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

ED 014 773   72   CG 000 976

THE ANALYSIS OF DIAGNOSTIC EFFECTIVENESS OF A FACET DESIGN BATTERY OF ACHIEVEMENT AND ANALYTICAL ABILITY TEST.

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

REPORT NUMBER BR-5-1409   PUB DATE 67
GRANT OEG-5-21-006
EDRS PRICE MF-$0.50 HC-$4.08   100P.

DESCRIPTORS- *TEST CONSTRUCTION, ACHIEVEMENT TESTS, APTITUDE TESTS, TESTS, DATA ANALYSIS, *STRUCTURAL ANALYSIS, GROUP INTELLIGENCE TESTS, RESEARCH PROJECTS, GRADE 8, *DIAGNOSTIC TESTS, GUTTMAN LINGOES MULTIDIMENSIONAL SCALOG., GUTTMAN LINGOES SMALLEST SPACE ANALYSIS

THE INTERNAL STRUCTURE OF A BATTERY OF ACHIEVEMENT AND INTELLIGENCE TESTS WAS ANALYZED TO ENHANCE THE DIAGNOSTIC VALUE OF THE BATTERY. CONSTRUCTION OF THE ACHIEVEMENT AND INTELLIGENCE TESTS WAS GUIDED BY A FACET DESIGN. THE RESEARCH HYPOTHESES WERE THAT--(1) STAGES OF DEVELOPMENT AND ACHIEVEMENT IN THE VARIOUS AREAS TESTED IN THE BATTERY EXIST, AND (2) CERTAIN KINDS OF SYSTEMATIC DIFFICULTIES CAN BE DIAGNOSED BY FACET DESIGN AND BY ANALYSIS OF TEST DISTRACTORS. DATA ANALYSIS WAS DONE USING THE GUTTMAN-LINGOES SCALOGRAM ANALYSIS I (MSA-I) AND THE GUTTMAN-LINGOES SMALLEST SPACE ANALYSIS I AND II (SSA-I AND SSA-II). THE STRUCTURE OF THE INTERRELATIONSHIPS HOLDING BETWEEN A BATTERY OF TESTS AND BETWEEN ITEMS OF A SUBTEST WERE INVESTIGATED USING SSA-I. RESULTS WERE CONSISTENT WITH EARLIER FINDINGS ON THE RELATIONSHIPS BETWEEN DIFFERENT SUBTESTS OF ANALYTICAL ABILITY AND APTITUDE ACHIEVEMENT. THE MSA-I ANALYSIS OF SUBTESTS AND OF INDIVIDUAL ITEMS REVEALED THE EXISTENCE OF STAGES OF ACHIEVEMENT FOR THE ARITHMETIC TEST. A SSA-II ANALYSIS OF THE ARITHMETIC TEST SHOWED THAT SUBJECTS TENDED TO MAKE SIMILAR TYPES OF ERRORS ON DIFFERENT ITEMS. (PS)

DIAGNOSTIC EFFECTIVENESS OF FACET DESIGNED TESTS

Cooperative Research Project No. OE-5-21-006

LOUIS GUTTMAN AND I. M. SCHLESINGER

THE ISRAEL INSTITUTE OF APPLIED SOCIAL RESEARCH

JERUSALEM / ISRAEL

BR: 5-1409
PA: ~~#~~ 72

THE ANALYSIS OF DIAGNOSTIC EFFECTIVENESS OF A FACET DESIGN BATTERY OF ACHIEVEMENT AND ANALYTICAL ABILITY TEST

Cooperative Research Project No. OE-5-21-006

Louis Guttman, Ph.D., Principal Investigator

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

The Israel Institute of Applied Social Research
Jerusalem, Israel

1967

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

to

The Henrietta Szold Institute
National Institute for Research in the Behavioral Sciences

ACKNOWLEDGEMENTS

We take pleasure in expressing our gratitude to the persons who have helped us in this study. Dr. Sara Smilansky and Mr. S. Zirlin of the Szold Institute were responsible for the construction of the test battery on which this present study is based. Of the present and previous staff of the Institute, Marlyn Grossman and Zehavit Weiser gave valuable assistance with the data analyses; U. Haim, H. Bardin, S. Seger and Naomi Mansoor were responsible for mechanical processing and computer analysis of the data; V. Klaff, and Haya Gratch aided in the stylistic editing of this final report. This study owes very much to the devoted efforts of the above.

The smallest space analyses in this report were made possible through a grant from the National Science Foundation to the University of Michigan, Grant No. GS-929.

CONTENTS

LIST OF FIGURES

Note: Description of figures usually appear on page preceding figure.

LIST OF TABLES

1. INTRODUCTION

1.1. General Objectives

The general purpose of the present study was to carry out a deeper analysis than usual of the internal structure which holds in a battery of achievement and intelligence tests, with a view to enhancing their diagnostic value.

The basis for this study was a battery of intelligence and achievement tests constructed at the Szold Institute, with the principal investigator of this project as advisor. This battery was administered in 1962 by the Israel Ministry of Education and Culture to a country-wide sample of 8,000 eighth graders. Various analyses of the test results have been carried out by the Israel Institute of Applied Social Research.

The test battery consisted of the following tests:

A. Achievement Tests

1. Hebrew Language
   - Reading comprehension (four subtests)
   - Vocabulary
2. Arithmetic: Percentages (six subtests)
3. Geometry (three subtests)

B. Analytic Ability Tests

Analogies, progressions, differences, similarities, etc. (16 subtests)

More detailed information about these tests will be given in the appropriate sections of this report. Examples of the various test items are given in the appendix.

The construction of these tests was guided by a facet design. The notion of facets and their application to intelligence tests has been discussed at length elsewhere.[2,5] In brief, two sets of elements, A and B are called facets, and their Cartesian space is the set of all pairs of elements <u>ab</u>, where <u>a</u> is an element of A and <u>b</u> is an element of B. A Cartesian space may consist of any number of facets, or sets of elements; with <u>n</u> facets, any one point in the Cartesian space has <u>n</u> component elements.

