R E P O R T   R E S U M E S

ED 010 590   72

DEVELOPMENT OF DIAGNOSTIC ANALYTICAL AND MECHANICAL ABILITY TESTS THROUGH FACET DESIGN AND ANALYSIS.

BY- GUTTMAN, LOUIS, SCHLESINGER, I.M.
ISRAEL INST. OF APPLIED SOCIAL RES., JERUSALEM

REPORT NUMBER CRP-OE-4-21-014   PUB DATE   66
REPORT NUMBER PL-480-1
EDRS PRICE MF-$0.18 HC-$3.84   96P.

DESCRIPTORS- *TEST CONSTRUCTION, *ITEM ANALYSIS, *PREDICTIVE MEASUREMENT, APTITUDE TESTS, DIAGNOSTIC TESTS (EDUCATION), *MEASUREMENT INSTRUMENTS, *SET THEORY, FACTOR ANALYSIS, VOCATIONAL EDUCATION, JERUSALEM, ISRAEL

METHODOLOGY BASED ON FACET THEORY (MODIFIED SET THEORY) WAS USED IN TEST CONSTRUCTION AND ANALYSIS TO PROVIDE AN EFFICIENT TOOL OF EVALUATION FOR VOCATIONAL GUIDANCE AND VOCATIONAL SCHOOL USE. THE TYPE OF TEST DEVELOPMENT UNDERTAKEN WAS LIMITED TO THE USE OF NONVERBAL PICTORIAL ITEMS. ITEMS FOR TESTING ABILITY TO IDENTIFY ELEMENTS BELONGING TO AN ORDERED SET (ANALOGY TESTING) WERE CONSTRUCTED ON THE BASIS OF FACET DESIGN, AND THE ELEMENTS OF THE SET WERE PRESENTED IN SIMPLE DIAGRAMMATIC DRAWINGS. THE TWO TESTS DEVISED IN THIS PROJECT WERE AN ANALYTICAL ABILITY TEST AND A MECHANICAL COMPREHENSION TEST. THE ATTRIBUTES, FUNCTIONS, AND ARRANGEMENTS EMPLOYED IN THE ANALYTICAL ABILITY TEST WERE DESCRIBED. THE ITEMS EMPLOYED IN THE ANALYTICAL ABILITY TEST WERE OF THREE CLASSES (1) CLASS A - WHERE FIGURES VARIED IN SIZE, DIRECTION, OR PLACE, (2) CLASS B - WHERE SHAPE WAS AN ATTRIBUTE WHICH VARIED, AND (3) CLASS C - WHERE FUNCTION WAS EXHIBITED BY DIFFERENT ATTRIBUTES OR DIFFERENT SETS OF VALUES OF AN ATTRIBUTE. DESCRIPTIONS WERE INCLUDED OF (1) DISTRACTORS OF THE ANALYTICAL ABILITY TEST, (2) ANALYTICAL ABILITY TEST BOOKLETS, (3) ITEMS OF THE MECHANICAL COMPREHENSION TEST, (4) MECHANICAL COMPREHENSION TEST BOOKLETS, AND (5) SOME NEW METHODS OF ITEM ANALYSIS. DATA COLLECTION INCLUDED THE ADMINISTRATION OF THE CONSTRUCTED TESTS TO A SAMPLE OF 637 PUPILS IN GRADES 7, 8, AND 9. CONCLUSIONS WERE (1) FACET DESIGN COULD BE EMPLOYED IN ITEM CONSTRUCTION, (2) THE DEGREE OF SUCCESS (TEST DEVELOPMENT) ACHIEVED WAS UNEVEN, DIFFERING WITH THE KIND OF TEST AND THE NATURE OF ANALYSIS CONDUCTED, AND (3) VARIOUS REGIONS OF CLASSES, ARRANGEMENTS, AND FUNCTIONS, AS WELL AS THE SUBTESTS OF THE MECHANICAL TEST SHOULD PROVE IMPORTANT FOR PREDICTIVE MEASUREMENT. (RS)

DEVELOPMENT OF DIAGNOSTIC ANALYTICAL AND MECHANICAL ABILITY TESTS THROUGH FACET DESIGN AND ANALYSIS

Cooperative Research Project No. OE-4-21-014

Louis Guttman, Ph.D., Principal Investigator

I.M. Schlesinger, Ph.D., Co-Investigator

The Israel Institute of Applied Social Research
Jerusalem, Israel

1966

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U.S. Department of Health, Education and Welfare

ACKNOWLEDGEMENTS

It is a pleasant duty to express our thanks to the persons who helped in this study in many ways. Of the present and previous staff of the Institute, J. Norman and S. Rosenbaum helped with the construction of items; Deborah Greenwald and Marlyn Grossman assisted with analysis and interpretation of the data; U. Haim, H. Bardin and S. Seger were responsible for the mechanical processing and computer analysis of the data; Batia Stein and Haya Gratch aided in the stylistic editing of this final report. The painstaking efforts and often great resourcefulness of the above has furthered the progress of this study materially.

We are also grateful to Mrs. Ruth Bauman for her great patience and skill in devising the standard figures of the analytical ability test and for supervising their production.

Collection of data was made possible through the cooperation of Mr. S. Adiel, the Deputy Director of the Ministry of Education and Culture and other officials at the Ministry, as well as of the many principals of schools where the tests were given. Dr. Gina Ortar of the Hebrew University School of Education kindly made available to us the sample of the adults for the analytical ability test. We wish to thank them all for their interest and readiness to help.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

1. OBJECTIVES AND RATIONALE OF THE STUDY

1.1. General Objectives

The present project constitutes a first attempt at implementing a new methodology of test construction and analysis, which is based on face theory. This theory has been developed by the principal investigator, who has also shown the fruitfulness of the facet approach for the reinterpretation of previous research on intellectual abilities (see references 4, 6, 8, 9, 11). However, apart from a few minor attempts (as yet unreported) this is the first time a methodology based on facet theory is systematically used in test development.

Development of this methodology was stimulated in part by the need for further progress in the field of mental testing. Little improvement has been made over the validity ceilings of standard tests attained by Spearman and Binet or in the original Alpha tests of the American army. In most cases, the validity ceilings do not rise above .5 or .6. It would seem that in order to arrive at substantially higher validities, a new approach in test construction would be necessary. As will be shown, the approach taken in the present study, holds promise of eventually bringing about improvement in this direction.

The immediate practical problem leading to this project, was the need for an efficient tool of evaluation for the purposes of vocational guidance and school use. Although it may be assumed that both general intelligence and mechanical ability are needed for success in vocational training, there is as yet insufficient information on the relationship existing between the different kinds of abilities in this area.

Tests developed in this project differentiate more sharply between general intelligence and mechanical aptitude, than do previous tests. Thus, decisions as to whether a particular student is better qualified for academic or for vocational training can be made more accurate, as can differential placement within vocational areas. Analyses of the interrelationships between different abilities that have been carried out in the present study represent a first step towards attaining this objective.

In view of the varied degrees of literacy in the populations in which the tests might be used, we restricted ourselves to pictorial tests. Two types of tests were developed in this project: (1) Analytical ability tests; these employed abstract figures and were completely nonverbal. (2) A mechanical comprehension test in which each item presented a situation pictorially about which a question was asked verbally. Various analyses were carried out on the data obtained with these tests. In some of these, new methods of analysis recently developed by the principal investigator were employed. As will be seen in the following, the results attest to the potential fruitfulness of the facet approach in predicting both the degree of difficulty and the structure of the interrelationship between various tests and subtests.

1.2. Rationale of Item Construction

The approach to mental ability tests which is taken in the present project has been described previously by the principal investigator (see references 4, 6, 8, 9). It was first developed during weekly meetings held during the course of several years with members of the Israel Army Psychotechnical unit, where a new type of definition of intelligence, involving *analytical ability*, was arrived at. This new definitional approach aimed, among other things, at ascertaining what common system underlies the testing programs conducted in England, America and other countries over the past sixty years. A definitional system of what all these programs have in common was developed. Here a "faceted definition" was employed (see reference 9).

Given any two sets of elements A and B, which we call *facets*, their Cartesian space is the set of all pairs of elements *ab*, where *a* is an element of A and *b* is an element of B. A Cartesian space may consist of any number of facets, or sets of elements; with *n* facets, any one point in the Cartesian space has *n* component elements.

A set of abilities of a given kind is defined simultaneously by a "faceted definition" which makes use of two or more facets, in such a way that any element of the Cartesian space defines one variety of the ability.

The definition which was finally arrived at is the following:

> An act of a subject is *intelligent* to the (extent) to which it is classified by a (tester) as (demonstrating) a *correct* perception of an unexhibited *logical* (aspect)

of a (relation) intended by the tester, on the basis of another (exhibited) logical (aspect) of that relation that is correctly perceived by the subject.

Italics are used to indicate the constant features of the concept of intelligence, while the parentheses indicate facets, which are sets of ideas rather than constants.

The substance of a question in an intelligence test can be characterized as exhibiting a logical aspect of a relation. If the subject's response is correct, it will demonstrate a correct perception of an unexhibited logical aspect of the same relation. Given this focus on logical aspects of relations, items which test analytical ability and items which test achievement can be distinguished. If the name or the selection rule of the relation is exhibited, the test is of achievement. If the rule is to be inferred, the test is of analytical ability.

To illustrate this distinction, consider the following two items:

(1) Who was the first president of Israel?

(2) A dog is to a puppy as a cow is to a ___________.

In the first question the selection rule of the relation - first presidency of a country - is exhibited. The Cartesian space is that of all first presidents against all countries. The exhibited relation is that each country is paired only with its own first president. One element of this relation, namely Israel, is also exhibited, and another aspect of the relation is required. It is assumed that the subject knows what a "president" means, what "country" means, and what "Israel" means. What is being tested is whether or not the subject possesses the information. This is therefore a test of achievement. By contrast, the second item exhibits neither the name nor the culling rule of a relation. Instead, it exhibits an ordered pair (dog-puppy) and the first element of a further ordered pair. Evidently, the intention is to see if the subject can infer a two-faceted Cartesian space, within which there is a binary relation, of which the two ordered pairs are elements. This item is, therefore, one of analytical ability.

The way in which the relation is exhibited is an important facet of the definition. Any of the senses could be used as a mode of communication, and each mode may define a particular kind of intelligence, as is evidenced by the intercorrelation matrix obtained in previous research (see references 8,9). Even within the visual mode, using paper and pencil tests, there are different kinds of languages: written speech, formalized language such as arithmetic and algebra, and pictorial language.

In focussing only on pictorial tests, we are essentially holding a major facet constant, which may help clarify the relation of the empirical results to the facets which are systematically varied in our tests.

The approach towards test construction outlined above allows a more systematic procedure of item construction (see section 1.3). Furthermore, the faceted definition according to which items are constructed, leads to a prediction of the structure of interrelationships between items and subtests. In general, the relationship between items within the framework of facet design should be expected to have its counterpart in the empirically obtained correlation matrix, where the size of the correlation is related to similarity of facet profiles. These notions will be further elucidated when the method of item construction and the results of the analyses are described (see especially sections 2.1 and 3.1.3).

1.3. Practical Advantages of the Approach Taken

Various practical advantages may be expected to accrue from the facet theory approach which the present project has tried to implement. These may be treated under three headings:

(a) internally systematic construction of items and distractors;

(b) enhancement of the possibility of revealing lawful empirical relationships among items and tests;

(c) using knowledge of the content and statistical structure for the parsimonious utilization of the test in the prediction of external criteria.

The construction of test items has typically been a more or less intuitive process. The investigator usually includes in a first trial those items he can think of and which, on the face of it, seem to him to belong to the universe of items he desires. The decision as to whether or not he has had the right hunch is made to depend on subsequent item analysis. By such a trial-and-error procedure, he is finally left with a set of items which he includes in his tests.

Section 1.3 (2)

The fallacy of using such statistical techniques as a criterion for content has been pointed out many years ago (see reference 2, pp.181-185). By contrast, the facet approach supplies the investigator with a tool for constructing items systematically. Intuition still has a part to play, since there is no way of formalizing all aspects of the process of arriving at a faceted definition. However, once such a definition has been set up, the investigator is in possession of the building-stones of the items he needs; the construction of items becomes thus a much more simple and straightforward process. No fallacious use of statistics is involved.

The suitability of an item for a given test or subtest is decided upon on a priori definitional grounds, instead of by subsequent statistical item analysis. Analysis of correlations between items is employed only to test an empirical hypothesis on the relation of the observed statistical structure to the faceted definitional design.

Not only the items themselves, but also their distractors can be systematically constructed on the basis of those facets which were employed in devising the items. This is an important step in formalizing the process of items construction, because distractors are clearly an integral part of the item.

An illustration of how distractors are constructed through facet design is given in the present study (section 2.2), where it is also shown that this method lends itself to the systematic construction of distractors differing in degree of attraction. This makes it possible to assign meaningful differential scores to each item, instead of merely scoring answers as being either "right" or "wrong" as is customarily done. By assigning such scores it may become feasible to employ tests with a much smaller number of items, yet without loss of discriminatory power.

The formalization of item construction is of special importance when parallel forms of a test have to be prepared. Unless a facet definition is employed, there is no way of hypothesizing beforehand whether items included in the two forms may be of comparable difficulty, or, more important, of justifying the assertion that they define the same kind of ability. By constructing items as well as distractors on the basis of a faceted definition, it becomes possible in principle to devise truly parallel forms, both in the sense of content and of statistical structure. This objective will be attained to the extent that the investigator has succeeded in including in his faceted definition all the behaviorally relevant aspects of items (see also section 3.1.5).

So far, the advantages of the approach taken for the construction of items and distractors has been discussed. Once the initial form of the test has been constructed in this manner, it may be expected that the facet design will open up opportunities for further development of the test. Analysis of the test results will show to what extent the facet structure of the items is reflected by their intercorrelational structure. This may lead to a further sharpening of the facet design as well as to the construction of additional types of items with other facet profiles.

Section 1.3 (4)

The facet approach should ultimately lead to a statistical mapping of the structure of the abilities tested, as well as to establishing their empirical relationship with various criteria. To the extent that the investigator succeeds in this, it will become possible to limit the number and the length of tests employed for purposes of prediction by using only those test-items which correlate most strategically with the criterion variable. Furthermore, a successful facet design should make it possible to devise shorter forms of a test without thereby appreciably reducing its reliability.

2. CONSTRUCTION OF TESTS AND PROCEDURES

2.1. Items of Analytical Ability Test

Certain decisions had to be made regarding the type of test which would be constructed within the framework of the present project. One possibility would have been to devise tests encompassing a wide range of abilities. Data on the relationships between different abilities can be found in the literature, and reanalyses of such results have been published by the principal investigator elsewhere (see references 8, 9). These appear to show the fruitfulness of the facet approach. Although further work with tests tapping a wide range of abilities needs to be done, it appeared desirable in the present project to limit ourselves to the area of analytical ability, using non-verbal items. It was finally decided to construct items testing the ability to identify elements belonging to an ordered set (so-called analogy tests are an example of this type), where the set consists of simple geometric drawings.

Even within this limited area, an indefinitely large number of items can be constructed on the basis of a facet design. However, should the facet approach lead to the prediction of the structure of interrelationships between them, a relatively small sample of items could be used to represent the universe. If such structural information could result *within* such a narrow area, this would certainly justify expecting clear structure among areas which differ more grossly among themselves.

The definitional differences among the items of the narrow area we have concentrated upon in this project, are small compared to the differences in the design between the tests analyzed previously, so that we have weaker expectations of finding a stable statistical structure in the empirical data.

When the type of analytical ability selected for this project was explored, it became apparent that it included a large number of subtypes. A general framework for this type of test was arrived at, which permits the construction of a variety of tests. Hence, an additional decision had to be made, as to what kinds of items should be included within the present project. In the following sections, the test items which were finally decided upon, will be described.

The test items were in some ways similar to existing analogy tests employing figures; a well-known example are Raven's Progressive Matrices. In the following, the final forms of the test items, arrived at after an extensive pilot study, will be described. Examples of items are given in the appendix.

