in any number of variables due to any particular principal component may be found by summing the squares of the covariances and dividing by the number of variables involved (for the variance of each test may be regarded as one).

If we carry out this operation in connection with the covariances of the special nine *with respect to the first principal component of* *the entire fourteen,* we find that the first principal component *for the fourteen* accounts for 45.3%[71] of the total variance *in the nine.* This value is astonishingly close to the value of the ''variance sum'' obtained for the first principal component of the nine, namely, 46.8%. That this close correspondence would be found might have been surmised from the correspondence of the individual covariances in the first two columns.[72] The implication of this seems to be that *to a large extent* the common factor in the nine is similar to the common factor in the fourteen.[73] Since common factors of considerable importance seem to be present in these two large batteries of tests, it becomes a matter of interest to observe which tests, or what groups of tests, best measure these factors.

The tests are arranged in order of their efficiency in measuring the factor common to all fourteen. The fact of major interest is that the magnitudes of these correlations (or covariances) with the first principal components are so uniform. However, there are differences; the range for the fourteen is .51 to .76; for the nine (based on raw r's), .56 to .75. The best single test is the Arithmetic Reasoning which correlates .757 with the first principal component. (This does not appear in the special nine.) The next five tests are but slightly less efficient: their correlations with the first principal component of the fourteen are all within .06 of the value for the Arithmetic Reasoning Test. These tests are Spatial Analogies, Verbal Analogies, Number Series Completion, Modified Kelley Spatial, and Numerical Generalizations. In the group of nine, these same tests are at the top, though in slightly different order; they fall within a range of .05. In this superior group the numerical group is represented either two or three times,[74] the spatial group twice, and the verbal group only once.

A more accurate measure of comparative group influences may be obtained from the *average %* variance due to the two first principal components (that of all 14 and that of the special 9) for each of the three material groups and each of the three structure groups. These averages are given in Table X. Three different criteria are used: (a) that based on the *special nine,* which has the great advantage of dealing with the same number of tests from each of the

[71] This value does not appear in Table IX.

[72] The maximum deviation occurs in the case of the M. K. Spatial Test: .713 *vs.* .741, a difference of only .028.

[73] This similarity, though suggestive, cannot legitimately be interpreted as evidence for Spearman's ''universal g factor.''

[74] Three, if the Arithmetic Reasoning Test is included.

TABLE X

AVERAGE PER CENT VARIANCES DUE TO THE FIRST PRINCIPAL COMPONENT OF ALL FOURTEEN VARIABLES AND THE FIRST PRINCIPAL COMPONENT OF THE SPECIAL NINE FOR EACH OF THE MATERIAL AND STRUCTURE GROUPS

(NOTE: the averages are in order of magnitude.)

*% Variances Due to First Prin. Comp. of Special 9**		*% Variances Due to First Principal Component of All Fourteen Variables†*			
Group	*Av. of 3 % Variances*	*Group*	*Av. of 4 or 5 % Variances‡*	*Group*	*Av. of 3 Highest % Variances*
Spat.	.523	Num.	.481	Num.	.526
Anal.	.481	Anal.	.446	Spat.	.498
Num.	.472	Spat.	.439	Anal.	.479
Cons.	.472	Gens.	.425	Cons.	.454
Gens.	.450	*Verb.*	.400	Gens.	.443
Verb.	.409	Cons.	.390	*Verb.*	.434

* The % variances here averaged are derived from the r's which appear in column 2 of Table IX.

† The % variances here averaged are derived from the r's which appear in column 1 of Table IX.