The facet approach in test construction makes it possible to arrive at items by a systematic <u>a priori</u> design, instead of by the usual process of designing test items which is largely based on intuition and on subsequently weeding out inappropriate items by means of statistical analysis of test results.

The data analyses were guided by two research hypotheses of a very general nature, but which could be tested directly with the empirical materials at hand:

A. There exist stages of development and of achievement in the various areas tested in our battery.

B. There are certain kinds of systematic difficulties which can be diagnosed by facet design and by analysis of test distractors.

These, then, were the working hypotheses formulated at the outset of the study. They are, of course, not the kinds of hypotheses which in themselves could lead to detailed predictions. The general predictions involved did guide the analysis of the data, leading to an understanding of the structure obtained.

1.2. Description of Types of Analysis

Customary methods of analysis were not suitable for the investigation of such hypotheses. Instead, new kinds of multivariate analyses were employed.

When the data were gathered in 1962, they were transferred to Powers-Samas punch cards at the Israel Institute of Applied Social Research. Work on this project started with the use of the mechanical computing equipment of the Institute. At the time the analyses were half completed, an important development in methods of multivariate data analysis took place. Electric computer programs for Smallest Space Analysis and for Multidimensional Scalogram Analysis were developed. These methods are briefly described in Sections 2.1, 4.1 and 5.1. They are so much more powerful than conventional methods of analysis that it was decided to make a fresh start and resubmit our data to the new program, which had meanwhile become operational on the Hebrew University 7040 computer. This necessitated transferring part of our data to IBM cards and entailed a considerable loss of time. But the cost involved is held to be more than justified in view of the increased scope and power of our analysis.

Test results of a sub-sample of 600 subjects were therefore transferred to IBM cards and analyzed by the new computer techniques. This sample was considered satisfactory for the type of analyses we had in mind.

Two types of analyses were employed for our data:

A. Correlational analyses involving not only pairwise comparisons between structures but a study of the structure of inter-relationships holding in a whole matrix of data. These were carried out by the newly developed G-L SSA-I and G-L SSA-II programs (see sections 2.1 and 5.1).

B. Scalogram Analyses were used both for one dimensional and for multidimensional structures. These have been carried out by the G-L MSA-I (see Section 4.1). This analysis is especially designed for the investigation of typologies since it reveals what profiles occur in the data. Occasionally, the term profile analysis will be used in this report.

In the first stages of this project it became apparent that the tests of our battery differ largely in the extent to which their construction conforms to a facet design, and hence in the extent to which it would be fruitful to submit them to the various kinds of analysis. In particular it was found that the Hebrew achievement tests would render little useful information through a fine grain analysis of their internal structure. Therefore, this report concentrates mainly on correlational analysis and profile analysis of the arithmetic achievement subtests and some of the analytical ability subtests. Intensive distractor analyses could be carried out only for the arithmetic achievement tests and for one of the analytical ability subtests. The results of these analyses should prove fruitful not only in the respective substantive information they provide, but also in pointing towards more effective ways of test construction in the areas in question.

Section 1.2

The units on which the above analyses were carried out are as follows:

(a) the subtests in a test battery;

(b) individual test items;

(c) distractors of test items.

This report, then, concentrates on the following analyses:

Correlational analyses of subtests

These are described in Sections 2.2 and 2.3. The smallest space resulting from our analysis could be easily interpreted in terms of the facets defining the various tests. Thus, school achievement, analytical ability, and aptitude-achievement fell into distinct contiguous regions, and within analytical ability, the tests fell into regions according to two facets: (a) the task given to the subject; and (b) the language of presentation. These results were obtained by analysis of the result of our test battery as well as of those of a previously published study. In the latter it also became apparent that analytical ability, aptitude-achievement, and school achievement tend to fall in concentric circles, with analytical ability in the centre, which is in line with a previous analysis of Thurstone's data.[3,4]

Correlational Analysis of individual items

Here, again, it was found that the facet definition of the arithmetic achievement items determine their position within the space, i.e. the correlational structure of the test is partly predictable by the content of test items as defined by an a priori faceted definition. These results are reported in Section 3.1.

Profile analysis of subtests

Various pairs of subtests in arithmetic were found to form quasi-scales. This is evidence of stages of achievement in arithmetic, which, however, form only a "partial order". It was further shown in Section 2.4 that information about these stages is obtainable neither from the distribution of scores in each individual subtest, nor from the more customary correlational analysis.

Profile analyses of test items

These analyses were carried out both on analytical ability subtests and on arithmetic subtests. In almost every one of these subtests, considerable constraints were found to operate on the occurrence of profiles of scores. The nature of the analysis for these two groups of subtests -- analytical ability and arithmetic -- was different in one important respect. In the case of the analytical ability test, there was no theoretical basis for predicting the nature of these constraints, because individual test items were not systematically constructed according to a faceted definition (Section 4.2). On the other hand, items in the arithmetic subtest differ from each other in three facets. This fact, combined with some intuitions regarding stages of achievement, leads to an analysis guided by two hypotheses and subsequently interpreted by the same. A partial order was found which may be taken as evidence of stages of acquisition within each subtest (Section 4.3).

Distractor analyses

Distractors are usually constructed on the basis of intuition as to what answer might be attractive, or else by obtaining attractive answers from tests which are first given in open-ended form. If distractors are constructed in systematic fashion, the subjects can be assigned differential scores on the type of wrong answers to which they are attracted. This may lead to a diagnostically useful typology of errors.

The smallest space analysis shows that in our arithmetic tests and to a certain extent in one of the analytical ability subtests different types of distractors fall into contiguous regions. This means that subjects who tend to make certain types of errors on one item tend to make the same kind of error on other items too. In the arithmetic test, this tendency was dependent to a large extent on the degree of mastery that the different subjects had achieved over the material. As they progress in the area tested, subjects first learn to avoid one type of error, next another kind. This was revealed by a multidimensional scalogram analysis which resulted in a scale of profiles consisting of error scores on three types of distractors and the score of correct answers.