An item of the analytical ability test consists of a three by three matrix of squares in which the middle square (i.e., row two, column two) is missing. Each square contains a figure. The task of the subject is to supply the missing square by choosing one of the alternative answers given. In order to accomplish this task the subject must presumably find the rule of construction of the item. This rule is exemplified by the squares of the matrix; in simpler items - in the right and left hand columns of the matrix and/or in the upper and lower row of the matrix. In more complex items, the rule is exemplified in a different way, as will be explained below. In this manner "a logical aspect of a relation" is exhibited (see the definition above, section 1.2). The subject shows his understanding of the rule by selecting the correct square from the given alternatives.

The rule of construction gives the arrangement of three sets of three figures in a three by three matrix. In each set, the values of the attributes of the three figures and the way in which these values are ordered within each set are described by functions.

Using example 1 of the appendix as an illustration, it can be seen that place (within the small square) and size are varied. Size varies in the order: small, large, medium; that is, by function b* (see below). Place varies in the order: upper left corner, upper left center, upper left corner; this order is termed function a* (see below).

The arrangement in example 1 is by rows and columns; that is, function b* orders size by columns and function a* orders place by rows.

In the following, the attributes, functions and arrangements employed in the analytical ability test are described.

Attributes

There are two types of attributes:

I. Attributes whose values can be ordered on a continuum such as size, number, etc.

II. Attributes which cannot be thus arranged, such as shape.

Section 2.1 (4)

Inspection of the examples in the appendix shows that figures of an item in this final version of the test vary in two or more of the following attributes:

Type I: Size, direction, and place (within the small square).

Type II: Shape (e.g. example 2 in the appendix)

<u>Functions</u>

In items where the figures varied in attributes of type I, the order of the values of this attribute are given by the function. Inspection of example 1 shows that in the top and bottom rows figures are arranged by size in the following order from left to right: small, large, medium. The subject is made to understand by means of the instructions that the middle row should exhibit the same characteristics as the top and bottom rows, and that the middle figure should be largest in size.

Within each item an attribute may have either <u>two</u> or <u>three</u> values. The two-value functions are either <u>a</u> or <u>a*</u>, and the three-value functions - <u>b</u> or <u>b*</u>. These are defined as follows:

<u>a</u>: f, f, g
<u>a*</u>: f, g, f
where $f > g$ or $f < g$

<u>b</u>: f, g, h
<u>b*</u>: f, h, g
where $h > g > f$ or $h < g < f$

It will be noted that <u>a*</u> and <u>b*</u> are permutations of <u>a</u> and <u>b</u> respectively.

Examples of these functions are:

a - See example 3 in the appendix. The two different values for direction are arranged in the columns in the order f, f, g, as given by function a.

a* - See example 1. There are two different values for the attribute 'place' - upper left corner and left center - arranged in the rows in the order f, g, f, as given by function a*.

b - See example 4. Three different values for direction (north, northeast, east) are arranged from left to right in the order f, g, h, as given by function b.

b* - See example 4. There are three different values of size, arranged in the columns in the order: small, large, medium, or f, h, g, as given by function b*.

The above applies only to type I attributes. Type II attributes, such as shape, cannot be said to vary by functions, since they cannot be ordered.

In the research leading to the construction of these tests, other functions were also experimented with. It appeared, however, that it was impractical to employ them in a three by three matrix, since such a matrix would not allow for an exhibition of the function, which would enable the testee to find an unequivocal solution.

In an earlier version of the test, an additional function was employed: f, g, h(f,g) - i.e., the third value, h, was a combination of f and g. This involved drawing a compound figure; for instance, if f and g are a small and a large triangle respectively, h might be a small triangle enclosed within or alongside a large triangle. However, when working with such items, it was found that the subject did not pay attention to the relationship defined by the function, but rather to the number of figures, or else, he saw in the compound figure a new and different shape. The use of this function was henceforward dispensed with. This experience serves to elucidate an important point regarding the construction of items, namely, that aspects of an item attended to by the subject may not be those defined by the facet design as intended by the researcher. Only careful pretesting or an analysis of the data can show whether this is the case. This point will be taken up again in the discussion of the results below (section 3.1.6).

Arrangements

Each function orders the values of a set of three figures. There are three such sets in a three by three matrix, and there are different ways in which these sets can be arranged in the matrix - by rows, by columns and diagonally:

Rows (R): Arrangement by rows is shown by example 4, for instance, in which each row has a different value of place (within the small square).

Columns (C): An illustration of arrangement by columns are direction and size in example 4. The values of these attributes differ in each column.

The arrangement in example 4 is, therefore, CCR (direction, size, and place).

Section 2.1 (7)

<u>Diagonals</u> (D or D'): Let x, y and z be values of the attribute. These can be arranged diagonally in either one of the following ways:

D:				D':		
x	y	z		x	y	z
z	x	y		y	z	x
y	z	x		z	x	y

The arrangement here is similar to that of a "latin square" and the diagonal may run from the upper left hand corner to the lower right hand corner (D) or else from the upper right hand to the lower left hand corner (D'). Any one of the values f, g, and h may be assigned to the positions in the above diagram that are indicated by x, y, and z.

The following arrangements of D show how the four functions defined previously were arbitrarily arranged diagonally. (Note that the permutations <u>a</u> - <u>a*</u> and <u>b</u> - <u>b*</u> have been defined for the case where the values are arranged either in rows or in columns). For the "latin square" arrangement these have been arbitrarily redefined so as to specify which value occupies the main diagonal: for <u>a</u>, this is the value occurring twice within the row or column (g); for <u>a*</u>, the value occurring once (f); for <u>b</u>, it is one of the extreme values (f or h); and for <u>b*</u>, the middle value (g). For D', a similar arrangement of functions was used (with diagonals reversed).

Function a			Function a*			Function b			Function b*		
g	g	f	f	g	g	f	g	h	g	h	f
f	g	g	g	f	g	h	f	g	f	g	h
g	f	g	g	g	f	g	h	f	h	f	g

Test Items

Items employed in the tests are of three classes:

Class A: Class A includes items in which the figures vary only in one or more of the three attributes: size, direction, or place. An item of this class can be defined by specifying the function and arrangement of each attribute varied. For instance, example 5 is defineable as:

place:	b	D'
direction:	a*	D
size:	b*	R

That is, place varies diagonally according to function b, direction by the other diagonal according to function a*, and size by rows according to function b*.

Items in which either two or three attributes vary were also included in the tests. The combinations of arrangements employed rendered the subtests described in Table 2-1.

Table 2-1

Subtests of the Analytical Ability Test - Class A.

	Two Attributes	Three Attributes
No diagonals	RC	CCR
One diagonal	DR	CRD
Two diagonals	D'D	D'DR

Thus, in the RC subtest there are two attributes, and the figures are arranged by rows and columns only (see example 1), and in the D'DR subtest there are three attributes, and the figures are arranged by each of the two diagonals as well as by rows (see example 5).

The universe of items defined by the above specifications was sampled and 48 test items were constructed.

In addition to the specifications outlined above, it was necessary to make certain arbitrary decisions, e.g., as to which sizes, directions and places to employ, whether f, g, g in items of function a should be arranged from top to bottom (right to left) or vice versa, and so on.

Class B: This class includes items where shape is an attribute which is varied. For instance, in example 2 three different shapes are arranged by columns. Items of this class may vary also in other attributes according to one of the functions and arrangements described above: e.g., example 2, in addition to having shape as one of the attributes varied, has figures varying in place: b* R; and in size: b R.

Class C: Solution of items of this class presumably requires a greater degree of abstraction. The main feature of these items is that the function is exhibited either by means of different attributes or by means of different sets of values of an attribute in the right and left hand columns, and the subject is explicitly instructed to disregard arrangement by rows. In example 6 for instance, in the left hand column, the function b is exhibited by means of the attribute place; in the right hand column the function is also b, and is exhibited by direction. From these two columns the subject must find the rule, i.e., that function b must be employed on any attribute. The figures in the middle column vary neither by place nor direction but by size. Employing function b, the subject will understand that the correct answer is distractor No.8.

Each column may exemplify more than one function as is the case in example 7. In the left hand column the figures vary by size (function b*), by place (function b), and by direction (function b*). In order to arrive at the correct choice in the middle column, the same functions must be employed, though not necessarily for the same attributes. It can be shown that No.1 is the only appropriate answer amongst the alternatives presented.

2.2. Distractors of the Analytical Ability Test

The usual method of constructing distractors calls for the creation of such alternatives as the subject is likely to choose; this is determined either by the intuition of the test constructor or, empirically, by first presenting the test items to a group of subjects in an open-ended form. By contrast, the method employed in the present study relies on the facet design used in item construction, which allows for the systematic construction of distractors. Since the choice of distractors must be assumed to determine to a large extent item difficulty as well as the relationships of the items to one another, the construction of truly parallel forms of the test is not possible without employing this approach, which is based on facet design (see section 1.3).

Another objective was that of obtaining distractors whose order of difficulty is predictable a priori, (to be verified, of course, by the test results). This was achieved by manipulating systematically the degree of similarity between a given distractor and the correct answer.

Take, for example, an item in which three attributes are varied. The correct answer presents, by definition, the correct value on all three attributes. The distractors most similar to this answer have the correct value on two of the three attributes and an incorrect value on the third attribute. A less similar distractor has a correct value on only one attribute. Finally, the distractor least similar to the correct answer has incorrect values on allthree attributes. Example 4 of the appendix, which is a Class A item, shows that there are eight distractors (including the correct answer) differing in respect to the attributes which have the correct value, as shown in Table 2-2.

For the incorrect value of the attribute a value was chosen which: (a) appeared in one of the 8 squares of the test item, and (b) was closest to the correct value (e.g., in example 4 where the correct answer is a figure which is large in size, the distractors which have incorrect values for this dimension are medium, not small in size.)

Table 2-2

Design of Distractors in Three-Attribute Items - Example 4

	Size	Direction	Place	Distractor No.
Correct answer	+	+	+	7
1 attribute wrong	+	+	-	2
1 attribute wrong	+	-	+	1
1 attribute wrong	-	+	+	4
2 attributes wrong	-	+	-	8
2 attributes wrong	-	-	+	6
2 attributes wrong	+	-	-	5
All attributes wrong	-	-	-	3

+ The distractor is correct on this attribute

- The distractor is wrong on this attribute

A distractor constructed by the above rules could be identical with one of the 8 squares of a test item, e.g. distractor No.2 in example 4 which is identical with the middle square in the top row. This might entice the subject to choose such a distractor, only for this reason. Care was therefore taken to have all items of the test, as far as possible, include an equal number of distractors identical with one of the squares.

The number of distractors for two-attribute items was also kept at eight, so as not to introduce an additional determinant of difficulty. This did not make it possible to have a design of distractors which was as "tidy" as the one used for the three-attribute items, but as far as possible the same principles were observed. The design of these items is illustrated in Table 2-3 (by example 1 of the appendix).

Table 2-3

Design of Distractors in Two-Attribute Items - Example 1

	Size	Place	Distractor No.
Correct answer	+	+	3
1 attribute wrong	+	-	4
1 attribute wrong	+	-	7
1 attribute wrong	-	+	8
1 attribute wrong	-	+	5
2 attributes wrong	-	-	1
2 attributes wrong	-	-	2
2 attributes wrong	-	-	6

+ The distractor is correct on this attribute

- The distractor is wrong on this attribute

As to distractors of items belonging to Classes B and C, the reader can satisfy himself by inspection that they have been constructed in a similar systematic way, but space does not allow us to go into details.

2.3. Analytical-Ability Test Booklets

In the pilot study, after all items were drawn by our draftsman, it soon became clear that the degree of precision required for the drawings for this kind of test is very great. This tended to make the drawing of items very expensive; therefore a different procedure was adopted for the next version of the experimental test. Instead of drawing each individual item, standard figures were designed by our draftsman. There were six different figures, and each of these was drawn in four sizes, four directions, and four places. The resulting 384 combinations were photographed and printed. Thus we had at our disposal 384 different kinds of labels, bound in small booklets. Moreover, sheets were prepared on which the empty frame of an item appeared.

Such materials may serve for the construction of an unlimited number of items; all that remains to be done is to paste the appropriate labels in the correct squares, and then mimeograph the sheet. The work of the draftsman can thus be dispensed with and, though it involves an initially larger expense, our procedure is actually more economical in the long run. Moreover, the use of these standard figures permits a far greater degree of precision than is usually attained otherwise. Another advantage of the procedures is that the differences between any given attribute remain constant between items. This eliminates the possibility of the draftsman inadvertently determining the degree of difficulty of an item by introducing different degrees of similarity between values.

Section 2.3 (2)

For the first phase of the study, four alternative forms of the test were prepared. Each form contained thirty-two items and four practice items. The items were allocated to the different forms so that each possible combination of two items of Class A would appear in at least one of the forms. (Practical considerations did not permit implementing such an allocation for items of Classes B and C.) This allowed us subsequently to compute correlations between any two items.

The four forms were designed to contain the following subtests; each subtest consisted of eight items.

<u>Form 1:</u> DR, D'D, CCR, CRD of Class A.

<u>Form 2:</u> RC, D'D, CRD, D'DR of Class A.

<u>Form 3:</u> RC, DR, CCR, D'DR of Class A.

<u>Form 4:</u> RC, DR of Class A; Class B; Class C.

In Class A, for the two-attribute subtests, eight combinations of attributes and their functions were chosen, and these remained the same in all three groups. Thus, an item having the following attributes and functions: size: <u>a*</u>; place: <u>b</u>, appeared in subtests RC, DR, and D'D. Similarly, items for the three-attribute subtests were equal in the attributes and their functions.

For the second phase of the study, another form, Form 5, was constructed, the items for which were chosen on the basis of experience gained in the first phase. Specifically, it was deemed advisable to include items of each one of the above groups. Further, additional items of Class C and Class B were constructed. Form 5, then, included three items each of RC, DR, D'D, CCR, CRD, and D'DR, six items of Class B and nine items of Class C (and three practice items).

In Form 5, as in the previous forms, the two-attribute items within Class A had parallel items. The same was true of the three-attribute items in Class A.

2.4. Items of Mechanical Comprehension Test

One of the major distinctions we have attempted to make is between "knowledge of physics" and "mechanical comprehension". For success in a vocation involving handling tools and the like, knowledge of physics is certainly helpful. But practical know-how with instruments is more immediately relevant, and such know-how may exist "viscerally" without explicit technical grasp of any physical laws that may be involved. One may be able to drive a nail well with a hammer without being able to state correctly any laws about force, acceleration, and the like.

Existing tests were examined from this point of view, and appear to contain a mixture of items. Some are on pure physics; others are on how to operate something mechanical - which could be answered correctly without any clear understanding of the physics involved.

The more purely "mechanical comprehension" items were culled by the rule that they should deal with some real-life process like moving something, lifting, revolving, rolling, dropping, breaking, etc., in which real-life instruments are involved and which are of common occurrence. This requirement ultimately defined one facet (Facet F, below) for item construction. This requirement is also practical for 8th-graders, for whom it would not be very appropriate to test "knowledge of physics"(even if this were relevant for vocational guidance.)

Existing items fulfilling the "mechanical comprehension" requirement were looked at closely to see what further facets seemed implicit in them. An earlier version of the faceted definition for this test was quoted in the research proposal. In the course of our work this definition was extended and sharpened so that it can now define any item of such a test and also be used to generate additional items.

The revised definition runs as follows:

Which of two or more (A: 1. absolute / 2. relative) states of (element B_i) will (C: 1. lead to / 2. result from) a relatively appropriate specified (D: 1. absolute / 2. relative) state of (element B_j) in a situation involving (E: 1. absolute / 2. relative) (mechanical activity of type F_i) where there (G: is / is not) an intermediate consideration.