‡ There were 5 variables in the verbal and the construction groups, 4 in all the others.

six groups to be compared, each of the nine variables being averaged in with just one material group and just one structure group, but the disadvantage of leaving out the Arithmetic Reasoning Test; (b) that based on *all fourteen* tests, which has the advantage of including the Arithmetic Reasoning Test and of covering a slightly broader range of abilities,[75] but the disadvantage of dealing with unequal numbers of tests in the six groups; and (c) that based on the *best three* from each of the groups in the entire fourteen, as judged by the % variance due to the first principal component; this has the advantage of equalizing the number of representatives from each of the six groups, but does not insure the averaging of each test once, and only once, with a material and with a structure group (as is the case with the special nine). It will be seen that the average % variance due to the principal components for the six groups do not differ from each other greatly, no matter which criterion is used, the greatest range being .409 to .523. However, what differences there are indicate that the spatial group is

[75] But see discussion above which indicates that the similarity between the common factor for the nine and the common factor for the fourteen is very great.

the most representative of the special nine, and that the numerical group is the most representative of the fourteen, by either criterion. This shift of supremacy is largely due to the fact that the Arithmetic Reasoning Test has been included in the numerical group, under the last two criteria. The analogies group ranks second if the first two criteria are used, and a close third if the third criterion is used. But this apparent superiority is probably fortuitous.[76] Most interesting is the consistently inferior position of the verbal group, either at the bottom or tied for bottom position.[77]

This relative disharmony between the verbal and the other tests is brought out by another aspect of Table IX. Under the caption "K_2" appear the correlations (covariances) of each of the special nine tests with the *second* principal component; that is, with the principal component of what remains after the influence of the first principal component has been eliminated. The striking fact here is that all correlations are negative *except those for the three verbal tests.* Thus, the verbal group, which was least representative of the first principal component, is most representative of the second principal component. However, this second component is relatively unimportant, since it accounts for but 9.5% of the variance sum, as compared with 46.8% due to the first.

The comparative weakness of the verbal group as a measure of the general ability tested by the variables of this study is interesting in connection with the weight commonly given to verbal items in tests of general intelligence.[78] It may be argued, however, that

[76] For, three of the six "nondescript groups" of Section 6 (composed, like the analogies group, of one member from each of the three material groups) have average % variances due to the first principal component of the special nine which are higher than the average for the analogies group. (These figures have not been reproduced here, but they may be checked readily from the figures in column 2 of Table IX.) In other words, the apparent superiority of the analogies group to the generalizations and "construction" groups may be due merely to a happy combination of tests from the three material groups; for the superior potency of the material aspects of the tests was demonstrated in Sections 5 and 6.

[77] Though the inferiority of the verbal group has not been reliably established, it is probably not wholly fortuitous; for only one of the six "nondescript groups" of Section 6 has an average % variance due to the first principal component of the special nine as low as that of the verbal group. (These figures have not been reproduced here, but they may be checked readily from the figures in Table IX, column 2.) This fact tends to reënforce the probability of genuine verbal inferiority which is suggested by Table X.

[78] See National Intelligence Tests, Otis Intelligence Scale, the I. E. R. Intelligence Scale C A V D, Thorndike's Intelligence Examination for High School Graduates, etc., etc. Terman has remarked of the revised Binet Test: "in the large majority of cases the vocabulary test alone will give an I. Q. within 10% of that secured by the entire scale." (26, p. 230.) Elsewhere he has expressed the need for non-verbal measures of intelligence: "A more

the numerical and the spatial tests together evoked a species of "mathematical ability" and by the force of their combined numbers "crowded the verbal tests out." But when we recall that the mean tetrad analysis of Section 5 gave evidence of a "cleavage" between the numerical and spatial groups quite as marked as that between the numerical and verbal groups,[79] this argument seems weak.[80] Perhaps a better explanation is the varied linguistic home background of the S's.[81] But, *if,* and in so far as, we assume that the battery of tests with which this study deals is a measure of general intelligence more adequate than those intelligence tests which stress verbal items and minimize numerical and spatial items,[82] doubt is thrown upon the correctness of the proportions of verbal, numerical, and spatial items which obtain in those tests.

---

serious effort should be made to devise non-verbal tests capable of bringing out differences in general intelligence on the higher levels." (25, pp. 127*ff.*) Thorndike does not consider his tests tests of *general* intelligence; for in 1921 he said: "Some of us have, I fear, claimed a generality for our measures . . . which it would be very hard to justify." . . . ". . . the ability measured by verbal tests is not the same as the ability measured by non-verbal tests; . . ." (27, pp. 125, 126). And in 1927 he prefers to define intelligence in terms of specific instruments for measuring it: thus, he speaks of "intellect C A V D," "C A V D O," etc. (28, pp. 65, 412*f*). Of especial interest is Peatman's analysis of the Thorndike Intelligence Examination for High School Graduates (17, pp. 18, 47, etc.).