* * * * *

While the main efforts of our study had to be directed towards an exploration of the utility of the new kinds of nonmetric analyses in the area of testing and their adaptation for diagnostic purposes, there are a number of substantive results which should prove to be of immediate practical value. Information about the developmental stages in learning arithmetic, for instance, may have repercussions on teaching methods. Likewise, the distractor analyses will be important in revealing the kinds of difficulties encountered by students in this test. As will be shown in more detail in later sections, such information is not revealed by current methods of test analysis.

2. ANALYSES OF RELATIONSHIPS BETWEEN SUBTESTS

2.1. The SSA-I Computer Program

In order to analyze the structure of interrelationships between the various subtests included in our battery, 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.(12) The order of these distances, but not their absolute sizes, is preserved, and the program derives the smallest possible Euclidean space for these items.

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 distance, 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.

2.2. Results of the Smallest Space Analysis

Correlational analysis was carried out for all tests included in this study. A list of the subtests is given in Table 2.1. This table

Table 2.1

Test Variables Included in Smallest Space Analysis

Code	Description of Test	Example No.*
	Hebrew Achievement Test	
Comp 1	Comprehension test, story A	-
Comp 2	Comprehension test, story A	-
Comp 3	Comprehension Test, story B	-
Comp 4	Comprehension Test, story B	-
Voc	Vocabulary test	-
	Arithmetic Achievement Test - percentages	
Y	Subject's task is to find yield	11
R	Subject's task is to find rate	12
P	Subject's task is to find principal	13
T	Items requiring multiplication or division by length of time	14
	Geometry Achievement Test	
Geom A	Questions on angles	15
Geom P	Questions on polygons	16
Geom S	Questions on solids	17
	Intelligence Tests	
Cp	Completion: part is given, subject is to find whole in which part is embedded	1
Cf	Completion: part is given, subject is to find whole in which part is embedded	2
Df	Differences: find which figure is different from given set	3
Sv	Similarities: Find which is similar to given set	18
Sf	Similarities: Find which is similar to given set	4
Sp_1	Similarities: Find which is similar to given set	5
Sp_2	Similarities: Find which is similar to given set	6
SI_p	Similarities: the pictures in the given set describe an incongruous situation, correct answer must describe the incongruous situation	7
SR_p	Similarities: Names of the pictures in the given set rhyme with each other	-
Ap	Analogies	19
Av_1	Analogies	8
Av_2	Analogies	20
Pp	Progressions	21
Pf	Progressions	9
Pd	Progressions	10
Pv	Progressions based on alphabetical order (e.g., acorn, building, code, danger...)	22

* Refers to number of example in appendix

Table 2.2. Intercorrelation Matrix of Test Variables *

Variable No.	Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	Comp 1																											
2	Comp 2	45																										
3	Comp 3	35	33																									
4	Comp 4	45	42	50																								
5	Voc	45	41	34	43																							
6	Y	39	31	27	35	37																						
7	R	41	35	31	38	40	58																					
8	P	40	34	31	39	40	54	61																				
9	T	38	37	32	41	40	57	65	67																			
10	Geom A	35	34	27	32	36	43	45	44	47																		
11	Geom P	39	33	29	39	35	42	48	47	50	50																	
12	Geom S	33	27	27	31	33	38	39	40	42	42	44																
13	Cp	22	21	22	27	25	24	24	24	25	22	21	21															
14	Cf	28	24	25	28	28	32	33	35	35	31	30	27	23														
15	Df	28	24	24	30	28	28	27	30	30	27	29	28	46	27													
16	Sv	35	30	30	38	33	32	33	34	36	32	32	30	32	33	34												
17	Sf	37	31	29	35	37	34	42	36	39	34	36	33	46	33	33	56											
18	Sp1	27	27	23	31	28	24	29	29	29	27	26	22	23	26	25	31	29										
19	Sp2	24	19	21	25	26	24	28	27	26	25	25	20	22	23	24	25	26	50									
20	SIp	30	28	26	34	34	34	33	34	35	31	32	28	26	29	31	32	35	38	36								
21	SRp	17	17	15	19	21	19	20	21	22	17	16	14	18	19	22	19	23	30	37	31							
22	AP	27	24	21	30	33	25	29	29	33	27	26	23	20	26	23	25	30	23	20	31	18						
23	Av1	31	29	26	34	30	30	33	32	34	31	32	28	22	27	29	33	31	35	39	44	26	29					
24	Av2	26	22	21	28	26	25	27	25	26	23	24	21	18	19	24	24	26	36	39	41	29	23	49				
25	Pp	23	21	16	23	28	22	24	24	26	25	23	20	17	18	24	26	25	23	21	30	19	27	26	22			
26	Pf	24	21	20	24	23	24	27	30	34	23	25	21	16	54	21	22	24	22	19	27	19	24	28	20	24		
27	Pd	33	29	28	33	30	28	31	30	30	28	25	28	20	28	29	29	32	39	43	41	30	27	45	44	24	27	
28	Pv	43	38	36	47	43	40	41	39	41	39	40	38	23	36	33	44	41	32	31	42	24	35	41	32	30	29	46
Variable No.		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27