Section 2.4 (3)

List of elements in facets B and F

B_1 (attributes of motion)	B_2 (geometric attributes)	F (type of activity)
1. length of path	1. place	1. moving
2. shape of path	2. shape	2. lifting/lowering
3. direction	3. volume	3. revolving
4. velocity	4. circumference	4. keeping in position
5. acceleration	5. length, height (where volume is not relevant)	5. rolling
6. duration	6. angle	6. dropping
7. amount of force	7. relative position	7. breaking
8. distribution of force (torque)	8. area	8. arresting movement
		9. starting movement
		10. keeping path

The definition can best be illustrated by examples taken from the test. For instance, let us contrast the definition of a question asking which arm of a rotating sprinkler will be the first to arrive at a given point (see the example given in the appendix), with that of a question asking whether a round or many-sided object will travel farther with the acceleration gained in traversing an inclined plane. The first question asks which position willresult from a specified shape (B_1: 2.7; C: 2; B_j: 2.2), while the second requires the shape which leads to a specified length of path (B_1: 2.2, C_1, B_j: 1.1). Now, in the second question, two different shapes are given and also a relatively greater length of path is asked about. Therefore, for both facets A and D, the second element ("relative") applies. On the other hand, in the sprinkler question, only one shape is given and therefore "absolute" in facets A and D is the appropriate description. The sprinkler question involves the mechanical activity of revolving (F_3) while the balls question deals with rolling (F_5). In the sprinkler question, in order to determine how the shape of the sprinkler leads to the specified position, the subject has to consider in which direction the sprinkler will move. There is therefore one intermediate consideration, that of direction. In the other question, the solution can be arrived at directly without any intermediate considerations. (Both questions deal in situations involving absolute mechanical activity (E_1); there are no questions involving relative mechanical activity in our test.)

The fruitfulness of the definition is attested to by the fact that it has resulted in the construction of items not previously included in mechanical comprehension tests. Examples are questions about the possible influence of a curved accelerating path on the motion of an object after it leaves the path, about the relative efficiency of hammers with handles of different lengths (see appendix), and about the influence of the screw threads on the direction the screw must be turned and the depth to which it can penetrate.

It became clear to us, however, that it would be impossible to construct items for all structuples (combinations of elements from all facets) of the above definition because some of them did not seem to correspond to any empirical reality. We therefore proceeded to construct items of as many structuples as was feasible and sampled from these 81 items for the test.

As shown in the examples given in the appendix, each item consists of a drawing, a question, and three alternative answers. The answer "same", "no difference", and the like, is the correct answer for very few items included in the test, as well as for one of the examples given at the beginning of the test.

2.5. Mechanical-Comprehension Test Booklets

Items to be included in the test were divided into three groups of 27 each. Three alternative forms of the test were constructed, each of which included two of the three groups. Each form thus included 54 items. This made it possible subsequently to compute correlation coefficients for any two of the 81 items.

2.6. Collection of Data

The three forms of the mechanical comprehension test and Forms 1 - 4 of the analytical ability test were administered to a sample of 637 pupils which was constituted as follows:

Elementary Schools		
Grade Seven	250 pupils	
Grade Eight	78 pupils	328
High School - Grade Nine		
Vocational	140 pupils	
General	169 pupils	309
		637

The vocational high school classes contained only boys; all other classes were mixed. A total of 417 boys and 220 girls were tested.

The mechanical and analytical ability tests were both given on the same day. Each test required 1/2 - 3/4 of an hour. The alternative forms of each test were distributed amongst the students according to the order of seating, students sitting next to each other always getting two different forms.

Form 5 of the analytical ability test was given at a later date to a sample of 511 adults of both sexes.

The subjects wrote their answers on answer sheets. The answers were subsequently transferred to punch cards.

2.7. Some New Methods of Analysis

In order to analyze the structure of interrelationships between items and between groups of items in both the analytical ability and the mechanical comprehension test, a new method of multivariate analysis was employed. A program of Smallest Space Analysis has been recently developed by Guttman and Lingoes (the G-L SSA-I). This is essentially a symmetric analysis resulting in a parsimonious Euclidean presentation which has a monotone relation to the originally given distances between variables (see references 10, 11, 14, 15). The order of these distances, but not their absolute sizes, is preserved, and the program derives the smallest possible Euclidean space for these items.

For the purpose of our test, the distance function D_{ij} used (where i and j are any two items), was defined as follows:

Let $e_{si} = \begin{cases} 1, & \text{if subject } s \text{ answers item } i \text{ correctly} \\ 0, & \text{otherwise} \end{cases}$

Then $D_{ij} = \underset{s}{E}(e_{si} - e_{sj})^2$

where E denotes the expected value over the indicated subscript. Expanding the right member we obtain

$$D_{ij} = p_i + p_j - 2p_{ij}$$

where p_i = the proportion of subjects who answered item i correctly

p_{ij} = the proportion of subjects who answered both item i and item j correctly.

D_{ij} varies between 0 and 1. The coefficient of similarity between two items may be defined as:

$$C_{ij} = 1 - D_{ij}$$

This coefficient also varies between 0 (perfect dissimilarity) and 1 (perfect similarity). It has properties making it especially suitable for the analysis of test items. If D_{ij} is used directly as a distance function, a group of items which forms a perfect scale will fall on a straight line. If the items of the test can be described adequately by a three-dimensional space, then our coefficient of similarity tends to make the first principal axis of the space present essentially the order of difficulty (p_i) of the items. Hence, the relationship between content of items (the description of which is usually attempted by coefficients such as Pearson's r) is being described by the two-dimensional space of the remaining two axes.

Section 2.7 (3)

In addition to printing out the coordinates on each of the principal axes for each item, the SSA-I program also prints out the corresponding Shepard diagram and coefficient of alienation. The Shepard diagram is essentially a scattergram where each point represents the distance between two items; one axis represents the original coefficient of similarity or distances (as defined by the above coefficient C_{ij}) and the other axis presents the distance in the n- dimensional space calculated by the program. The coefficient of alienation refers to this relationship between distances and varies between 0 and 1.

3. FINDINGS

3.1. Analytical Ability Test

3.1.1. Relative Difficulty of Subtests

It was found that the facet structure of the items could predict the comparative difficulty of subtests, as measured by the mean number of items (0 to 8) answered correctly.

Class A: In constructing the items it was assumed that items varying in two attributes would be more difficult than those varying in three. In addition, arrangements by diagonal were expected to create more difficulty than arrangements in either horizontal or vertical direction. These two predictions give rise to an hypothesis of partial order of difficulty, which is illustrated in Figure 3-1, where subtests are arranged from top to bottom in increasing order of difficulty.

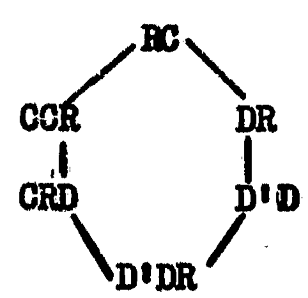


Figure 3-1. Hypothesized Relative Difficulty of Class A Items

Inspection of Figure 3-2 shows that the prediction is fully borne out. The bar-diagrams give the distribution of correct answers in the average

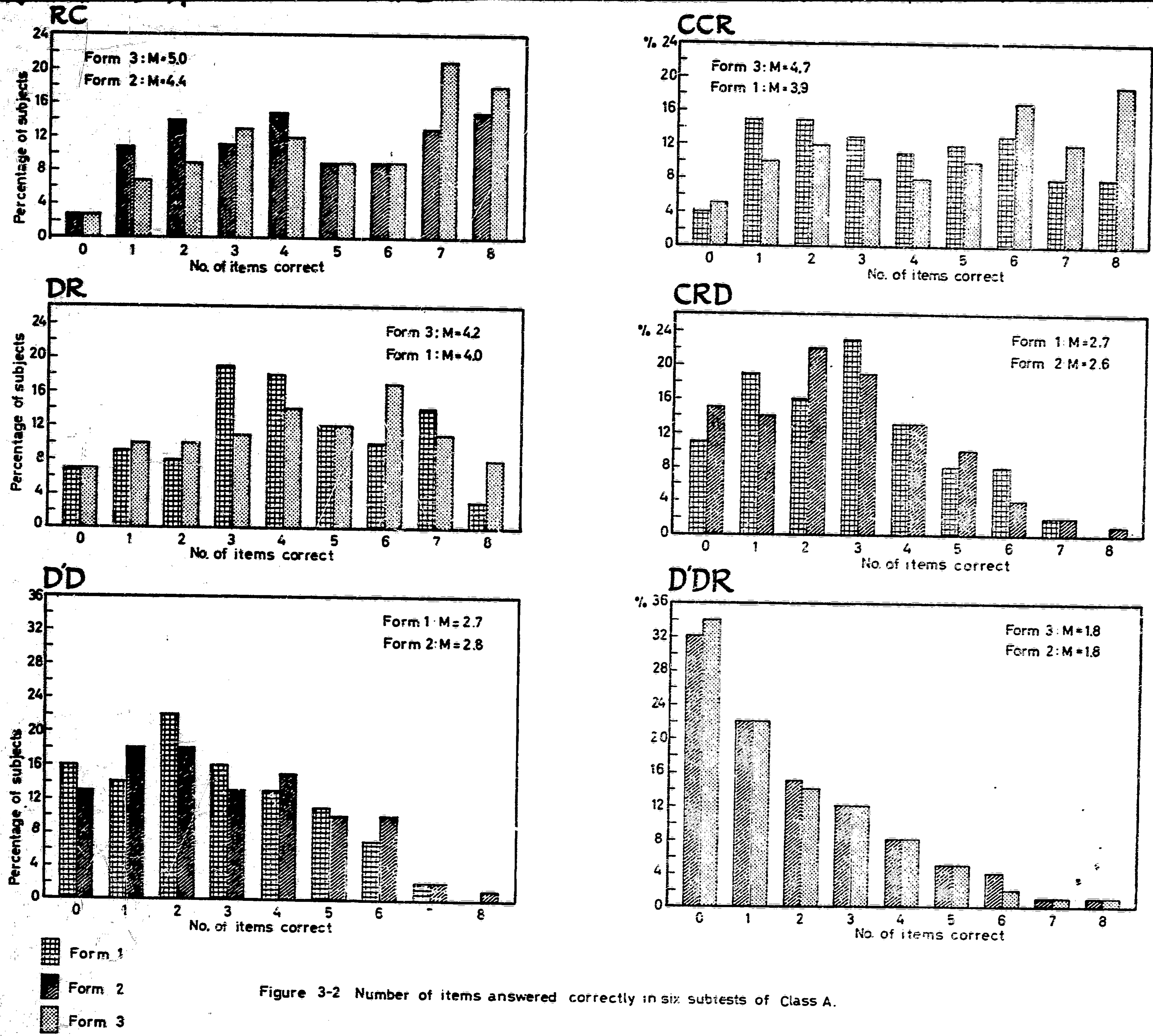


Figure 3-2 Number of items answered correctly in six subtests of Class A.

number of items answered correctly for different Class A subtests. Data for two alternative forms for each group of items in this Class are included, making it evident that the findings were replicated and can be regarded as reliable.

As stated, only a partial order could be hypothesized. Thus, no predictions were made about the difficulty of CCR as compared to DR or that of CRD as compared to D'D. The relative difficulty of the members of these pairs of groups can be determined empirically. In our case, however, the differences were very small and inconsistent in direction. This suggests that CCR is roughly equal in difficulty to DR, and CRD to D'D.

The order of difficulty shown in Figure 3-2 was replicated in Form 5 where all six subtests of Class A were included (with three items in each subtest).

<u>Class B</u>: Items in Class B in which shape was varied, were on the whole easier than Class A items. This was to be expected, since the shape of the figure is intuitively more salient than such attributes as its size or direction, and figures differing in shape appear more different from each other than figures differing, e.g., only in size.

In Form 4, the average number of items answered correctly in Class B (4.3 items) was between that for group RC (4.9) and DR (3.9). Within Class B, items of varying degrees of difficulty are to be found. These items vary also in other attributes in addition to shape (see section 2.1).

Section 3.1.1 (3)

In Form 5, some of the Class B items have arrangements by rows and columns only, as is the case with Class B items in Form 4. These were again intermediate in difficulty between subtests RC and DR of Class A. However, another kind of Class B items included in Form 5, in which two attributes were diagonally arranged, was much more difficult than these subtests; the number of subjects which solved these items correctly was not much higher than that which solved correctly the most difficult items of Class A, D'DR, which also had two attributes diagonally arranged.

Class C: In Form 4, Class C items were the most difficult; the average number of items answered correctly was 3.2, as compared to 4.9, 4.3 and 3.9 for the other groups in this form (see above). Since the means tended to remain fairly constant between the different test forms, it can be said that Class C items with a mean of 3.2, are more difficult than groups CCR and DR and less difficult than CRD and D'D (see Figure 3-2).

In Form 5 a new kind of Class C item was introduced. In these, two or three attributes were varied in each column. These new items turned out to be very difficult, more so than Class B or the groups in Class A; only one of these items was answered correctly by over 20% of the subjects and some were answered correctly by only 5.1%. It appears that for most practical purposes these exceptionally difficult items will not be useful since they possibly differentiate only a few extreme cases. The question as to the nature of the abilities required for solving these items is of theoretical interest and still awaits further investigation.

3.1.2. Scalability of Subtests

An interesting feature of the relationship between subtests becomes apparent by inspection of their joint distribution matrices. These show that for some subtests it is not only the case that subtest x is more difficult than subtest y, but also that the correct solution of all or most items of subtest x, is a necessary condition for solving items of subtest y (with few exceptions). Table 3-1 shows an example (subtests CRD and DR in Form 1).

Table 3-1

Joint Frequency Distribution for Numbers of Correct Responses (DR and CRD)

CRD \ DR	0-1	2-3	4-5	6-8	Total
0-1	19	25	6	1	51
2-3	7	17	28	13	65
4-5	-	4	12	19	35
6-8	-	-	4	12	16
Total	26	46	50	45	167

As shown in the above table, there are only four cases (i.e. less than 2.4% of the total) who failed to solve four or more questions of subtest DR and who nevertheless solved four or more questions of subtest CRD. We can say that DR and CRD form a kind of primitive scale. The ability to understand items of DR is therefore in a certain sense a "prerequisite" for an understanding of items of CRD. It should be noted that although this result is a plausible one, it is by no means obvious. Compare, for instance, the joint distribution of answers for Classes B and C (in Form 4). Here it will be seen that although Class C is more difficult than Class B, these subtests do not form the kind of scale that was obtained for DR and CRD (see Table 3-2).