[79] See Table VI–B above.

[80] See also the work of Rogers which involved both numerical and spatial tests (18, p. 96). She found no evidence for a general "mathematical ability." Nor did Spearman in reworking her data with new techniques (20, pp. 230–232).

[81] It will be recalled that in 84% of the S's homes some language other than English was spoken at least part of the time. However, the average age of the S's was 19½ years and 75% of them had had at least 9 years in the schools of New York City before entering college, which should to a considerable extent lessen the influence of the varied linguistic home background.

[82] In the absence of validating criteria I should hesitate to make this assumption, but, in view of the breadth and the intensiveness of the testing program, the assumption is not entirely devoid of plausibility.

CHAPTER IV

INTERPRETATION AND SUMMARY

1. Interpretation

The practical implications of this study are clear. To those concerned with the construction of internally consistent test batteries, objective similarity in tests, either material or structural, is suspect; but material similarity is more likely than structural similarity to result in groups factors. However, objective similarity of any kind is no guarantee that a corresponding group factor will appear; for the competing influences of other types of similarity, objective or implicit, may nullify the suggested bond.

From the theoretical point of view, it is interesting to examine the rough factor pattern which is tentatively suggested by the results of this investigation taken as a whole. In a few words, the pattern might be characterized thus: *a general factor, multiple overlapping group factors of various orders and various degrees of potency, and specifics.* Both the uniformity of the intercorrelations and the results from the application of Hotelling's method of principal components point to a general factor of considerable importance running through the entire group of fourteen variables. The mean tetrad results indicate that, superimposed on this common foundation, are diverse group factors which are determined largely by material or structural similarity, but which, nevertheless, give no promise of immutability. (In this connection, recall the noticeable variations in certain mean tetrad values which occur when the ij combinations are varied: see Table V.) In other words, group factors are *to some extent* a function of the test grouping:[83] though material and structural similarity act as powerful cohesive forces, a structural group factor, *e.g.*, may be disrupted if any of the component tests are lured away by more influential material affinities; and *vice versa.*

Though the exact *subjective* nature of both the general and the group factors remains a matter of speculation, the group factors

[83] Undoubtedly group factors are also to some extent a function of the "breadth" of the tests used. Spearman has suggested that the marked "cleavage" which appeared between the verbal and numerical groups in Schneck's study (19) was in part due to the "narrowness" of some of the functions tested. No such marked cleavage appears between these two groups in the present study.

located in this study make it plausible to infer that tests which are objectively similar, either with respect to material or with respect to structure, will tend to evoke mental processes which are to some extent similar; and tests which are objectively different, either with respect to material or structure, will tend to evoke mental processes which are in some way different. Whatever the nature of these processes may be, they seem clearly to involve something *in addition* to what is involved in the general factor.

It is difficult to find any objective feature common to all of the tests which throws much light on the subjective nature of the general factor. Though a certain amount of verbal ability was required in all the tests, even in the spatial tests, ''verbal facility'' can hardly be urged as an adequate characterization of the general factor, in light of the relatively low average % variance in the verbal group due to the general factor. Nor is the average % variance of any of the other groups sufficiently high to justify a characterization of this factor as ''spatial facility,'' ''analogical ability,'' etc. There is, however, one feature at least that all the variables have in common: they all involve the ability to see relationships, though the relationships involved may differ in kind and in degree of complexity.

2. Summary

With the purpose of determining whether the structure (or form) of a mental test may not be important in the production of group factors, as the material (or content) has been shown to be, fourteen tests, all but one of them difficult power tests, were given to an unusually homogeneous group of 186 college students in an eight-hour testing program. The tests could be arranged either in three material groups (*i.e.,* four numerical, four spatial, and five verbal tests) or in three structure groups (*i.e.,* four analogies, four generalizations, and five ''construction'' tests).