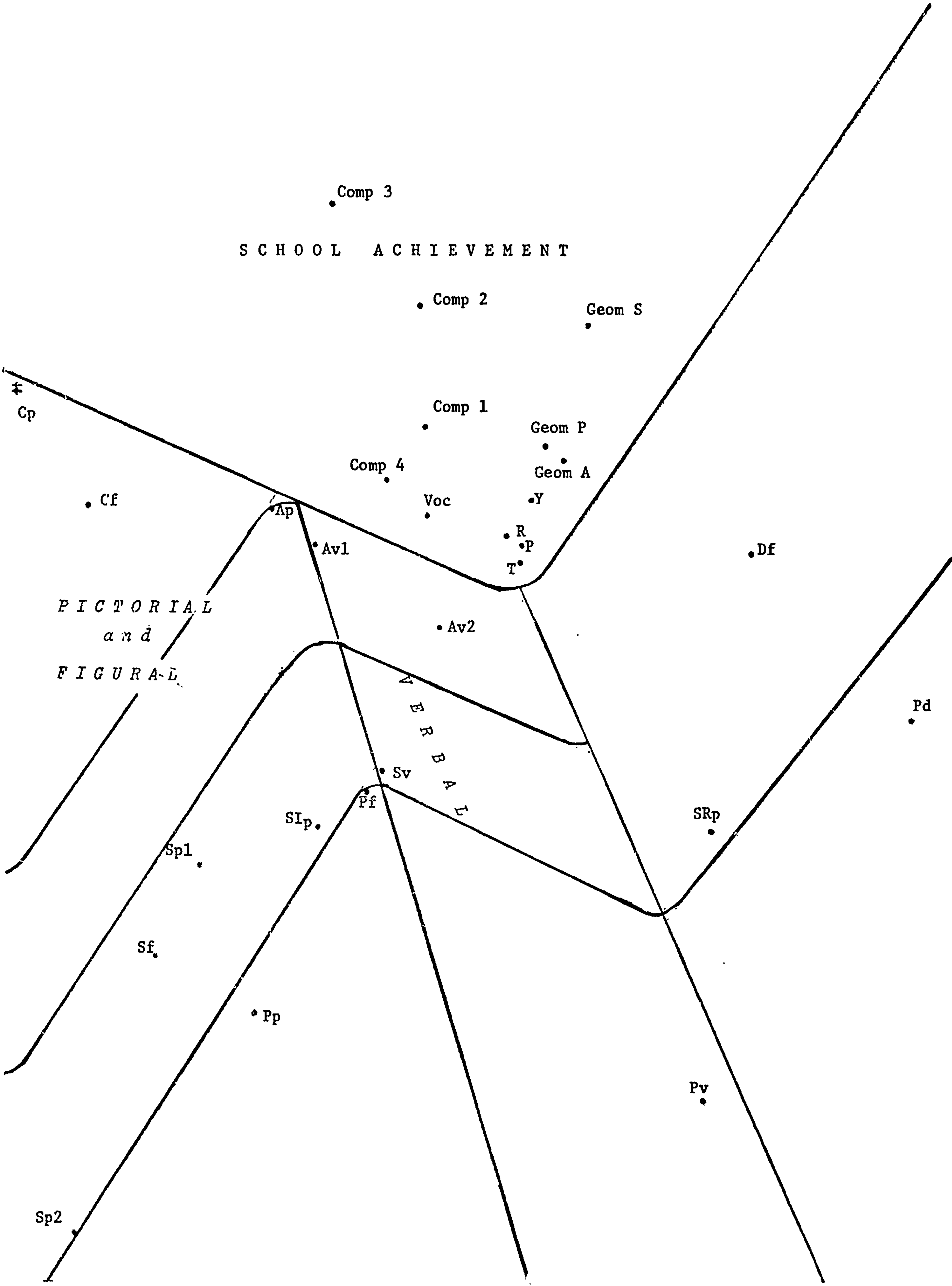
* Decimal points are omitted.

also lists examples of items for these tests in the appendix (except for the Hebrew tests). Table 2.2 contains the product moment correlations obtained between the test variables. (For Table 2.1 and 2.2 see pages 10, 11)

Each subtest in Table 2.1 has been assigned a code. For the ability tests, the code is based on the two facets employed to classify them, namely the subject's task, indicated by the first (capital) letter, and the language of presentation, indicated by the second (small) letter of the code (v = verbal, p = pictorial, f = figural, and d = digits).

A rather fair fit was obtained with the SSA-I: the coefficient of alienation was .199 for two dimensions, and .142 for three dimensions. Figure 2.1 shows the space diagram of the two dimensional space as printed out by the computer. The picture rendered by the three-dimensional space gives an essentially similar configuration. The computer program also prints out three two-dimensional projection diagrams for a three-space. For simplicity, the two dimensional picture is presented here, as it portrays the essentials of the configuration of the tests.

Fig.2.1. The two-space resulting from Smallest Space Analysis I representing the interrelationships between the subtests of Table 2.1. (see figure on next page)


Comp 3
SCHOOL ACHIEVEMENT
Comp 2
Geom S
Cp
Comp 1
Geom P
Comp 4
Geom A
Cf
Ap
Voc
Y
R
P
Av1
T
Df
PICTORIAL
and
FIGURAL
Av2
VERBAL
Pd
Sv
Pf
SIp
SRp
Sp1
Sf
Pp
Pv
Sp2

Inspection of the figure reveals several contiguous regions which have been marked off by us in the diagram.

First, there is a region containing scores on achievement tests in school subjects. This divides into three sub-regions: Hebrew (Com 1, 2,3,4 and Voc), Arithmetic (R,P,Y,T), and Geometry. Notice that points representing school subjects fall closer together than points in the other regions of ability tests. This reflects the fact that correlations between school subjects are higher than those between intelligence tests. The highest correlations are those among subtests in arithmetic: Y,R,P,T.

Further, the space diagram contains a region of intelligence tests. The boundary lines between these tests go in two directions. In one direction we have a set of curved parallel lines dividing the two-space into four regions according to the task posed to the subject: Progressions (Pf, Pp, Pv, Pd), Similarities (Sp_2, Sf, Sp_1, SIp, Sv, SRp), Analogies (Ap, Av_1, Av_2) and Completion (Cp and Cf).

Cutting across the above lines run boundaries distinguishing between languages of presentation. Pd, the test in which the subject is presented with items in the form of digits, lies in a region by itself (in the centre, right-hand side of the figure). The space diagram distinguishes further between verbal tests and non-verbal tests. There seems to be no distinction, at least for our data, between pictorial tests and figural tests. The former might have been expected to be closer to the verbal tests, since meaningful pictorial content might be regarded as closer to verbalization than non-contentful figures, but the present data do not bear this out.

Only two tests fall outside their proper region: Df and SRp. Note that to solve items of the latter test the subject must name the picture so as to find the one which does not rhyme. This translation into verbal form virtually turns SRp into a verbal test, and, in fact, as seen in the diagram, it is close to the verbal region.