Table 3-2

Joint Frequency Distribution of Number of Correct Responses for Classes B and C

Class B \ Class C	0-1	2-3	4-5	6-8	Total
0-1	13	3	2	1	19
2-3	11	10	11	3	35
4-5	8	15	12	6	41
6-8	2	20	17	12	51
Total	34	48	42	22	146

Table 3-3 gives the joint frequency distribution of all pairs of subtests in Forms 1-4. Since some of the subtests appeared in more than one form, overlapping prevented us from including them in Table 3-3. These three subtests are presented in Table 3-3(a) given in the appendix. Explanation is still lacking as to what factors relate to the difference in scalability. At any rate, the notion of subtests forming scales seems

Table 3-3. Joint Frequency Distributions of Number of Correct Responses in the Subtests of Forms 1-4

Subtest	Correct Answers	RC 0-1	RC 2-3	RC 4-5	RC 6-8	DR 0-1	DR 2-3	DR 4-5	DR 6-8	D'D 0-1	D'D 2-3	D'D 4-5	D'D 6-8	CCR 0-1	CCR 2-3	CCR 4-5	CCR 6-8	CRD 0-1	CRD 2-3	CRD 4-5	CRD 6-8	D'DR 0-1	D'DR 2-3	D'DR 4-5	D'DR 6-8	Class B 0-1	Class B 2-3	Class B 4-5	Class B 6-8	Class C 0-1	Class C 2-3	Class C 4-5	Class C 6-8	Totals F.1	Totals F.2	Totals F.3	Totals F.4
RC	0-1					8	6	1	-	21	1	1	-	10	4	1	-	18	5	-	-	23	-	-	-	13	7	1	-	17	1	3	-		23	15	21
	2-3					15	13	5	1	16	16	8	1	12	15	6	1	20	19	2	-	33	6	2	-	4	8	9	4	10	9	5	1		41	34	25
	4-5					4	6	10	12	13	18	6	2	1	7	9	15	8	21	10	-	21	12	5	1	2	15	8	4	4	14	9	2		39	32	29
	6-8					-	7	24	44	2	16	27	17	-	5	13	57	2	23	26	11	12	27	14	9	-	5	23	43	3	24	25	19		62	75	71
DR	0-1	8	15	4	-					19	4	2	1	15	7	4	1	19	7	-	-	24	3	-	-	10	12	3	2	21	4	2	-	26		27	27
	2-3	6	13	6	7					14	27	5	-	8	15	5	4	25	17	4	-	26	5	1	-	7	11	8	7	8	17	6	2	46		32	33
	4-5	1	5	10	24					11	21	14	4	-	7	13	20	6	28	12	4	24	10	6	-	2	11	23	16	5	18	21	8	50		40	52
	6-8	-	1	12	44					5	11	19	10	-	2	7	48	1	13	19	12	14	22	14	7	-	1	7	26	-	9	13	12	45		57	34
D'D	0-1	21	16	13	2	19	14	11	5					17	21	6	5	29	17	2	1	45	4	3	-									49	52		
	2-3	1	16	18	16	4	27	21	11					14	17	21	11	18	26	16	3	34	14	3	-									63	51		
	4-5	1	8	6	27	2	5	14	19					1	8	10	21	4	19	13	4	10	20	8	4									40	42		
	6-8	-	1	2	17	1	-	4	10					-	1	2	12	-	3	4	8	-	7	7	6									15	20		
CCR	0-1	10	12	1	-	15	8	-	-	17	14	1	-					24	7	1	-	19	3	1	-									32		23	
	2-3	4	15	7	5	7	15	7	2	21	17	8	1					19	23	4	1	28	3	-	-									47		31	
	4-5	1	6	9	13	4	5	13	7	6	21	10	2					4	20	13	2	13	13	2	1									39		29	
	6-8	-	1	15	57	1	4	20	48	5	11	21	12					4	15	17	13	28	21	18	6									49		73	
CRD	0-1	18	20	8	2	19	25	6	1	29	18	4	-	24	19	4	4					40	6	2	-									51	48		
	2-3	5	19	21	23	7	17	28	13	17	26	19	3	7	23	20	15					37	21	8	2									65	66		
	4-5	-	2	10	26	-	4	12	19	2	16	13	4	1	4	13	17					11	17	7	3									35	38		
	6-8	-	-	-	11	-	-	4	12	1	3	4	8	-	1	2	13					1	1	4	5									16	11		
D'DR	0-1	23	33	21	12	24	26	24	14	45	34	10	-	19	28	13	28	40	37	11	1														89	88	
	2-3	-	6	12	27	3	5	10	22	4	14	20	7	3	3	13	21	6	21	17	1														45	40	
	4-5	-	2	5	14	-	1	6	14	3	3	8	7	1	-	2	18	2	8	7	4														21	21	
	6-8	-	-	1	9	-	-	-	7	-	-	4	6	-	-	1	6	-	2	3	5														10	7	
Class B	0-1	13	4	2	-	10	7	2	-																												19
	2-3	7	8	15	5	12	11	11	1																												35
	4-5	1	9	8	23	3	8	23	7																												41
	6-8	-	4	4	43	2	7	16	26																												51
Class C	0-1	17	10	4	3	21	8	5	-																	13	11	8	2								34
	2-3	1	9	14	24	4	17	18	9																	3	10	15	20								48
	4-5	3	5	9	25	2	6	21	13																	2	11	12	17								42
	6-8	-	1	2	19	-	2	8	12																	1	3	6	12								22
Total	Form 1 —					26	46	50	45	49	63	40	15	32	47	39	49	51	65	35	16																
	Form 2 ---	23	41	39	62					52	51	42	20					48	68	38	11	89	45	21	10												
	Form 3 - - -	15	34	32	75	27	32	40	57					23	31	29	73					88	40	21	7												
	Form 4 ······	21	25	29	71	27	33	52	34																	19	35	41	51	34	48	42	22				

to be worthy of further investigation and is currently being applied in another study which deals with achievement tests. Achievement may progress by more sharply defined scales than analytical ability.

3.1.3. Interrelations of Items

The coefficient of similarity described above (section 2.7), was employed in the first analysis of the analytical ability test to represent the relationships between individual items. A three-dimensional configuration with a rather good fit was obtained by the SSA-I program. The following coefficients of alienation were found for the different forms: Form 1: .145, Form 2: .116, Form 3: .107, Form 4: .162, Form 5: .075.

Using the coefficient of similarity, the first axis of the space essentially represents the degree of difficulty (see section 2.7). The second and third axes were studied in order to find out how the facet structure of items relates to their configuration, apart from difficulty. Specifically, there are two kinds of relationships which could be expected:

(1) Items which are similar to each other in n facets may be expected to be closer to each other in the two-dimensional space than items similar only in a proper subset of these facets. This is an application of what has been called the principle of contiguity (see references 4, 11) and leads to the prediction of structures such as the simplex and the circumplex (see references 4, 6).

(2) The two-dimensional space can be partitioned into contiguous regions each of which contains (with few exceptions) those and only those items which are similar in one or more facets. Roughly speaking this may be interpreted as showing that items of one region are on the whole closer to each other than to those of another region, though this does not rule out the possibility that some items belonging to one region are closer to those of another region than to the remaining items of their own region.

While the correlation coefficient pertains to individual variables or items, the concept of contiguous regions pertains to sets of variables or items, and, being based on the correlation concept, is a concept of a higher order. The empirical use of the concept of contiguous regions has been facilitated recently through the availability of new methods of computer analysis such as those described in section 2.7.

Investigation of the two-space showed that our test results did not conform to the expected relationship stated first. Furthermore, parts of the configuration of these sets were not stable from form to form (i.e., items relatively close to each other in the two-dimensional space in one of the forms might be relatively far apart in the other, and vice versa.)

What is the explanation of this finding? The rather large number of subjects filling in each form (145 or more for each one of the Forms 1-4 and 511 for Form 5) argues against the possibility of the differences being caused by sampling error alone. The retest reliability of each form individually is also evidenced by the fact that its many items can be represented in a three-dimensional space; there is a definite structure to each form separately, and structure provides a lower bound to retest reliability (see references 1, 3, 5). Further evidence in this direction comes from indications of external validity. As reported in section 3.2.4, correlations of subtests of the analytical test with scores on the mechanical comprehension test ranged from .33 to .58 (with one half of the computed correlations coefficients being .50 or more).

There remains, then, the explanation that the different configurations in alternative forms are due to the differences in the sequence of items appearing therein.

In spite of areas of instability, contiguous regions could be found - replicated over forms - which contained items having some facet elements in common.

Within Class A, arrangements of items tended to result in groups of items forming a region, but in most cases such regions were partly overlapping rather than contiguous.

Furthermore, the function employed in the construction of the items had an effect on their configuration. A region containing items in which function a or a* predominated and another region containing those in which functions b or b* predominated could be discerned in all alternative forms. The difference between a and a* and that between b and b* (i.e., the different permutation of values) did not seem to affect the correlational structure. To describe this finding more precisely let us introduce the terms a-items and b-items which are defined as follows: An a-item is one in which either all attributes or two out of three attributes vary according to function <u>a</u> or <u>a*</u>. A b-item is defined, mutatis mutandis, in the same manner. Now it can be stated that in all forms there were two contiguous regions one of which contained all a-items and one of which contained all b-items (with very few exceptions).

It will be noted that every item varying in three attributes can be classified as being either an a-item or a b-item. So can two-attribute items in which both functions are either "a-functions" (containing only <u>a</u> or <u>a*</u>) or "b-functions". Not so those two-attribute items which contain one "a-function" and one "b-function"; here neither <u>a</u> nor <u>b</u> predominates. It was found, however, that in the DR subtest the region into which the item fell depended largely on whether the attribute arranged diagonally varied according to an "a-function" or a "b-function".

In order to find out how two-attribute items of the RC subtest fitted to the overall configuration, one of the investigators independently judged which of the attributes of these items (size, direction or place) was <u>salient</u>. It was argued that, e.g., when the direction of the figure appeared intuitively to be the most salient attribute, the subject would, on first looking at the item, pay attention to direction and try to solve this item accordingly by choosing a distractor which would exhibit the value of direction seemingly most appropriate to him. It was then found that those items of RC in which the salient attribute varied by function <u>a</u> or <u>a*</u> fell almost invariably within the region of a-items whereas those in which the salient attribute varied by function <u>b</u> or <u>b*</u> fell within the region of b-items. This pattern was not found for D'D items.

In conclusion, it can be stated that the function facet exerts an influence on the structure of intercorrelations. As the foregoing will have made clear, however, only certain aspects of this structure can so far be explained by the facet design: the final structure of the correlation matrix stilleludes our analysis. Apparently the difficulty of items plays a very large role in this structure. In the present test, where items formed a rather homogeneous set, the second and third dimensions which represent the similarity between them when difficulties are held constant (see section 2.7), account for only a minor part of the variance. It seems, then, that similarity between items is so great that those differences between them which are defined by the facet structure will have a small impact compared to that of other factors. In other words, tapping a wide variety of abilities should be expected to yield a more stable structure than the more homogeneous group which constitutes the present test. The structure of the latter may be susceptible to subtle influences not picked up by our facet definition. While it is in the nature of the case that these influences cannot be identified with certainty, there appeared to be indications of perceptual factors playing a role. This notion will be discussed in section 3.1.6.

3.1.4. Interrelations of Subtests

Analysis of the intercorrelations between individual items has shown that the resulting structure can be interpreted to a certain extent in terms of the functions specified by the facet design, whereas other facets did not seem to be represented in the structure (section 3.1.3). An additional analysis was carried out to determine whether correlations between total scores obtained on various subtests are such as to render a structure which is interpretable in terms of the facets which serve to define these subtests, namely, arrangement of values and class of item (see section 2.1). Data obtained for Form 5 were used in this analysis, because this form contained the greatest variety of subtests: there were 12 subtests of two or three items each.

Product-moment correlations were computed for all pairs of total scores of these 12 subtests. There were three subtests containing Class C items which differed in the number of attributes varied in each column (see section 2.1). It was found that these had almost invariably close-to-zero correlations with all other subtests. The three Class C subtests were not even strongly related to each other, the correlation coefficients being .20, .08, and .03.

The intercorrelation matrix for Class A and Class B subtests was submitted to SSA-I. The resulting two-dimensional space had a good fit (coefficient of alienation .106) and is easily interpretable. As shown in Figure 3-3, the two-space can be partitioned in two directions. The boundary of one partition runs more or less horizontally across the figure, dividing the three Class B subtests (i.e., subtests in which shape is an attribute varied) from Class A subtests. The other (close to vertical) partition, determines the number of diagonal arrangements occurring in the items of the subtests. The region to the left of the figure includes

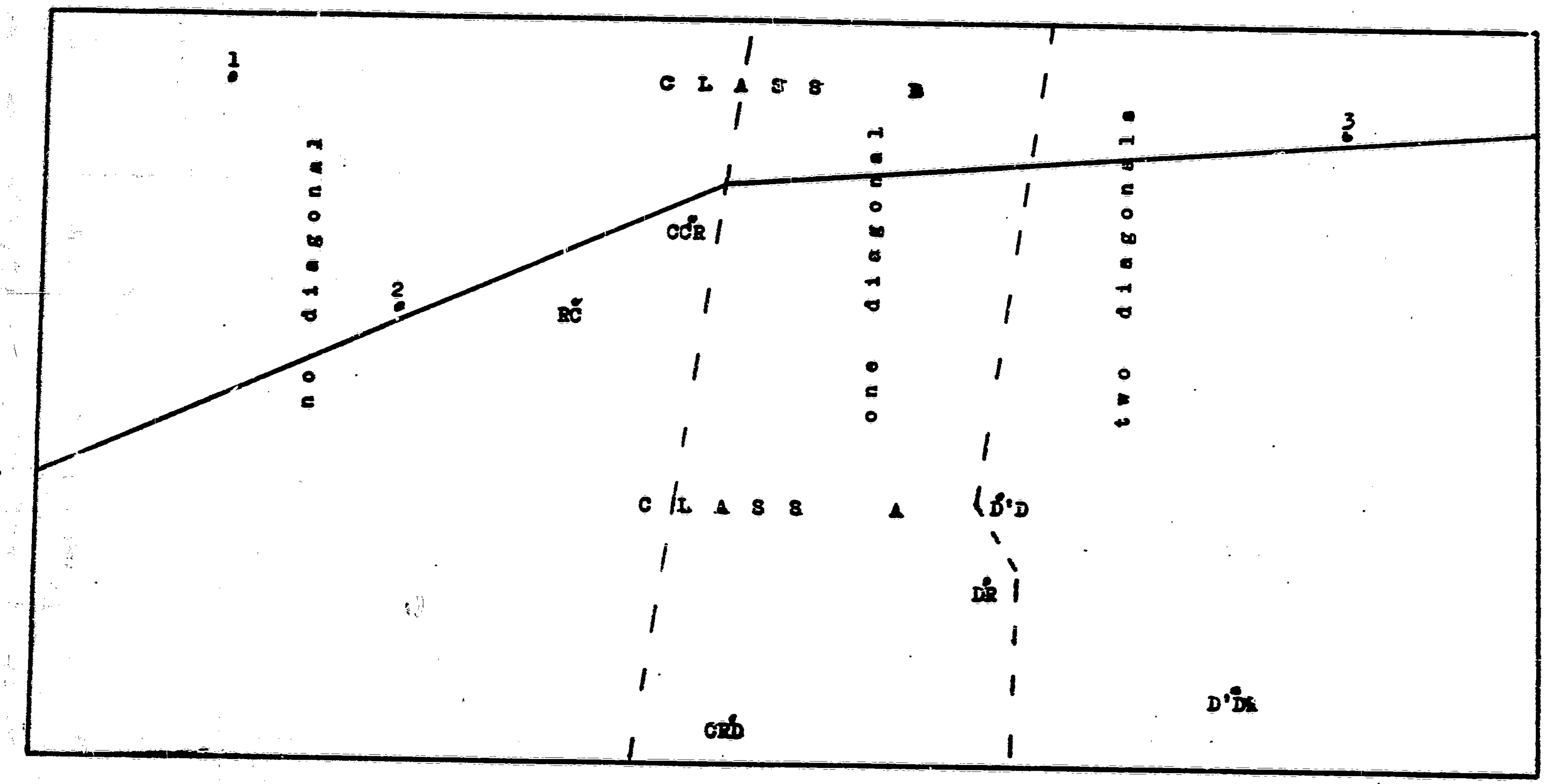


Fig. 3-3. The interrelations between nine subtests of Classes A and B

subtests 1 and 2 of Class B and two Class A subtests that have no diagonal arrangement; the middle region includes two Class A subtests with one diagonal arrangement; and the region to the right includes one Class B subtests and two Class A subtests with two diagonal arrangements.

This simplicial structure of diagonal arrangements was replicated in an analysis of the Class A subtests in Forms 1-3. Each of these forms included four of these subtests (with 8 items in each subtest). It was found that "no-diagonal" subtests correlated more with "one-diagonal" subtests than with "two-diagonal" subtests, and that, in fact, there was little overlap between these groups of correlation coefficients. These results are shown in Table 3-4.

Table 3-4

Correlation Coefficients between Subtests of Forms 1-3 having Two, One, or No Diagonally Arranged Variables*

	No diagonal		One diagonal		Two diagonals	
	RC	CCR	DR	CRD	D'D	D'DR
No diagonal						
RC	-	.79	.71	.69	.66	.49, .59
CCR	.79	-	.68, .75	.66	.57	.52
One diagonal						
DR	.71	.68, .75	-	.70	.57	.60
CRD	.69	.66	.70	-	.62, .60	.57
Two diagonals						
D'D	.66	.57	.57	.62, .60	-	.67
D'DR	.49, .59	.52	.60	.57	.67	-

* Since these coefficients are based on three different samples, some cells have more than one coefficient.

It may be concluded that both arrangement of items and their class affect the structure of interrelationships. It is noteworthy, though, that the structure is affected only by diagonal arrangements, whereas the difference between arrangement by rows and arrangement by columns does not seem to play any part. Again, this was replicated in the analysis of Forms 1-4. Furthermore, neither in Form 5 nor in Forms 1-4 was there any indication of an effect of the number of attributes varied on the correlational structure.