Application of the simple tetrad criterion failed to give convincing evidence of a third order or fourth order group factor in any group of four or five objectively similar variables. However, the use of ''directed mean tetrads'' gave clear *evidence of first order group factors between pairs of tests similar either in material or in structure, material similarity being the more important.*

The directed mean tetrads also revealed a reliable ''cleavage'' between the following groups: numerical and spatial, numerical and verbal, verbal and spatial, analogies and generalizations. The

most marked cleavage appeared between the verbal and the spatial groups. (The existence of a "cleavage" between two groups by no means indicates a complete absence of relationship.)

The interpretation of (directed) mean tetrads given by Kelley (14) was extended to cover some of the tetrad patterns dealt with by Tryon (30, pp. 415–419), which are more complex than those covered by Kelley. (See Section 4 of Chapter III above.)

Using Hotelling's recently developed "method of principal components," evidence suggesting *second order group factors* in groups of tests similar in material or in structure was found. Again material similarity was more influential than structural similarity.

This new technique also revealed the presence of an *important general factor* running through all fourteen tests, which was better measured by the spatial and numerical test groups than by the verbal group.

In the section on "Interpretation" (preceding section) a rough factor patter was suggested, and some tentative speculations were ventured as to the subjective implications of the group factors found.

BIBLIOGRAPHY

1. Anastasi, A. A group factor in immediate memory. *Arch. Psychol.* No. 120, 1930.
2. ————. Further studies on the memory factor. *Arch. Psychol.* No. 142, 1932.
3. Bailor, E. M. Content and form in tests of general intelligence. N. Y.: Teachers College, Columbia Univ., 1924.
4. Brigham, C. C. A study of error. N. Y.: College Entrance Examination Board, 1932.
5. Brolyer, C. R. A formula for the mean tetrad. *J. Gen. Psychol.* 6, 1932, 212–214.
6. Davey, C. M. Comparison of group verbal and pictorial tests of intelligence. *Brit. J. Psychol.* 17, 1926, 27–48.
7. Dunlap, J. W., and Kurtz, A. K. Handbook of statistical nomographs, tables, and formulas. Yonkers (N. Y.): World Book Co., 1932.
8. Findley, W. G. Specialization of verbal facility at the college entrance level. N. Y.: Teachers College, Columbia Univ., 1933.
9. Fisher, R. A. Statistical methods for research workers. London: Oliver, 1932 (4th ed.).
10. Garrett, H. E. Statistics in psychology and education. N. Y.: Longmans, 1929.
11. ————, and Anastasi, A. The tetrad difference criterion and the measurement of mental traits. *Ann. N. Y. Acad. Sci.* 33, 1932, 233–282.
12. Hart, B., and Spearman, C. General ability, its existence and nature. *Brit. J. Psychol.* 5, 1912, 51–84.
13. Holzinger, K. J. Statistical résumé of the Spearman two-factor theory. Chicago: Univ. of Chicago Press, 1930.
14. Kelley, T. L. Crossroads in the mind of man. Palo Alto (Calif.): Stanford Univ. Press, 1928.
15. ————. Statistical method. N. Y.: Macmillan, 1924.