These relations which emerge in the space diagram of the Smallest Space Analysis may of course be checked against the correlation coefficients that serve as input to the analysis. The SSA-I merely facilitates the interpretation of the matrix.

On the basis of the faceted definition of intelligence, a distinction has previously been made between analytical ability and achievement tests.[2,3] By the term "Achievement", a certain type of intelligence was referred to and not school achievement. To avoid confusion, the term aptitude-achievement will be used instead, in the following.

To illustrate the distinction between analytical ability and aptitude-achievement, consider the following two items:

(1) The young of the cat are called ______.

(2) Hen is to chicken as cat is to ________.

In (1) the relation which has to be applied in solving the question is given ("the young of"), and assumed to be understood by the subject. By contrast, in (2) the relation has to be elicited. The former is an instant of an aptitude-achievement test item, and the latter, one of analytical ability.

Section 2.2

Consider now the subtests of our intelligence test battery. Analogy items, test items asking about similarity or about differences and so-called progressions, are all analytical ability test items; the relation in question is not given but must be educed by the subject from the item. The only aptitude-achievement tests are the completion tests, Cf, and Cp. Here the part of the figure is given and the subject is required to point out which is the "whole" including this part; see Examples 1 and 2 in the appendix. The distinction between these two kinds of intelligence is by no means always an easy one to make. The above two tests are actually on the border line between analytical ability and aptitude-achievement.

In a reanalysis of Thurstone's intercorrelation matrices(3) it has been shown that analytical ability tends to occupy the central region of the space, and aptitude-achievement -- the surrounding regions. It is of interest that the two tests of the present battery, which may be looked upon as aptitude-achievement tests (Cf and Cp) are rather far apart from the other intelligence tests. Though they are suggestive, these results cannot be regarded as a satisfactory replication of the previous analysis in view of the small number of aptitude-achievement tests involved. To obtain such a replication, another analysis was conducted of the intercorrelation matrix of a recently published study, and this will be reported in Section 2.3.

In conclusion, it may be said that the faceted definition of our tests is definitely reflected in the structure of intercorrelations as shown by the Smallest Space Analysis.

2.3. Analysis of Data From Hoeger's Study Of Ability and Achievement Tests

To obtain additional information on the structure of analytical ability and its relation to aptitude-achievement and to school achievement, the results of a recent study(9) were reanalyzed. In the latter study, there were several tests which were given to 519 high-school students in a West German city, and these can be classified into analytical ability tests on the one hand and aptitude-achievement on the other. The results of these tests, together with school grades in 11 subjects, had been submitted to a factor analysis. A description of the tests and school subjects is given in Table 2.3. In the right hand column, each intelligence test is given a code so as to facilitate identification of test content and comparison of the tests with those in our own study. This code identifies the task of the subject (first letter, capital) and the language of presentation of the test (second letter). The tasks of the subject were: C - completion; S - similarities; D - differences; P - progressions; etc. Languages of presentation are: v - verbal; f - figural; d - digits.

Product moment correlations between the above 20 variables as reported by Hoeger are given in Table 2.4. The table also includes the six factors resulting from Hoeger's factor analysis. The matrix of correlations obtained between the 20 variables of Table 2.3 was submitted to a Smallest Space Analysis, the G-L SSA-I, which resulted in a rather good fit (coefficients of alienation were .182 for two dimensions and .130 for three dimensions.

Table 2.3

Variables Included in Hoeger's Study

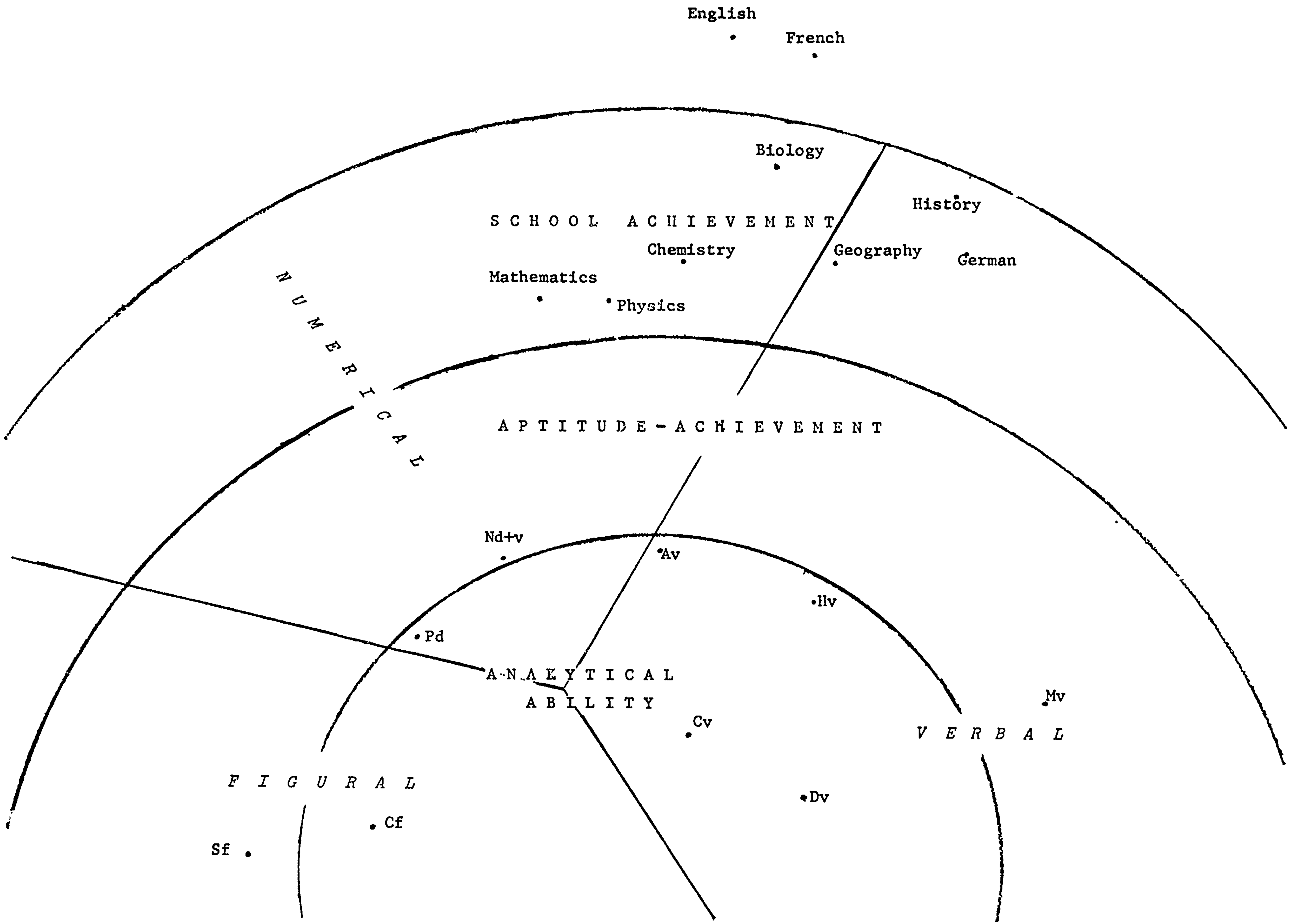
Code	Description of Variable
Cv	Complete one missing word in sentence
Dv	Find which word is different from given set of words
Av	Word analogies
Hv	Give superordinate of two words (e.g., rose-tulip)
Pd	Numerical progressions
Cf	Find which of five geometric figures (circles, squares, etc) can be put together from given parts of figure
Mv	Subject memorizes 25 words each belonging to one of the following categories: flowers, tools, artifacts, birds, animals; then he is asked questions of the following form: The word beginning with the letter a was: ..... (a flower, a tool, a bird.....)
Nd+v	Verbally formulated arithmetic problems
Sf	Match cubes presented in different orientations in space
	German
	History
	Geography
	English
	French
	Mathematics
	Physics
	Chemistry
	Biology
	Arts
	Music