3.1.5. Relative Attraction of Distractors

Distractors for the analytical ability test were designed in a systematic manner as has been described above (section 2.2). This design made it possible to predict their differential attractions. In particular, distractors differing from the correct answer in one attribute were chosen more often than distractors differing in two attributes, and the latter type of distractors would be preferred over those differing from the correct answer in three attributes. In Table 3-5, the average number of subjects who chose distractors which were incorrect in one, two or three attributes is given. The average was computed by dividing the total number of times a particular type of error was made (i.e., errors in one, two or three attributes) by the total number of times a distractor with one, two or three attributes incorrect appeared in the subtest. The data were averaged separately for the different subtests of Class A - CR, DR, etc. - and were drawn from Forms 1, 2, 3, and 4.

Section 3.1.5 (2)

Table 3-5

Average Number of Subjects Choosing a Distractor Incorrect in One, Two, or Three Attributes, in the Different Subtests

Number of attributes incorrect	Subtest					
	RC	DR	D'D	CCR	CRD	D'DR
One	13.0	14.2	17.5	18.1	22.5	23.4
Two	4.6	5.2	9.2	5.7	12.0	7.7
Three	(not applicable)			1.8	3.3	6.3

The systematically manipulated similarity between the distractor and the correct answer is, then, a very powerful factor of its attraction. There were notable exceptions to this general rule, however. For some items the wrong answers would tend to involve choice of a distractor differing in two attributes from the correct answer, and these were preferred to distractors more similar to it (i.e., differing from the correct answer in one attribute only). Examination of these items suggests a reason for this. In several cases in which distractors with two incorrect attributes were preferred, these distractors were similar in all respects to the two figures of the matrix adjoining it, in either horizontal or vertical direction. Apparently, those respondents who gave these answers tended to look for a distractor in allrespects similar to the row and the column of the matrix. They made what will be termed adjacency-errors.

Section 3.1.5 (3)

An adjacency-error refers to cases in which one of the distractors was exactly alike, or very similar to,[1] the other forms in either the middle row or the middle column, or both, if the two pairs in the horizontal and vertical are both alike (e.g., distractor No.6 in example 3 of the appendix and distractor No.6 in example 5).

Table 3-6 shows that there was a tendency on the part of the subjects to make adjacency-errors; the average number of students who chose the adjacency-error is compared with the average number choosing other distractors. (Thus, if there was only one adjacency-error in a particular subtest, the number given is the absolute number of subjects choosing the adjacency-error distractors). The comparison is made for the subtests of Class A within Forms 1 and 3. All adjacency-errors in one subtest of eight items were considered as a whole. The average was computed by dividing the total number of times an adjacency-error was chosen, over the number of possible adjacency-error distractors.

---

1 A distractor is 'similar' when the two squares in either the middle row or the middle column, though having the same values for size and direction, have slightly different values for the attribute place, though neither of these values can be extreme (i.e., not in one of the corners of the small square). In addition for a distractor to be 'similar', there must be only one distractor which is exactly like one of the two squares in the middle row or column, and no distractor similar to the other square.

Section 3.1.5 (4)

Table 3-6

Comparison between Average Number of Students Choosing Adjacency-Errors and Average Number Choosing Other Distractors

Subtest in Form 1	Subtest in Form 3	Average Choosing Adjacency-Error	Average Choosing Other Distractors
	RC	45.0	8.0
	DR	78.0	9.9
DR		87.0	10.3
D'D		43.0	13.3
	CCR	42.8	6.3
CCR		68.5	5.8
CRD		43.0	14.4
	D'DR	35.0	13.9

To show how the availability of adjacency-errors interacts with similarity in influencing the subject's choice, a further breakdown of answers was made in Table 3-7 which compares the average number of subjects choosing a distractor incorrect in one, two or three attributes when: (a) there is no available adjacency-error; (b) the adjacency-error leads to choice of a distractor incorrect in one attribute; (c) the adjacency-error leads to choice of a distractor incorrect in two attributes. (It so happened that there was no adjacency-error possible in a distractor with three attributes incorrect.) The averages were computed as in Table 3-6.

Table 3-7

Average Number of Subjects Choosing Incorrect Distractor, according to Possible Adjacency-Error

Subtest in Form 1	Subtest in Form 3	Available Adjacency-Error Incorrect in 1 Attribute — Distractor actually chosen incorrect in: 1 attr.	2 attrs.	3 attrs.	No Available Adjacency-Error — Distractor actually chosen incorrect in: 1 attr.	2 attrs.	3 attrs.	Available Adjacency-Error Incorrect in 2 Attributes — Distractor actually chosen incorrect in: 1 attr.	2 attrs.	3 attrs.
	RC	17.0	3.3	-	10.2	6.7	-	-	-	-
DR		23.8	3.7	-	14.2	6.3	-	-	-	-
	DR	24.8	3.7	-	15.2	4.1	-	-	-	-
D'D		-	-	-	18.0	7.8	-	14.3	17.0	-
	CCR	20.4	2.3	1.0	13.9	4.3	2.0	7.0	13.3	3.0
CCR		28.3	3.3	1.3	9.9	4.9	2.5	8.0	24.0	1.0
CRD		-	-	-	22.5	10.0	3.7	24.0	17.3	4.0
	D'DR	-	-	-	22.6	7.9	6.2	23.2	14.0	3.3

The finding that its similarity to the correct answer affects the attraction of a distractor has important practical implications, since it now seems feasible to assign differential scores for each item by taking into account the type of distractor chosen. It appears, however, that possible adjacency-errors have to be taken into account as well.

3.1.6. Perceptual vs. Logical Processes

Items of our test were built according to a rule of construction the discovery of which was one of the tasks assigned to the subject (section 2.1). The question may be raised as to whether the subject actually tried to elicit the rule underlying the item, or whether he attempted to find the solution by means of a different process.

The effect of the facet design of items on the correlation structure should be expected to be maximal when subjects actually think in terms of these facets, i.e., when they solve items by way of the construction rule. Such a process of solving items will be termed logical process. Inspection of Raven's Progressive Matrices - which served as a starting point for work on our pictorial analytical ability test - suggests that other processes may be at work in this test. Some of Raven's distractors seem to be quite unlikely a choice because they contained figural elements not included in, or very dissimilar to, those in the matrix. Conversely, one or two distractors of an item may immediately suggest themselves as possible correct solutions because of their similarity to the figures in the matrix. When the subject is influenced by considerations such as these in the solution of items, we shall say that perceptual processes are at work. While it seems reasonable to assume that such processes are operative in the solution of Raven's matrices, there is no formal evidence to this effect. At any rate, in constructing our test, it seemed desirable to reduce the operation of perceptual processes to a minimum so as to maximise the effectiveness of the facet structure.

Section 3.1.6 (2)

It soon became apparent that by constructing the distractors of our test in a systematic fashion (see section 2.2) there were fewer possibilities for perceptual processes to take place. Thus, intuitively, none of our distractors appear to be as unlikely a choice as some of Raven's distractors.

However, the systematic construction of distractors by no means ensures that perceptual processes are ruled out completely. In the first experimental version which was employed in our pilot study, the number of figures served as an attribute and this apparently favored the operation of perceptual processes (see discussion in section 2.1). This impression was reinforced by a small informal experiment, which indicated that in this version, judged similarity of the correct answer to the figures adjacent to the missing figure was negatively related to the degree of difficulty of the item. In this version, the missing figure was in the third row third column, i.e., in the lower right hand corner. In order to reduce this effect in the revised version, it was thought advisable to omit instead the figure of row 2 column 2, i.e. the middle figure. Further, the revised version did not employ number as an attribute, nor was the function f,g,h(f,g) included (see section 2.1). It remained to be seen how successful these steps were in reducing the operation of perceptual factors.

The first indication of the operation of perceptual factors in the revised version (which is the one on which all the analyses reported in the previous sections were carried out) comes from one of the findings pertaining to subjects' preferences among the different distractors of an item. As stated (section 3.1.5), when the distractor which was most similar to the correct answer was <u>less</u> preferred, the more preferred distractor frequently was similar to the figures adjacent to the missing square. Wherever this occurs, it seems that the subject did not attempt a logical analysis of the test item, but rather used perceptual processes instead.

Further analysis of subjects' errors substantiates this claim. It was attempted to find out whether making an adjacency-error was a "trait" of a certain sample of subjects; i.e., whether the subject whose answer to one item is determined by adjacency willtend to make the same type of error in other items as well. It was found that this was indeed the case. Four of the forms were examined with regard to the tendency to make an adjacency-error. Within each form, all Class A items in which such an error was possible were compared. In Form 1, there were 7 such items; in Form 3, 9 items; in Form 4, 5 items, and in Form 5, 9 items, giving a total of 98 comparisons. In every single case the tendency to make an adjacency-error on one question in the pair correlated with the tendency to make a similar error on the other question. The fact that the correlation was always replicated is evidence for the reliability of the finding. These results are shown in Table 3-8(a-d).

As a further test of the operation of perceptual processes, the configuration of items was examined, to determine whether those items in which the same attribute is salient would tend to fall within

Table 3-8 Contingency of Adjacency-Errors

(a) Form 1.

Item No.	Type of Error	4		6		8		10		18		21		30		
		adj.	other	adj.	other	adj.	other	adj.	other	adj.	other	adj.	other	adj.	other	Total
4	adj.	-		42	24	47	19	35	31	50	16	30	36	29	37	66
	other			30	71	29	72	25	76	37	64	13	88	14	87	101
6	adj.	42	30	-		48	24	33	39	52	20	31	41	31	41	72
	other	24	71			28	67	27	68	35	60	12	83	12	83	95
8	adj.	47	29	48	28	-		42	34	51	25	31	45	30	46	76
	other	19	72	24	67			18	73	36	55	12	79	13	78	91
10	adj.	35	25	33	27	42	18	-		37	23	26	34	22	38	60
	other	31	76	39	68	34	73			50	57	17	90	21	86	107
18	adj.	50	37	52	35	51	36	37	50	-		29	58	30	57	87
	other	16	64	20	60	25	55	23	57			14	66	13	67	80
21	adj.	30	13	31	12	31	12	26	17	29	14	-		25	18	43
	other	36	88	41	83	45	79	34	90	58	66			18	106	124
30	adj.	29	14	31	12	30	13	22	21	30	13	25	18	-		43
	other	37	87	41	83	46	78	38	86	57	67	18	106			124
	Total	66	101	72	95	76	91	60	107	87	80	43	124	43	124	167

Table 3-8 Contingency of Adjacency-Errors

(b) Form 3

Item No.	Type of Error	9		11		12		14		16		18		23		31		35		Total
		adj.	other	adj.	other	adj.	other	adj.	other	adj.	other	adj.	other	adj.	other	adj.	other	adj.	other	
9	adj.	-		21	32	20	33	23	30	21	32	18	35	33	20	23	30	15	38	53
	other			17	85	14	88	24=	78	28	74	22	80	45	57	23	79	9	93	102
11	adj.	21	17	-		21	17	20	18	21	28	18	20	26	12	21	25	13	25	38
	other	32	85			13	104	27	90	28	89	22	95	52	65	17	92	11	106	117
12	adj.	20	14	21	13	-		23	11	21	13	19	15	24	10	21	13	12	22	34
	other	33	88	17	104			24	97	28	93	21	100	54	67	25	96	12	109	121
14	adj.	23	24	20	27	23	24	-		23	24	17	30	29	18	21	26	14	33	47
	other	30	78	18	90	11	97			26	82	23	85	49	59	25	83	10	98	108
16	adj.	21	28	21	28	21	28	23	24	-		25	24	33	16	19	30	15	34	49
	other	32	74	17	89	13	93	26	82			15	91	45	61	27	79	9	97	106
18	adj.	18	22	18	22	19	21	17	30	25	15	-		26	14	19	21	15	25	40
	other	35	80	20	95	15	100	23	85	24	91			52	63	27	88	9	106	115
23	adj.	33	45	26	12	24	54	29	18	33	45	26	52	-		35	43	19	59	78
	other	20	57	52	65	10	67	49	59	16	61	14	63			11	66	5	72	77
31	adj.	23	23	21	17	21	25	21	25	19	27	19	27	35	11	-		17	29	46
	other	30	79	25	92	13	96	26	83	30	79	21	88	43	66			7	102	109
35	adj.	15	9	13	11	12	12	14	10	15	9	15	9	19	5	17	7	-		24
	other	38	93	25	106	22	109	33	98	34	97	25	106	59	72	29	102			131
	Total	53	102	38	117	34	121	47	108	49	106	40	115	78	77	46	109	24	131	155

Table 3-8 **Contingency of Adjacency-Errors**

(c) **Form 4.**

Item No.	Type of Error	4 adj.	4 other	9 adj.	9 other	12 adj.	12 other	14 adj.	14 other	26 adj.	26 other	Total
4	adj.	-		10	25	11	24	14	21	14	21	35
	other			8	103	15	96	28	83	45	66	111
9	adj.	10	8	-		11	7	12	6	9	9	18
	other	25	103			15	113	30	98	50	78	128
12	adj.	11	15	11	15	-		14	12	16	10	26
	other	24	96	7	113			28	92	43	77	120
14	adj.	14	28	12	30	14	28	-		22	20	42
	other	21	83	6	98	12	92			37	67	104
26	adj.	14	45	9	50	16	43	22	37	-		59
	other	21	66	9	78	10	77	20	67			87
	Total	35	111	18	128	26	120	42	104	59	87	146

Table 3-8 Contingency of Adjacency-Errors

(d) Form 5.

Item No.	Type of Error	5 adj.	5 other	7 adj.	7 other	10 adj.	10 other	11 adj.	11 other	13 adj.	13 other	18 adj.	18 other	21 adj.	21 other	23 adj.	23 other	26 adj.	26 other	Total
5	adj.	-		9	23	12	20	15	17	14	18	16	16	5	27	9	23	7	25	32
	other			6	473	34	445	37	442	50	429	85	394	7	472	27	452	13	466	479
7	adj.	9	6	-		9	6	10	5	7	8	10	5	6	9	8	7	6	9	15
	other	23	473			37	459	42	454	57	439	91	405	6	490	28	468	14	482	496
10	adj.	12	34	9	37	-		18	28	19	27	17	29	5	41	9	37	4	42	46
	other	20	445	6	459			34	431	45	420	84	381	7	458	27	438	16	449	465
11	adj.	15	27	10	42	18	34	-		22	30	18	34	8	44	11	41	8	44	52
	other	17	442	5	454	28	431			42	417	83	376	4	455	25	434	12	447	459
13	adj.	14	50	7	57	19	45	22	42	-		23	41	5	59	9	55	8	56	64
	other	18	429	8	439	27	420	30	417			78	369	7	440	27	420	12	435	447
18	adj.	16	85	10	91	17	84	18	83	23	78	-		5	96	15	86	10	91	101
	other	16	394	5	405	29	381	34	376	41	369			7	403	21	389	10	400	410
21	adj.	5	7	6	6	5	7	8	4	5	7	5	7	-		7	5	5	7	12
	other	27	472	9	490	41	458	44	455	59	440	96	403			29	470	15	484	499
23	adj.	9	27	8	28	9	27	11	25	9	27	15	21	7	29	-		9	27	36
	other	23	452	7	468	37	438	41	434	55	420	86	389	5	470			11	464	475
26	adj.	7	13	6	14	4	16	8	12	8	12	10	10	5	15	9	11	-		20
	other	25	466	9	482	42	449	44	447	56	435	91	400	7	484	27	464			491
	Total	32	479	15	496	46	465	52	459	64	447	101	410	12	499	36	475	20	491	511

a contiguous region. It was felt that analysis of subjects' incorrect answers would disclose which attribute of the test figures, in each instance, was the most salient. If the distractor chosen by the subject was identical to the correct response in one attribute, it could reasonably be hypothesized that the subject was attentive to, and based his choice of response upon, this attribute alone. The use of one attribute and disregard of the others was taken to indicate the saliency of the particular attribute in the specific item.

This test was applied to all incorrect responses to each Class A item in Forms 1, 2, 3 and 5. It was found, through inspection of the two-space, that items with direction as their salient attribute indeed tended to fall into one region, but there were not enough data, nor were the data sufficiently consistent, to permit any conclusions on this matter.