16. Paterson, D. G., *et al.* Minnesota mechanical abilities test. Univ. Minn. Press, 1930.
17. Peatman, J. G. A study of the factors measured by the Thorndike intelligence examination for high school graduates. *Arch. Psychol.* No. 128 1931.
18. Rogers, A. L. Experimental tests of mathematical ability and their prognostic value. N. Y.: Teachers College, Columbia Univ., 1918.
19. Schneck, M. R. The measurement of verbal and numerical abilities. *Arch. Psychol.* No. 107, 1929.
20. Spearman, C. The abilities of man. N. Y.: Macmillan, 1927.
21. ————. Disturbers of tetrad-differences. *J. Educ. Psychol.* 21, 1930, 559–573.
22. Stephenson, W. Tetrad-differences for non-verbal subtests. *J. Educ. Psychol.* 22, 1931, 167–185.
23. ————. Tetrad-differences for verbal subtests. *J. Educ. Psychol.* 22, 1931, 255–267.
24. ————. Tetrad-differences for verbal subtests relative to non-verbal subtests. *J. Educ. Psychol.*, 22, 1931, 334–350.
25. Terman, L. M. Intelligence and its measurement: a symposium. *J. Educ. Psychol.* 12, 1921, 127–133.
26. ————. The measurement of intelligence. Cambridge (Mass.): Houghton, 1916.
27. Thorndike, E. L. Intelligence and its measurement: a symposium. *J. Educ. Psychol.* 12, 1921, 124–127.
28. ————, *et al.* The measurement of intelligence. N. Y.: Teachers College, Columbia Univ., 1927.
29. Tryon, R. C. Multiple factors *vs.* two factors as determiners of abilities. *Psychol. Rev.* 39, 1932, 324–351.
30. … determiners of abilities.