Inspection of the space diagrams showed, however, that "Arts" and "Music" tended to cluster together in one corner of the space diagram, away from the points representing other variables. This is a consequence of the relatively small correlation coefficients of these two variables with the others. Clearly, these two variables go off in a direction not common to the other variables.

Accordingly, the matrix was run again through SSA-I without "Arts" and "Music". This resulted in a slightly better fit (coefficients of alienation .162 and .103 for two and three dimensions respectively), and more importantly, gave a clearer picture, with the points representing variables spread out over a larger area of the space diagram. The space diagram for two dimensions is reproduced here in Figure 2.2.

The first point of interest here is that analytical ability tests occupy the central region and aptitude-achievement tests fall outside this region. While there is only a small number of the aptitude-achievement tests involved, this seems to be an additional confirmation of the earlier analysis of Thurstone's data (Section 2.2). School achievement is removed still further from analytical ability. In retrospect, such a result is of course quite plausible. Figure 2.2 thus shows three concentric circles: analytical ability, aptitude-achievement, and school achievement.

Fig. 2.2. The two-space resulting from Smallest-Space-Analysis I representing the interrelationships between test variables of Hoeger's study. (see figure on next page)


English
French
Biology
History
SCHOOL ACHIEVEMENT
Chemistry
Geography
German
Mathematics
Physics
NUMERICAL
APTITUDE-ACHIEVEMENT
Nd+v
Av
Hv
Pd
ANALYTICAL
ABILITY
Cv
Mv
VERBAL
FIGURAL
Cf
Sf
Dv

As was the case in our test battery (Section 2.2), school achievement tests tended to fall much closer together than the aptitude-achievement and the analytical ability tests. This reflects the fact that the relationship between achievement in different school subjects is stronger than that between different kinds of intelligence.

Among the school subjects we find the following sub-regions:

1) Foreign languages: English and French.
2) Science: Chemistry, Physics, Mathematics, and Biology. Of these, Biology is closest to foreign languages.
3) History and Geography.
4) German, the native language of the subjects tested. This is closer to History and Geography than to foreign languages.