It may be concluded, therefore, that perceptual processes are often resorted to by subjects taking the test; this is evidenced by the correlational structure as well as by the subjects' preference for the specific distractors. To the extent that such perceptual processes are not taken into account in the facet design, they might be expected to interfere with the predicted correlational structure (see section 3.1.3). Hence, the operation of such factors should be taken into consideration when constructing the test. This applies largely to tests employing figures as test material. In verbal tests, perceptual processes obviously should not be expected to operate, but it still remains to be seen whether these tests are subject to a different kind of interfering factors.

3.2. Mechanical Comprehension Test

3.2.1. Relative Difficulty of Items

The first analysis of our results was intended to determine to what extent the facet structure of items can predict their degree of difficulty. Five elements from two of the facets appear to be related to degree of difficulty: F: 3, B: 1.1, B: 1.2, B: 1.7, B: 2.6. They are respectively: mechanical activity of revolving (difficult), length of path (moderately difficult), shape of path (easy), amount of force (difficult), and angle (easy). But the strongest relationship seems to exist between facet G and degree of difficulty. The median percentage of subjects who answered correctly items which can be solved without an "intermediate consideration" was 69% whereas the median proportion of those items requiring "immediate consideration" was 43%.

The relationship between facet structure and difficulty indicates that our facets defined psychologically relevant aspects of the items.

3.2.2. Interrelations of Items

Analysis by the G-L SSA-I program resulted in a three-dimensional space with a reasonably good fit (coefficients of alienation 0.1457, 0.2637, 0.1524 for Forms 1, 2 and 3 respectively). As stated above (section 2.7), the first axis of this space relates to the degree of difficulty of the items. When the space of the remaining two axes was studied, the resultant configuration was not found to be interpretable in accordance with our facet design. Further, as was the case with the analytical ability test, the configuration was unstable in part, since it differed among the alternative forms of the test. This may possibly be explained as the result of the different sequences of items employed in the different forms.

Even when this factor was taken into account, however, and only those subgroups of items which showed a stable pattern from one form to another were examined, the item configuration of the two-space still could not be attributed to the facet profiles. It must therefore be concluded that various facets of the test -- if at all behaviorally relevant -- interact in a complex fashion, and that this does not permit prediction of correlation structure, at least until very extensive further investigations are carried out with a great number of items employing our facet design.

3.2.3. Subtests

While it is not yet possible to predict relationships between mechanical comprehension items on the basis of our facet design, a more modest analysis of the configuration yielded by the SSA-I program seems to give good results. Certain subgroups of items were found, each of which satisfied the following conditions:

(a) The group of items pertained to a certain sub-area of mechanical comprehension; (there were questions involving free movement; the results of movement on the body being moved; the influence of length of radius on the resultant movement); and

(b) the group formed a quasi-scale;

(c) which was stable over alternative forms.

Table 3-9 lists these groups. The numbers in each group are the item numbers in the form of the test for which the coefficient of reproducibility was computed. The order of items in this table is from the most difficult to the easiest. Coefficients of reproducibility are given for each group.

Table 3-9

Subtests of Mechanical Comprehension and their Coefficients of Reproducibility

Content	Form	Items No.	Coefficient of Reproducibility
Free movement	1	21, 27, 44, 18	.94
Free movement	1	40, 19, 17	.88
Free movement	2	23, 54, 48, 19	.94
Free movement	2	38, 34, 39	.91
Free movement	2	26, 17, 24	.89
Movement and its effects	1	42, 38, 15	.94
Movement and its effects	2	21, 29, 52	.94
Length of radius	2	10, 56, 11, 14	.89
Length of radius	2	51, 42, 27	.91

The immediate practical advantage of this grouping lies in that a new version of the test (which yields several sub-scores), can now be constructed. These sub-scores are meaningful in that score profiles can be reconstructed from them. Further investigation with the sub-scores can be expected to reveal more important theoretical relationships with outside variables than do total scores on mechanical comprehension tests. In fact, such a revised version of our mechanical comprehension test is now being prepared and will shortly be in use at vocational guidance centers throughout the country.

3.2.4. Relationship between Mechanical Comprehension and Analytical Ability

Correlations were computed between the total score of the mechanical comprehension test and subtests of the analytical test, for Forms 1, 2 and 3 of the latter. (Subjects taking these forms took Forms 1, 2 and 3 respectively of the mechanical test. The subjects taking Form 4 of the analytical test were equally divided between Forms 1 and 3 of the mechanical test - see section 2.6 - and therefore, the samples were too small for the computation of reliable correlation coefficients). Table 3-10 shows these correlations.

Table 3-10

Product-Moment Correlations between Total Scores of the Mechanical Comprehension Test and Subtests of Analytical Ability (Class A Items)

	Analytical Ability Subtests					
	RC	DR	D'D	CCR	CRD	D'DR
Form 1	-	.38	.38	.47	.33	-
Form 2	.58	-	.50	-	.55	.46
Form 3	.55	.58	-	.54	-	.41

It will be seen from the table that the highest correlations were those between mechanical comprehension and (with one exception in Form 1) subtests RC and DR, (i.e., the easy subtests, which include two-attribute items with at most one diagonally arranged function) while the lowest correlations were those with subtest D'DR (i.e., the most difficult subtest; see section 3.1.1).

It should be pointed out that the size of these correlation coefficients bear evidence to the external validity of both tests concerned, and therefore also to their reliability.

More detailed analyses of the relationship between mechanical comprehension and analytical ability are now being carried out on the basis of subtests of the mechanical comprehension test.

To obtain additional information on the relationship between mechanical comprehension and different kinds of analytical ability, a reanalysis was undertaken of the results of a study by Kraak (see reference 13). Kraak published the correlation matrix for ten subtests of mechanical comprehension, a verbal analogy test and a test of numerical progressions.

An SSA-I space shows that the mechanical tests closest to the verbal analysis test, are two subtests which show diagrams or pictures of machines and in which the subject is called upon to explain their operation. However, another subtest showing such a diagram, is far removed from the verbal analogy test. (This subtest shows a three-armed T-formed lever, two arms of which are held by springs, and the subject is asked how a weight on one of the arms would influence the lever.) Four tests of spatial perception form contiguous regions of the two-dimensional space resulting from the SSA-I. These are closest to the numerical progression test.

4. CONCLUSIONS

In the present study a first attempt has been made to implement the facet theory approach in test construction and data analysis. The experience gained here makes it possible to evaluate the fruitfulness of this approach, and the promise held out by it for further progress in mental testing.

The first point to be made here is in regard to the possibilities of systematic item construction through facet design. Two kinds of tests were devised in this study: an analytical ability test and a mechanical comprehension test. While each employed a facet definition, it appeared that the two differed in the extent to which use could be made of their respective definitions. In the analytical test, the facet design could be used to specify the items down to every detail. This was not so in the mechanical comprehension test, where the designs could be used only to determine the kind of relationships between attributes (such as force, speed, and volume) which the items pertained to. It was not possible to build items for each one of the profiles which could be generated from the facet definition of the mechanical test, whereas no such limitation was found in regard to the analytical tests, which called for the finding of analogies and employed arbitrarily chosen geometric figures.

Further, the distractors of the analytical tests could be systematically designed by means of facets (. section 2.2). Analysis of the results showed that we were successful in constructing distractors which differed systematically in "attraction", opening up new possibilities for increasing the discriminatory power even of short tests (section 3.1.5). This advantage was not shared by the mechanical test, where it proved difficult to design distractors by facets.

Considering the nature of the tests concerned, all this is just what one would expect. It only goes to show that tests may differ widely in the ease with which facet design may be applicable to their construction. On the other hand, it seems safe to say that whatever the nature of the test, facet design can be employed in item construction at least to some extent; the experience with the particularly unwieldly mechanical ability questions seems to be evidence for this.

That the facets employed in the construction of the present test were psychologically relevant is shown, first of all, by the fact that they could predict the relative degree of difficulty of test items. This is true for the analytical ability test and, to a certain extent, for the mechanical comprehension test.

Another question - which also concerns the psychological relevance of our facets - is that of the obtained relationship between the facet structure of the test and the statistical structure of its results. Here again, large differences were found between the two kinds of tests. For reasons which have been discussed, the data from the mechanical test did not yield an intercorrelation matrix which was interpretable in terms of the facets employed (section 3.2.2). By contrast, several of the facets of the analytical ability test were reflected through items forming regions in the space described by the computer program using smallest space analysis (sections 3.1.3, 3.1.4). Even here though, the success achieved was only partial, and many details of the statistical structure did not prove amenable to explanation through the facet profiles of the items.

In considering possible reasons for this, a factor which acted as "noise" was hypothesized: the operation of what has been termed "perceptual processes" (section 3.1.6). Supplementary analysis of the data lent support to this explanation. This result might prove to be important for future work on the construction of analytical ability test of the kind developed in this project, where, it is suggested, this factor should be taken into account.

In summing up the lessons of this project, we might say that this first attempt of carrying out all steps of test development systematically by means of facet design seems to show the fruitfulness of the approach taken. The degree of success achieved was uneven, differing with the kind of test and the nature of analysis conducted. At times it was possible to reveal the reasons for the difficulties encountered, and thus valuable experience has been gained for future work on test development based on facet theory.

The findings in regard to regions representing classes, arrangements and functions, and those regarding the contingency of adjacency-errors in the analytical test as well as in the subtests of the mechanical tests should ultimately prove of importance for purposes of prediction. In future research these findings could be the basis for developing typologies of respondents. Once appropriate criteria have been established, it will be possible to investigate the relationships between type of respondent and measures of academic achievement and vocational success.

References

1. Guttman, Louis. A Basis for analyzing test-retest reliability. Psychometrika, 1945, 10, pp.255-282.

2. Guttman, Louis. "Relation of Scalogram Analysis to Other Techniques" In: S.A. Stouffer et al. Measurement and Prediction, Princeton, New Jersey: Princeton University Press, 1950, pp.172-212.

3. Guttman, Louis. Reliability Formulas that do not assume Experimental Independence. Psychometrika, 1953, 18, pp.225-239.

4. Guttman, Louis. "A new approach to factor analysis: the radex" In: P.F. Lazarsfeld (Ed.), Mathematical Thinking in the Social Sciences, The Free Press, 1954, pp.258-348.

5. Guttman, Louis. Reliability Formulas for Noncompleted or Speeded Tests. Psychometrika, 1955, 20, pp.113-124.

6. Guttman, Louis. Empirical Verification of the Radex Structure of Mental Abilities and Personality Traits. Educational and Psychological Measurement, 1957, 17, pp.391-407.

7. Guttman, Louis. What Lies Ahead for Factor Analysis? Educational and Psychological Measurement, 1958, 18, pp.497-515.

8. Guttman, Louis. "The Structure of Interrelations among Intelligence Tests" In: Proceedings of the 1964 Invitational Conference on Testing Problems, Princeton, New Jersey: Educational Testing Service, 1965.

9. Guttman, Louis. A Faceted Definition of Intelligence. Scripta Hierosolymitana: Studies in Psychology, Jerusalem: The Hebrew University, 1965.

10. Guttman, Louis. A General Nonmetric Technique for Finding the Smallest Euclidean Space for a Configuration of Points. Psychometrika, (in press).

11. Guttman, Louis. "Order Analysis of Correlation Matrices" In: R.B. Cattell (Ed.), Handbook of Multivariate Experimental Psychology, Rand McNally, (in press).

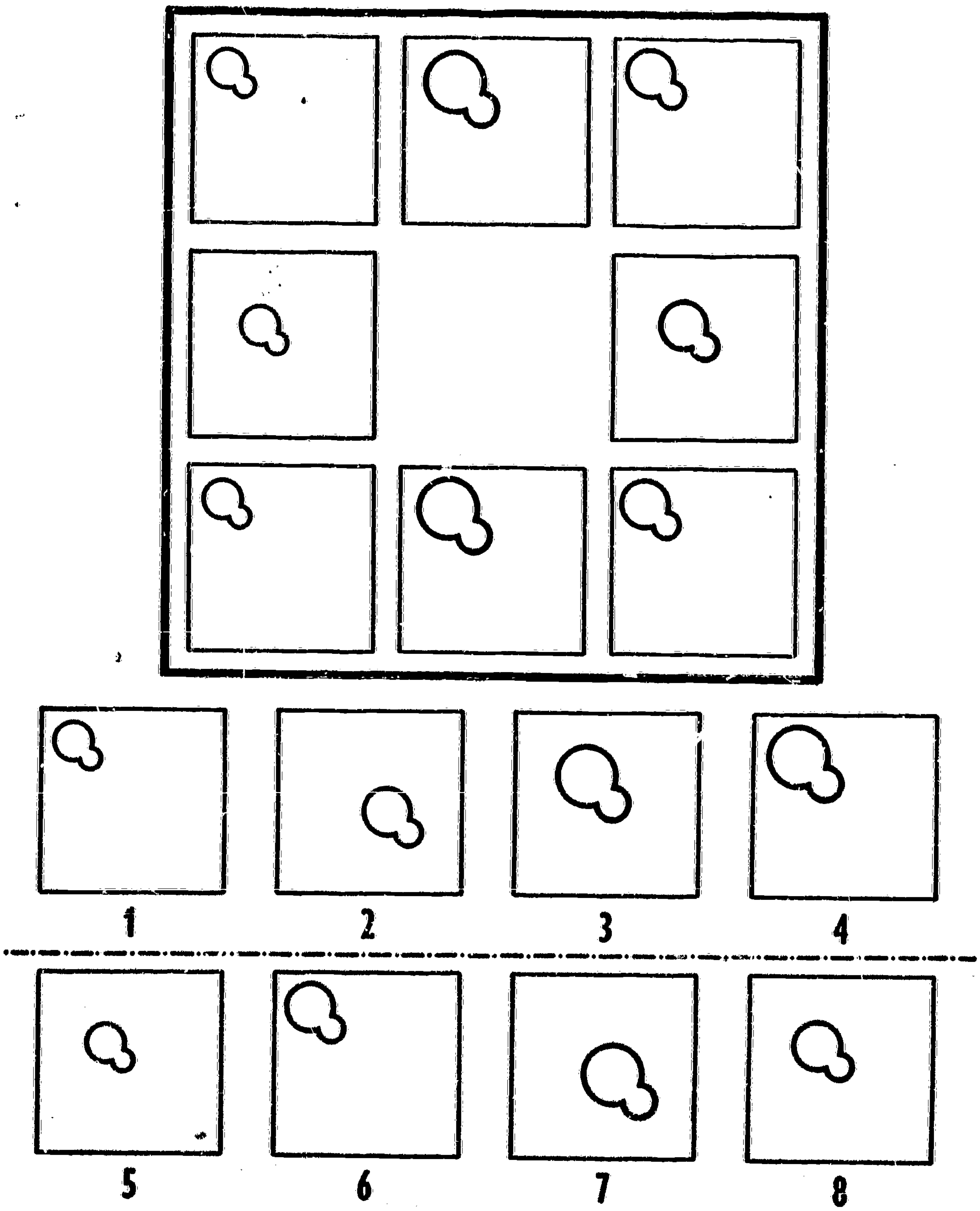
12. Guttman, Ruth, and Guttman, Louis. Cross-cultural Stability of an Intercorrelation Pattern of Abilities: A Possible Test for a Biological Basis. Human Biology, 1963, 35, pp.53-60.

13. Krauk, B. Räumliches Vorstellen als Voraussetzung technichen Denkens. Zeitschrift für experimentelle und angewandte Psychologie, 1961, 8, pp.78-107.