150.8
A673
no. 156
1933
Archives of psychology

ARCHIVES OF PSYCHOLOGY

List of numbers, continued from inside front cover

79. Memory Value of Advertisements: E. R. Brandt. $1.25.
80. Critical Examination of Test-Scoring Methods: R. G. Anderson. $1.00.
81. Thermal Discrimination and Weber's Law: E. A. K. Culler. $1.75.
82. Correlational Analysis of Typing Proficiency: L. Ackerson. $1.50.
83. Recall as a Function of Perceived Relations: C. B. Key. $1.25.
84. Consistency of Rate of Work: C. E. Dowd. $1.00.
85. Experimental Investigation of Recovery from Work: S. L. Crawley. $1.25.
86. Facilitation and Inhibition: T. N. Jenkins. $1.00.
87. Variability of Performance in the Curve of Work: J. D. Weinland. $1.00.
88. Mental Hygiene Inventory: S. D. House. $1.50.
89. Mental Set and Shift: A. T. Jersild. $1.25.
90. Experimental Investigation of Rest Pauses: C. W. Manzer. $1.25.
91. Routine and Varying Practice as Preparation for Adjustment to a New Situation: L. W. Crafts. $1.00.
93. Speed and Other Factors in "Racial" Differences: O. Klineberg. $1.50.
94. Relation of Reaction Time to Intelligence, Memory, and Learning: V. W. Lemmon. 80c.
95. Is the Latent Time in the Achilles Tendon Reflex a Criterion of Speed in Mental Reactions? G. H. Rounds. $1.25.
96. Predictive Value of Tests of Emotional Stability Applied to College Freshmen: E. G. Flemming. $1.00.
97. Vocabulary Information Test: A. L. Weeks. $1.00.
98. Effect of Temporal Arrangements of Practice on the Mastery of an Animal Maze: S. A. Cook. 80c.
99. Recognition Time as a Measure of Confidence: G. H. Seward. $1.00.
100. Precision and Accuracy: G. W. Hartman. 80c.
101. Group Test of Home Environment: E. M. Burdick. $1.50.
102. Effect of Material on Formal Syllogistic Reasoning: M. C. Wilkins. $1.25.
103. Effect of Incentives on Accuracy of Discrimination: H. C. Hamilton. $1.25.
104. Validity of Norms with Special Reference to Urban and Rural Groups: M. E. Shimberg. $1.25.
105. Blood Pressure Changes in Deception: M. N. Chappell. 80c.
106. Experimental Comparison of Psychophysical Methods: W. N. Kellogg. $1.25.
107. Measurement of Verbal and Numerical Abilities: M. M. R. Schneck. $1.00.
108. Perseverative Tendency in Pre-School Children. A Study in Personality: H. M. Cushing. $1.00.
109. Preliminary Study of the Effect of Training in Junior High School Shop Courses: L. D. Anderson. 80c.
110. Music Appreciation: M. J. Adler. $1.50.
111. Motivation in Fashion: E. B. Hurlock. $1.00.
112. Equality Judgments in Psychophysics: W. N. Kellogg. $1.00.
113. Illusions in the Perception of Short Time Intervals: N. Israeli. 80c.
114. Further Studies of the Reading-Recitation Process in Learning: Skaggs, Grossman, Krueger & Krueger. 80c.
115. Factors Affecting the Galvanic Reflex: R. C. Davis. $1.00.
116. Infant's Feeding Reactions During the First Six Months: R. Ripin. 80c.
117. Measurement of Mental Deterioration: H. Babcock. $1.25.
118. Phenomenon of Postural Persistence: L. S. Selling. $1.00.
119. American Council on Education Rating Scale: F. F. Bradshaw. $1.00.
120. Group Factor in Immediate Memory: A. Anastasi. $1.00.
121. Individual Differences in the Sense of Humor and Temperamental Differences: P. Kambouropoulou. $1.00.
122. Suggestibility in Normal and Hypnotic States: G. W. Williams. $1.00.
123. Analytical Study of the Conditioned Knee-Jerk: G. R. Wendt. $1.25.
124. Race Differences in the Organization of Numerical and Verbal Abilities: J. W. Dunlap. $1.25.
125. Errors of Measurement and Correlation: E. E. Cureton. $1.25.
126. Experience Factors, Test Scores and Efficiency of Women Office Workers: N. Bird. $1.00.
127. Delayed Reactions of Infants: C. N. Allen. 80c.
128. Factors Measured by the Thorndike Intelligence Examination for H. S.; J. G. Peatman. $1.00.
129. Educational Success and Failure in Supernormal Children: J. Regensburg. $1.75.
130. Effect of Practice on Visual Perception of Form: J. P. Seward. $1.00.
131. Relation to College Grades of Some Factors other than Intelligence: D. Harris. 80c.
132. Study of Psychological Differences Between "Racial" and National Groups in Europe: O. Klineberg. $1.00.
133. Emotional Differences of Delinquent and Non-Delinquent Girls of Normal Intelligence: A. Courthial. $1.25.
134. Learning and Retention of Pleasant and Unpleasant Activities: H. Cason. $1.25.
135. Investigation of Brightness Constancy: R. B. MacLeod. $1.25.
136. The Rorschach Test Applied to Feeble-Minded Group: S. J. Beck. $1.00.
137. Retention after intervals of Sleep and of Waking: E. B. van Ormer. $1.00.
138. Stimulus Temperature and Thermal Sensation: F. Heiser. $1.00.
139. Energy Cost Measurements on Curve of Work: H. J. Schubert. $1.00.
140. Technique for the Measurements of Attitudes: R. Likert. 80c.
141. Speed Factor in Mental Tests: P. H. DuBois. 80c.
142. Further Studies on the Memory Factor: A. Anastasi. $1.00.
143. An Experimental Study on Variability of Learning: S. E. Asch. $1.00.
144. Development of Inventory for Measurement of Inferiority Feelings at H. S. Level: R. B. Smith. $1.50.
145. The Psychological Effects of Oxygen Deprivation: R. A. McFarland. $1.50.
146. Relation of Subliminal to Supraliminal Learning: O. A. Simley. 80c.
147. Effects of Noise upon Certain Psychological and Physiological Processes: F. L. Harmon. $1.25.
148. Conditioned Responses in Children: G. S. Razran. $1.50.
149. Resemblance of Parents and Children in General Intelligence: M. C. Outhit. $1.00.
150. Influence of Oral Propaganda Material . . . : K-C. Chen. 80c.
151. Negative or Withdrawal Attitude: H. Pallister. 80c.
152. Judgment in Absolute Units as a Psychophysical Method: J. Bressler. $1.00.
153. Visual Illusions in the Chick: C. N. Winslow. $1.25.
154. Measuring Teaching Efficiency Among College Instructors: G. W. Hartmann. 80c.

DIRECTORY OF AMERICAN PSYCHOLOGICAL PERIODICALS

American Journal of Psychology—Ithaca, N. Y.; Cornell University.
Subscription $6.50. 624 pages annually. Edited by M. F. Washburn, K. M. Dallenbach, Madison Bentley, and E. G. Boring. Quarterly. General and experimental psychology. Founded 1887.