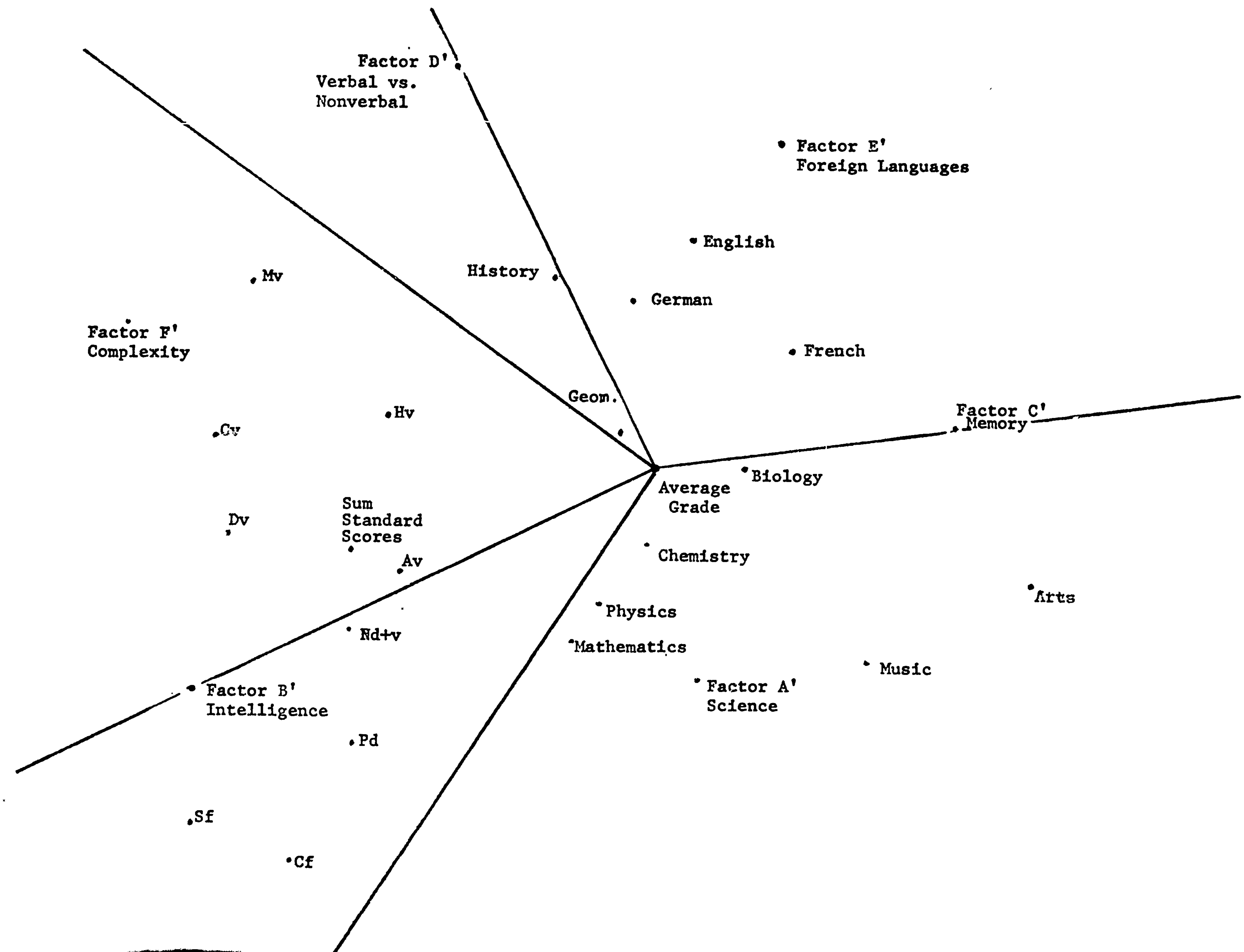
It is of interest that "Music" and "Arts" are closer to Science than to Languages or to History and Geography; this is shown by the space diagram of the previous analysis in which these two variables are included. (See Figure 2.3 below.)

The analytical ability tests fall into three regions according to language of presentation. Of the two tests involving digits, the arithmetic problems test, Nd+v, in which the problem is presented in verbal form, is closer both to the verbal region and to school grades than the number progressions test, Pd, which does not involve words. As expected, these tests are closer to the Sciences than to Geography, History, and Language grades. In the verbal region, the analogy and concept formation tests, Av and Hv, are closer to school grades than the other tests. Tests which are presented in the form of figures, Sf and Cf, are on the whole closer to the region of digit tests than to that of verbal tests, and are least related to the school subjects.

Hoeger submitted his correlation matrix to a factor analysis, using the centroid method with orthogonal rotation. It is instructive to compare the results of this factor analysis with those of the SSA-I. We therefore ran on the SSA-I program another matrix which included the six principal common factors revealed by Hoeger's analysis, (see Table 2.4). In this matrix, loadings of the variables on the factors are the correlations between the variables and the factors. The correlations between factors are taken to be zero (because of the orthogonality of the common factors). In addition to the common factors, all variables which went into the factor analysis were included, viz. the 20 variables of Table 2.3, the sum of standard scores of the intelligence tests, and the average score of school grades (see Table 2.4). The two-dimensional space diagram which resulted from the SSA-I gave a rather poor fit (coefficient of alienation .254), which is due in part to the fact that music and arts scores were included. Still, this diagram -- which is presented in Figure 2.3 -- gives an easily interpretable picture.

Fig. 2.3. The two-space resulting from Smallest Space Analysis I representing interrelationships between variables and six principal common factors of Hoeger's study; see Table 2.4.

(see figure on next page)



Factor D'
Verbal vs.
Nonverbal
Factor E'
Foreign Languages
English
History
German
Mv
Factor F'
Complexity
French
Geom.
Hv
Cv
Factor C'
Memory
Biology
Average
Grade
Sum
Standard
Scores
Dv
Av
Chemistry
Arts
Physics
Nd+v
Mathematics
Music
Factor A'
Science
Factor B'
Intelligence
Pd
Sf
Cf

Table 2.4. Intercorrelation Matrix of Variables in Hoeger's Study

Variable No.	Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	Cv																											
2	Dv	26																										
3	Av	36	35																									
4	Hv	29	22	34																								
5	Mv	21	19	17	22																							
6	Nd+v	24	24	49	24	19																						
7	Pd	27	17	21	18	22	46																					
8	Cf	24	19	27	22	08	35	30																				
9	Sf	14	13	13	11	15	18	27	39																			
10	Sum Stand.Scores	51	46	60	58	47	60	55	56	44																		
11	German	17	16	25	16	18	14	14	-06	-06	17																	
12	History	13	16	16	14	16	15	07	-04	01	20	35																
13	Geography	18	11	18	18	20	16	13	05	08	23	30	38															
14	English	12	09	17	21	09	16	03	-03	03	20	31	18	25														
15	French	08	09	13	19	11	16	10	02	00	19	31	26	26	53													
16	Mathematics	14	14	29	15	10	35	28	14	09	33	23	25	35	17	16												
17	Physics	21	18	32	20	03	34	24	16	07	33	25	26	28	27	27	61											
18	Chemistry	15	14	30	20	11	27	24	12	07	31	30	36	42	25	31	52	55										
19	Biology	11	11	21	17	15	13	10	03	00	19	31	35	51	28	35	35	39	54									
20	Arts	-03	-03	06	06	00	-03	00	07	02	02	08	09	10	05	33	09	12	17	21								
21	Music	06	16	05	12	08	04	12	03	03	11	21	08	22	11	19	12	16	24	27	21							
22	Average Grade	17	21	32	25	16	30	23	08	07	36	51	55	60	53	59	62	65	71	72	28	43						
23	Factor A'	06	10	29	04	-06	33	27	13	02	12	25	31	37	16	21	72	73	68	48	15	22	74					
24	Factor B'	50	44	57	49	30	49	42	57	38	89	04	00	06	-01	-01	14	19	14	03	-01	10	01	00				
25	Factor C'	05	02	-12	15	17	-15	05	07	15	10	25	20	38	17	42	-05	-02	24	45	40	39	47	00	00			
26	Factor D'	23	19	28	17	21	01	-17	-29	-26	05	39	41	34	21	09	05	07	17	30	-12	00	28	60	00	00		
27	Factor E'	04	11	18	27	03	17	-04	-03	-05	22	18	06	00	61	58	11	25	10	11	15	06	38	00	00	00	00	
28	Factor F'	80	02	-01	02	33	33	41	09	26	34	15	24	20	21	13	18	00	04	00	-15	-09	19	00	00	00	00	00
	Variable No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27