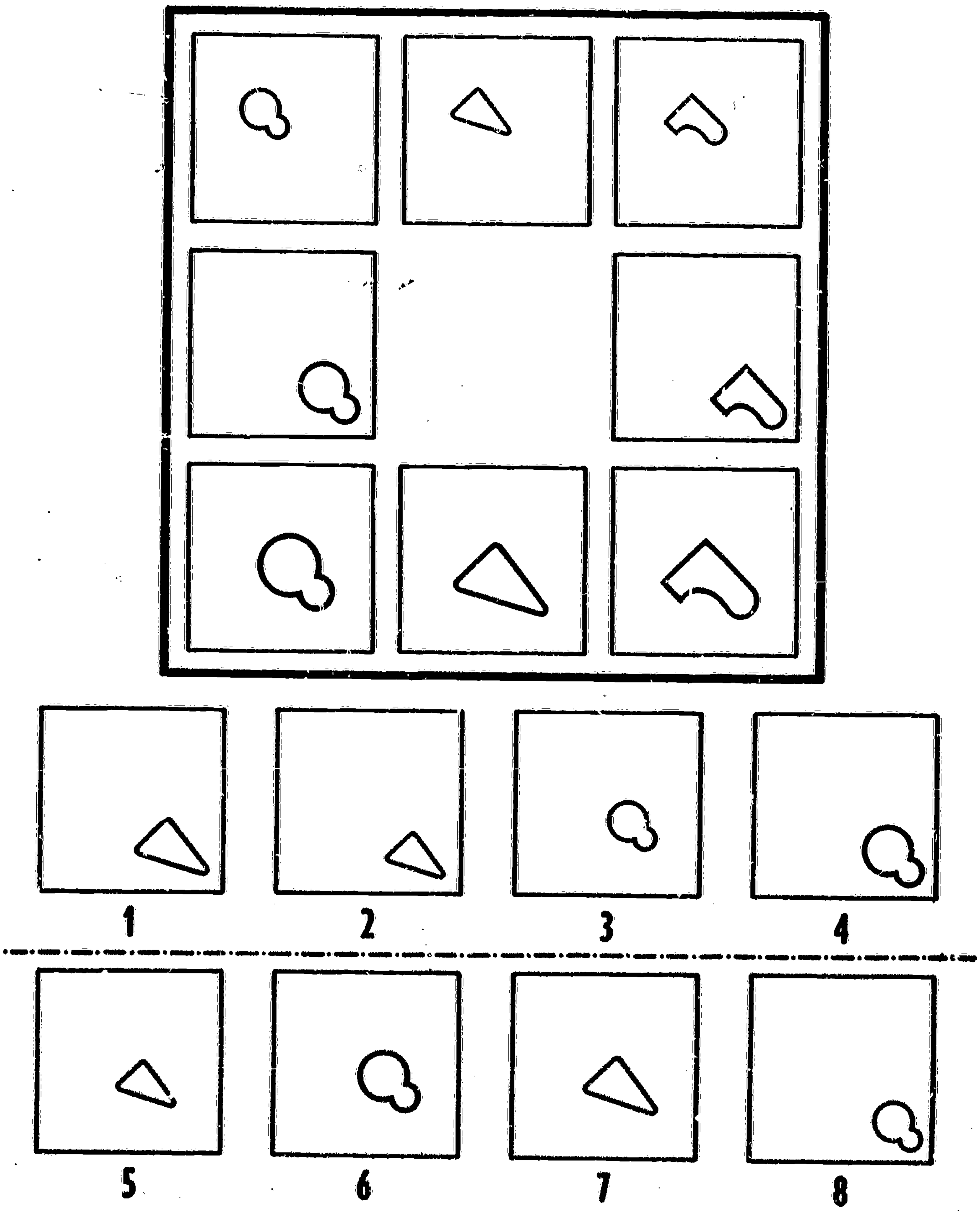
14. Lingoes, James C. New computer developments in pattern analysis and nonmetric techniques. IBM Journal, (in press).

15. Lingoes, James C. An IBM-7090 program for Guttman-Lingoes smallest space analysis-I. Behavioral Science, 1965, 10, pp.183-184.