Journal of Genetic Psychology—Worcester, Mass.; Clark University Press.
Subscription $14.00 per year; $7.00 per volume. 1000 pages annually. (2 volumes.) Edited by Carl Murchison. Quarterly. Child behavior, animal behavior, and comparative psychology. Founded 1891.

Psychological Review—Princeton, N. J.; Psychological Review Company.
Subscription $5.50. 540 pages annually. Edited by Howard C. Warren. Bi-monthly. General psychology. Founded 1894.

Psychological Monographs—Princeton, N. J.; Psychological Review Company.
Subscription $6.00 per volume. 500 pages. Edited by Herbert S. Langfeld. Published without fixed dates, each issue one or more researches. Founded 1895.

Psychological Index—Princeton, N. J.; Psychological Review Company.
Subscription $4.00. 300–400 pages. Edited by Walter S. Hunter and R. R. Willoughby. An annual bibliography of psychological literature. Founded 1895.

Psychological Bulletin—Princeton, N. J.; Psychological Review Company.
Subscription $6.00. 720 pages annually. Edited by Edward S. Robinson. Monthly (10 numbers). Psychological literature. Founded 1904.

Archives of Psychology—New York, N. Y.; Columbia University.
Subscription $6.00. 500 pages per volume. Edited by R. S. Woodworth. Without fixed dates, each number a single experimental study. Founded 1906.

Journal of Abnormal and Social Psychology—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $5.00. 448 pages annually. Edited by Henry T. Moore. Quarterly. Abnormal and social. Founded 1906.

Psychological Clinic—Philadelphia, Pa.; Psychological Clinic Press.
Subscription $3.00. 288 pages. Edited by Lightner Witmer. Without fixed dates (9 numbers). Orthogenics, psychology, hygiene. Founded 1907.

Psychoanalytic Review—Washington, D. C.; 3617 10th St., N. W.
Subscription $6.00. 500 pages annually. Edited by W. A. White and S. E. Jelliffe. Quarterly. Psychoanalysis. Founded 1913.

Journal of Experimental Psychology—Princeton, N. J.
Psychological Review Company, 700 pages annually. Experimental. Subscription $7.00. Bi-monthly. Edited by S. W. Fernberger. Founded 1916.

Journal of Applied Psychology—Athens, Ohio.
Subscription $5.50. 400 pages annually. Edited by James P. Porter. Bi-monthly. Founded 1917.

Journal of Comparative Psychology—Baltimore, Md.; Williams & Wilkins Company.
Subscription $5.00 per volume of 450 pages. Three volumes every two years. Edited by Knight Dunlap and Robert M. Yerkes. Founded 1921.

Comparative Psychology Monographs—Baltimore, Md.; The Johns Hopkins Press.
Subscription $5.00. 400 pages per volume. Knight Dunlap, Managing Editor. Published without fixed dates, each number a single research. Founded 1922.

Genetic Psychology Monographs—Worcester, Mass.; Clark University Press.
Subscription $14.00 per year; $7.00 per volume. 1000 pages annually. (2 volumes.) Edited by Carl Murchison. Monthly. Each number one complete research. Child behavior, animal behavior, and comparative psychology. Founded 1925.

Psychological Abstracts—Eno Hall, Princeton, N. J.; American Psychological Association.
Subscription $6.00. 700 pages annually. Edited by Walter S. Hunter and R. R. Willoughby. Monthly. Abstracts of psychological literature. Founded 1927.

Journal of General Psychology—Worcester, Mass.; Clark University Press.
Subscription $14.00 per year; $7.00 per volume. 1000 pages annually. (2 volumes.) Edited by Carl Murchison. Quarterly. Experimental, theoretical, clinical, and historical psychology. Founded 1927.

Journal of Social Psychology—Worcester, Mass.; Clark University Press.
Subscription $7.00. 500 pages annually. Edited by John Dewey and Carl Murchison. Quarterly. Political, racial, and differential psychology. Founded 1929.