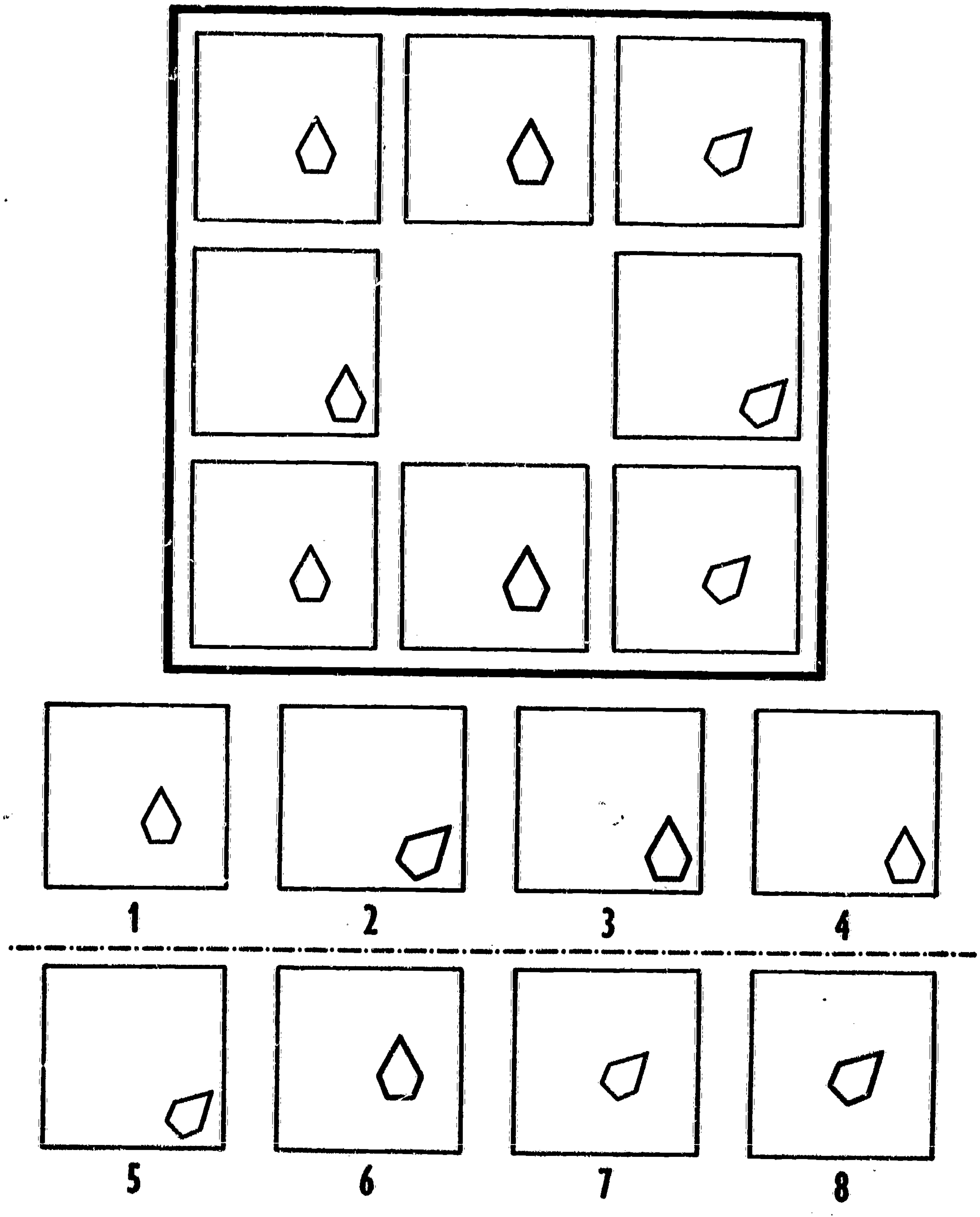
APPENDICES



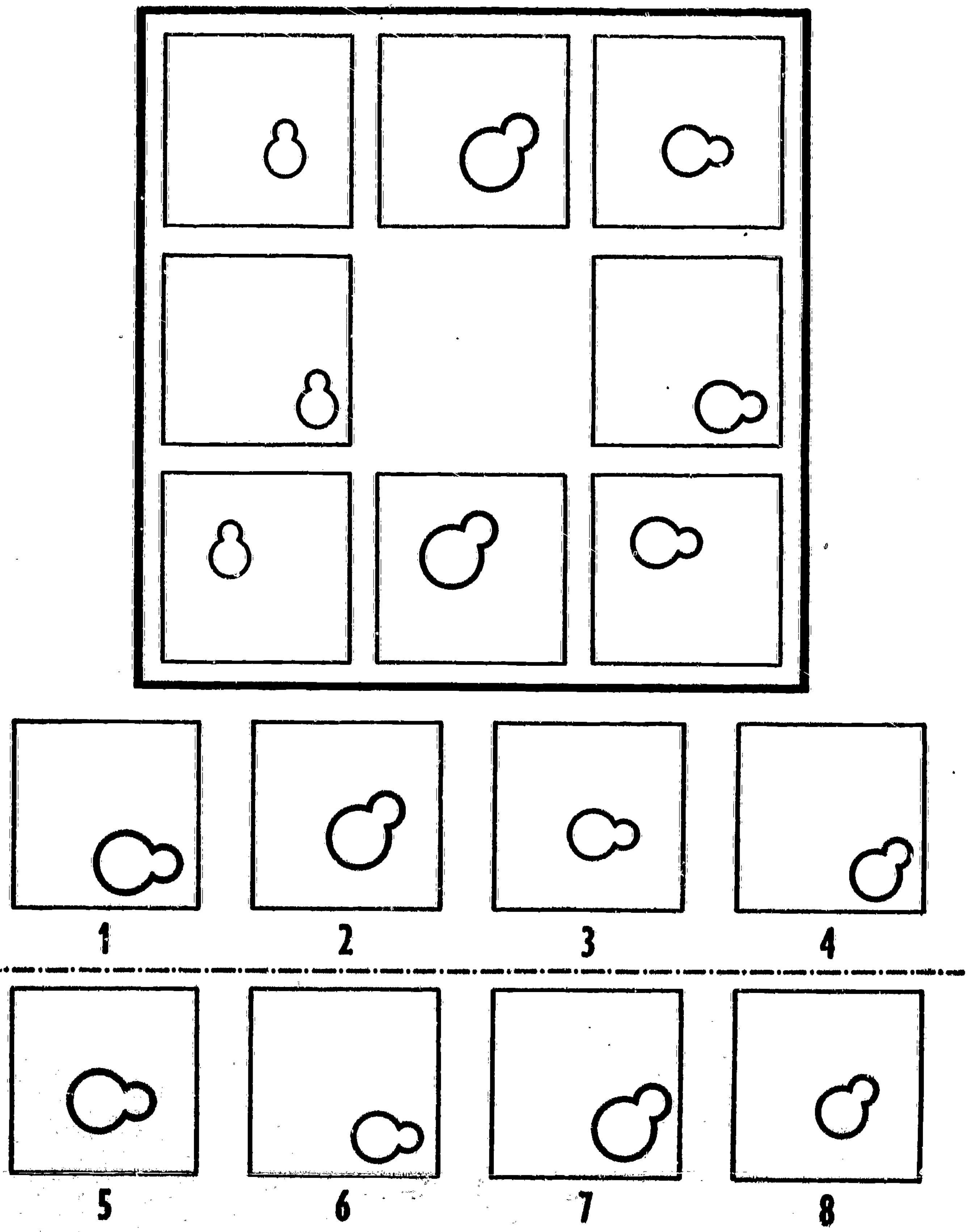
Example 1



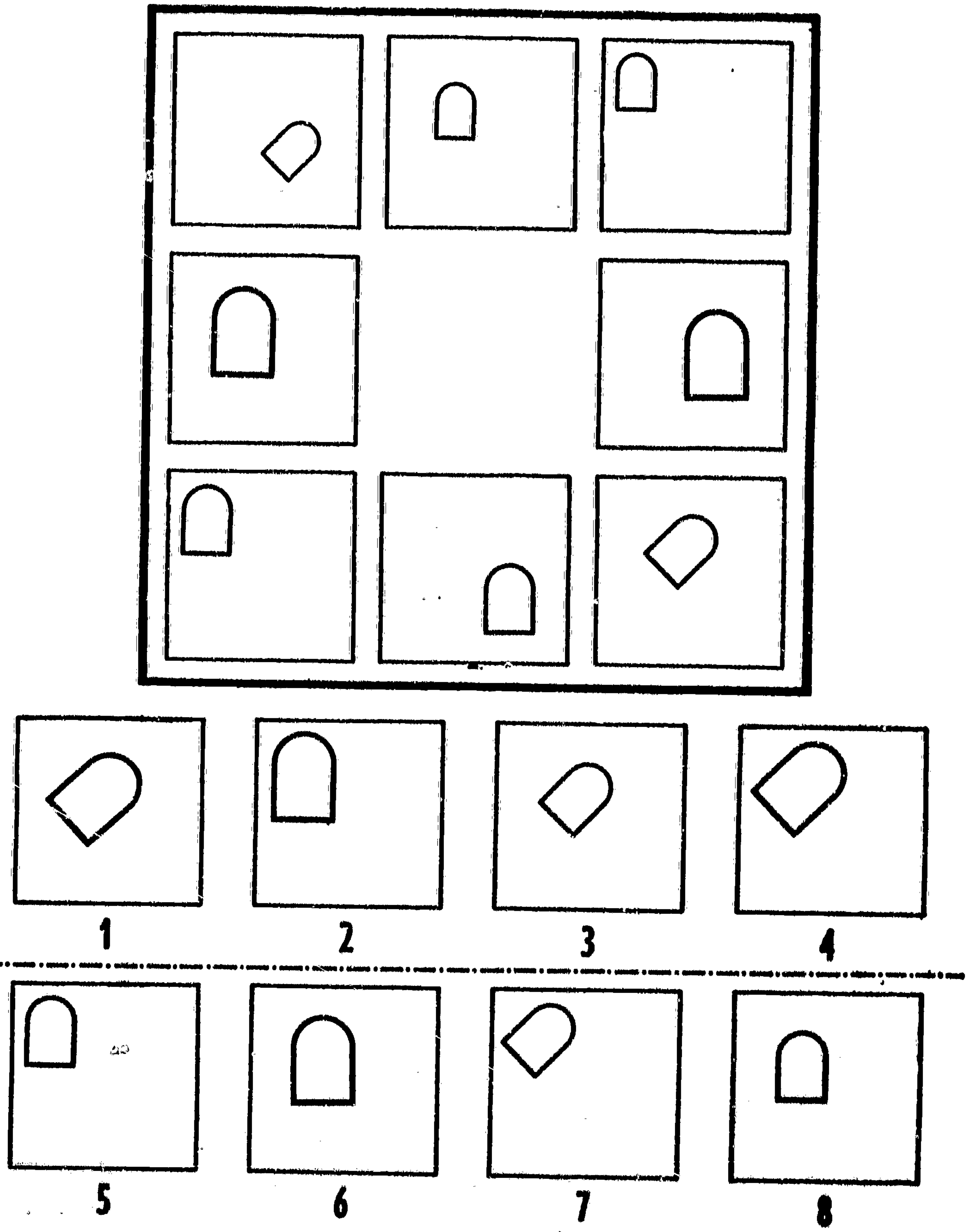
Example 2



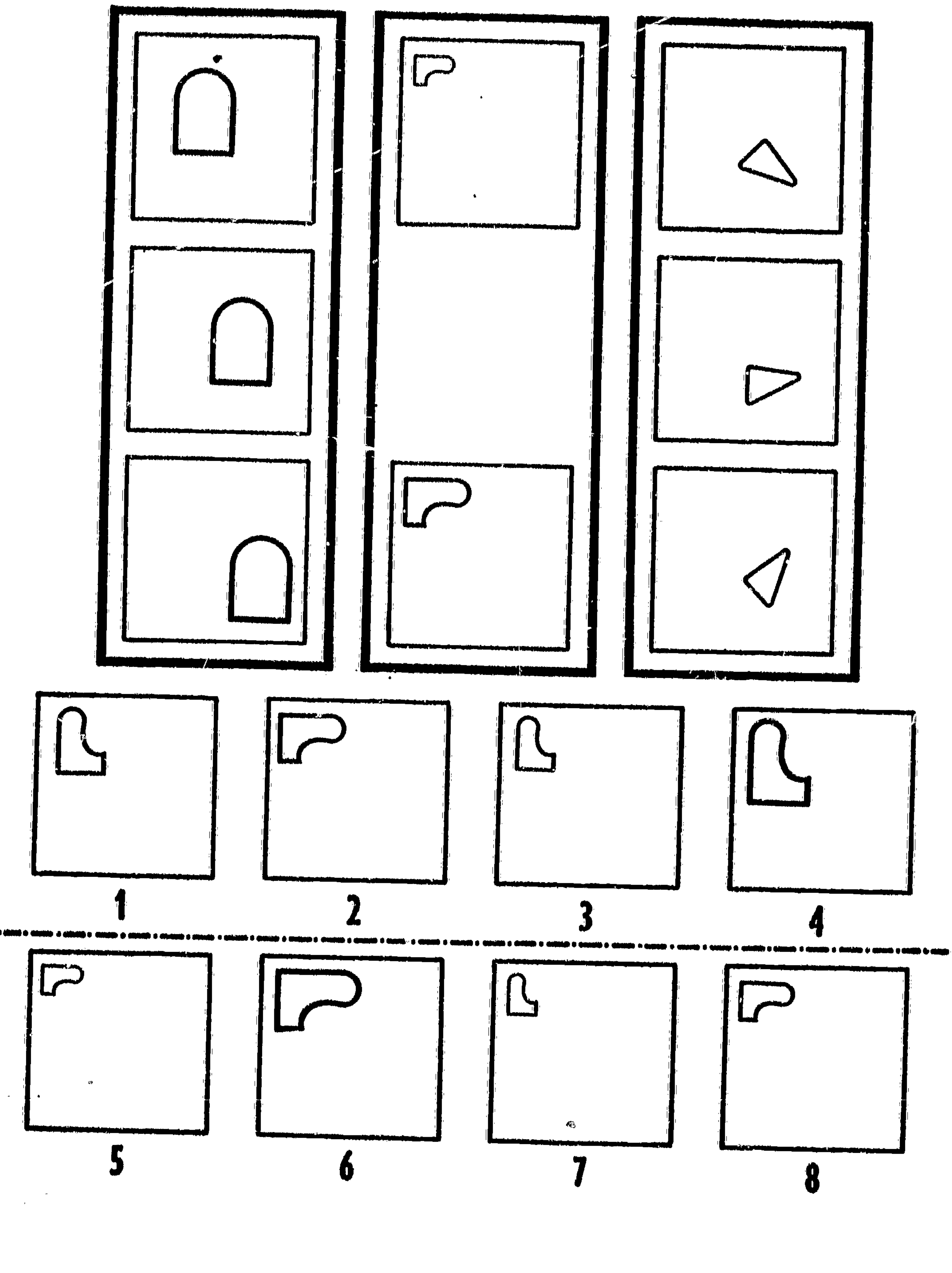
Example 3



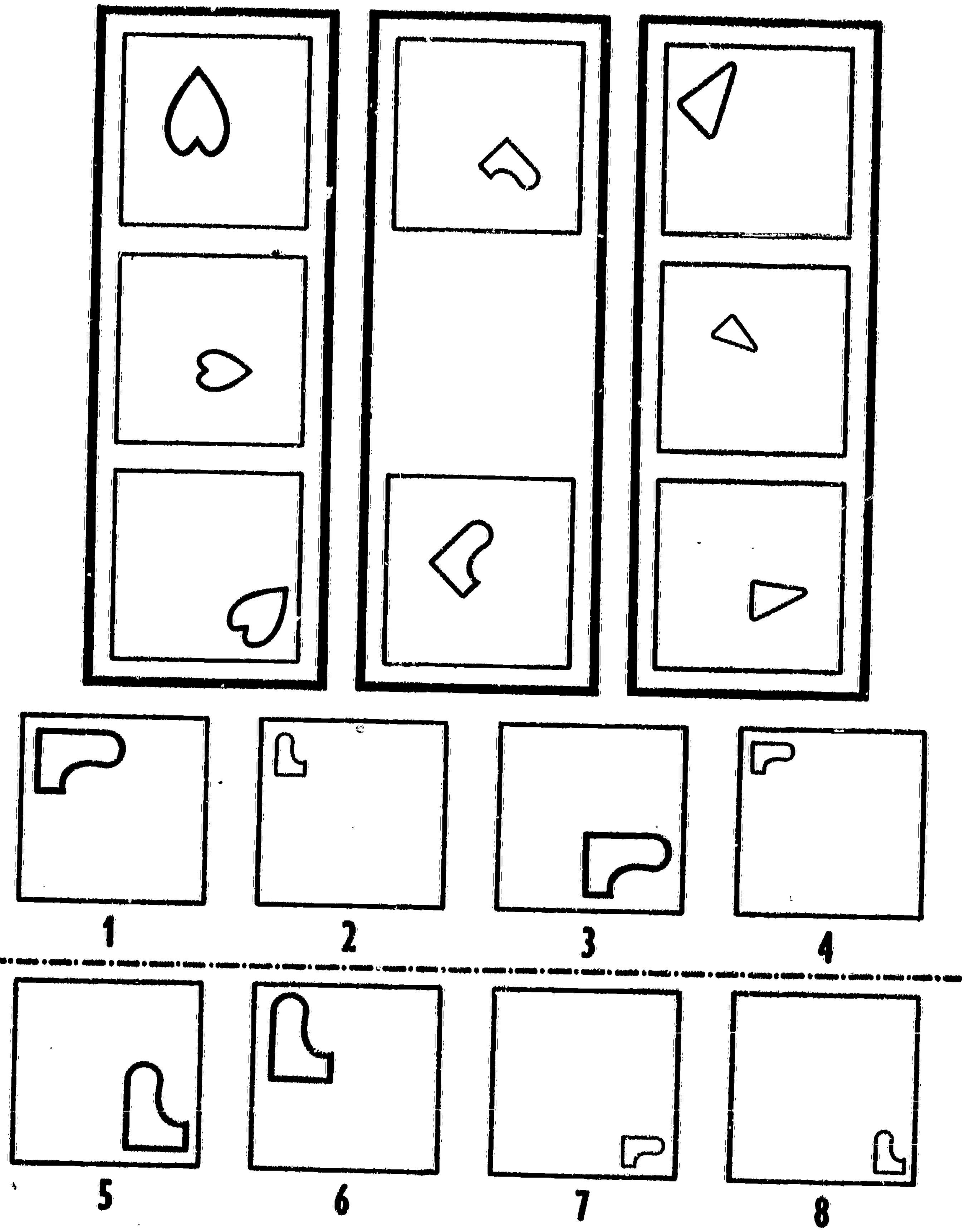
Example 4



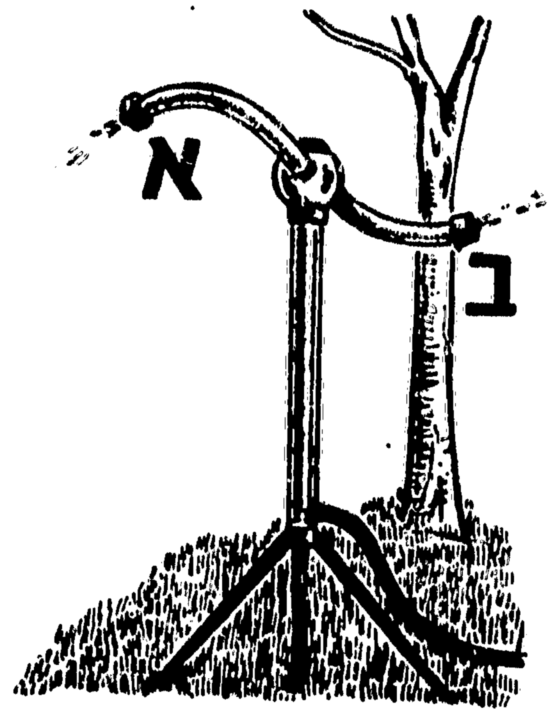
Example 5

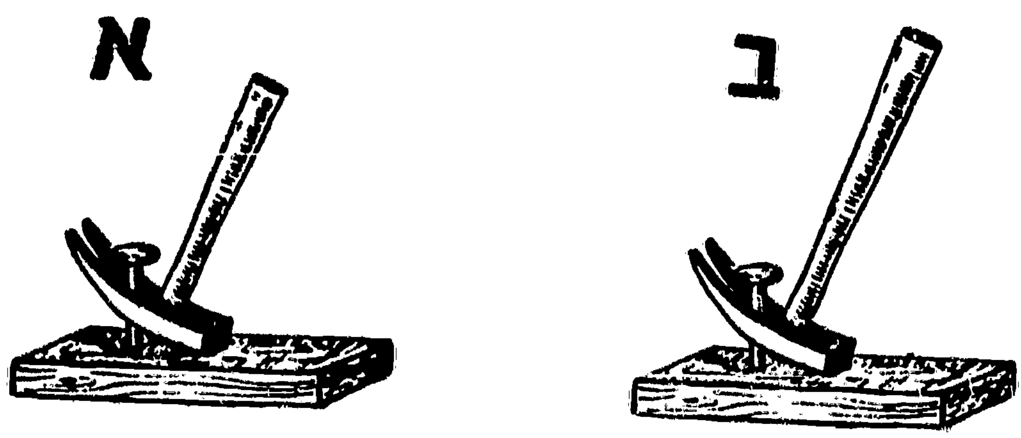


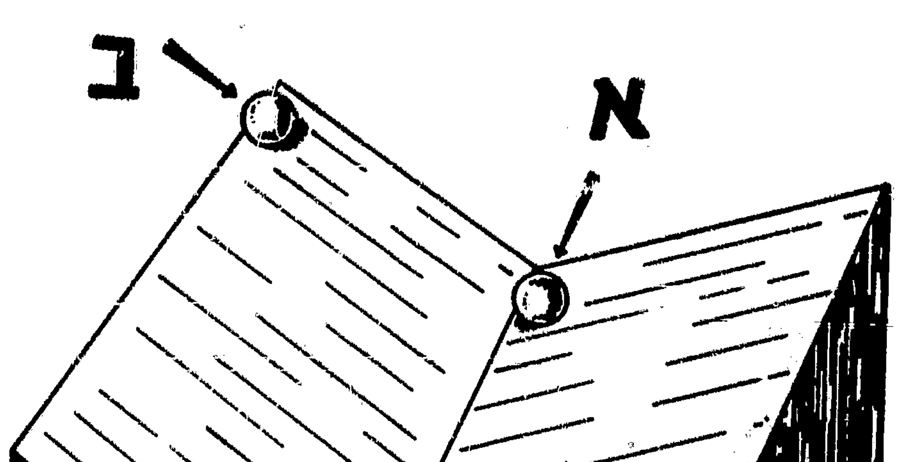
Example 6



Example 7


א
ב


א
ב


ב
א

From which outlet will the water reach the tree first,

1. From outlet A (א)
2. From outlet B (ב)
3. It cannot be established

With which hammer will it be easier to pull out the nail?

1. With hammer A (א)
2. With hammer B (ב)
3. It makes no difference

Which ball will roll in an arc?

1. Ball A (א)
2. Ball B (ב)
3. Neither of them

Example of Mechanical Comprehension Test Item

Table 3-3(a)

Joint Frequency Distribution of Number of Correct Responses

(Subtests which were not included in Table 3-3)

Form 1

DR \ CCR	0-1	2-3	4-5	6-8	Total
0-1	12	12	1	1	26
2-3	14	18	10	4	46
4-5	5	17	14	14	50
6-8	1	-	14	30	45
Total	32	47	39	49	167

Form 2

D'D \ CRD	0-1	2-3	4-5	6-8	Total
0-1	32	16	4	-	52
2-3	11	29	10	1	51
4-5	5	16	16	5	42
6-8	-	7	8	5	20
Total	48	68	38	11	165

Table 3-3(a)-(Cont.)

Form 3

RC \ DD'R	0-1	2-3	4-5	6-8	Total
0-1	12	3	-	-	15
2-3	29	4	1	-	34
4-5	20	7	4	1	32
6-8	27	26	16	6	75
Total	88	40	21	7	156

Form 4

DR \ RC	0-1	2-3	4-5	6-8	Total
0-1	13	7	4	3	27
2-3	5	11	9	8	33
4-5	3	7	12	30	52
6-8	-	-	4	30	34
Total	21	25	29	71	146

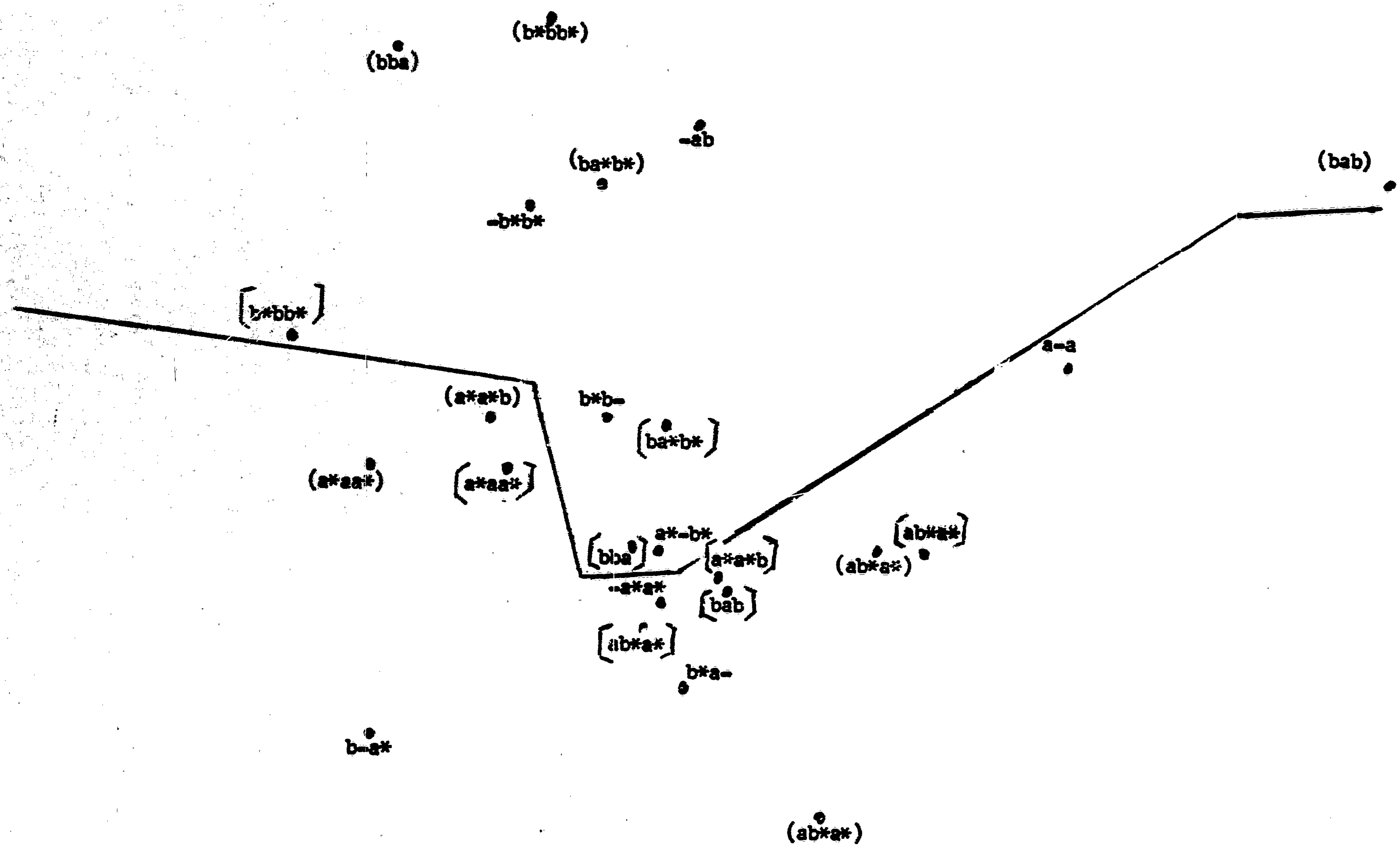


Fig. A-1. Space-diagram showing interrelations between items of subtests CRD (in parentheses), D'DR (in square brackets), and RC, of Form 2. Letters refer to functions of place, direction, and size, in this order. (A dash indicates that the corresponding attribute is held constant). Items fall into two approximately contiguous regions as explained in section 3.1.3