The figure is partitioned into five sectors, the boundaries of which radiate from the point representing the subjects' average school grades. These five sectors correspond essentially to the regions identified in the previous analysis:

(1) History and Geography;

(2) Languages (which here includes the native language, German);

(3) Science, with Arts and Music at the far end of the sector;

(4) digits and figural tests;

(5) verbal tests.

The six common-factors tend to be outermost points forming a circle. (Remember that their correlations with each other are zero, and that they should therefore be far apart from each other). The figure shows that the factors provide little information beyond supplying labels to those regions which have already been identified by our previous smallest space analysis, where no common factors were included. Three of the factors fall within sectors:

<u>Factor A'</u>, "Science" falls in the Science sector

<u>Factor E'</u>, "Foreign languages" falls in the Foreign language sector

<u>Factor F'</u>, called "Complexity" by Hoeger, is in the verbal intelligence sector.

The three remaining factors fall on the boundary line between sectors:

<u>Factor B'</u>, described by Hoeger as the ability to cope with the intelligence tests, fall on the boundary between the two sectors of intelligence tests.

Factor C', the "Memory" factor, falls between the two school subject sectors.

Factor D', is interpreted by Hoeger as a bi-polar factor with achievement involving verbal ability as the positive pole, and non-verbal achievement as the other pole; this factor falls on the boundary between the Language sector and the History and Geography sector.

It may be concluded that, in the present case at least, the SSA-I as a way of presenting data is preferable to factor analysis, because it renders a much more easily interpretable picture. Also, by viewing the space as a whole, as well as divided into regions identifiable by a facet design, we are spared the unnecessary quest after "meaningful coordinate axes" which dominates so much of current factor analysis. One does not need any special rotation or reference to coordinate axes in order properly to interpret Figures 2.2 and 2.3.

2.4. Scalability of Arithmetic Subtest

The six subtests of arithmetic achievement are shown in Table 2.5. The table also shows the mean number of errors on these subtests, which are arranged in ascending order of difficulty, when the number of items in each subtest is taken into account.

Table 2.5

Mean Number of Errors in Each of the

Six Arithmetic Subtests (N = 600)

Question asked	Code	Example No. *	Mean no. of Errors	No. of Items	Mean No. of Errors per Item
Yield	Y	11	0.817	4	0.204
Rate	R	12	1.125	4	0.281
Yield (with time operation)**	YT	24	0,911	3	0.304
Principal	P	13	1.582	4	0.395
Principal (with time operation)	PT	23	0.950	2	0.475
Rate (with time operation)	RT	14	1.436	3	0.479

* Numbers refer to examples in the Appendix
** Items involving time operation - items about interest for a period other than one year, and thus requiring an additional arithmetical operation.

The question may now be raised whether these subtests, in addition to differing in difficulty, also form stages of learning in the sample tested. In other words, we may ask whether acquiring the ability to solve items of one subtest presupposes a mastery of another subtest. Inspecting the joint distribution matrices of the subtests, it becomes apparent that some of them form a primitive scale. Consider, for instance, subtests P and Y. Table 2.6 shows that there are only 12 subjects out of the 600 in our sample, or 2 per cent, who have solved more than one of the four items of subtest P while solving less than two items of the four on subtest Y. These two subtests, then, may be said to form a primitive scale.

Table 2.6

Joint Frequency Distribution of Number of

Correct Responses in Subtests Y and P.

Y \ P	0	1	2	3	4	Total
0	8	3	2	1	1	15
1	19	19	5	2	1	46
2	18	24	19	16	3	80
3	10	27	37	43	15	132
4	7	32	58	82	148	327
Total	62	105	121	144	168	600

As has been shown previously[5], this result is not always obtained even where subtests differ in degree of difficulty. An example of two of our tests which do not seem to scale appreciably is given in Table 2.7.

Table 2.7

Joint Frequency Distribution of Number of

Correct Responses in Subtests PT and R

PT \ R	0	1	2	3	4	Total
0	10	30	39	45	63	187
1	5	12	33	51	95	196
2	0	4	8	36	169	217
Total	15	46	80	132	327	600

To the extent that two subtests are scalable in the sense described above, this may be interpreted as two distinct stages of acquisition. This information, therefore, is of considerable practical importance.

It should be borne in mind that this property of scalability does not imply a high correlation coefficient. In fact, it has been shown previously(1) that the joint correlation coefficient of two variables forming a scale may approach zero, and its size depends on the frequency distribution of the two variables. In the case of our subtests, the smaller the correlation coefficient the greater the extent of scalability of the subtests. This can be shown by comparing a rough index of scalability with the product moment correlation coefficient. To obtain this index, scores on all subtests were dichotomized as follows:

Subtest of four items each (Y, P, and R): 0 and 1 correct answers vs. 2,3 and 4.

Subtest of three items each (TY and RT): 0 and 1 correct answers vs. 2 and 3 correct answers.

Subtest PT with two items: 0 correct answers vs. 1 and 2 correct answers.

The joint frequency distributions of the various subtests can now be put into fourfold distribution tables. An example is Table 2.8 which shows the data of Table 2.6 after dichotomization.

Table 2.8

Data of Table 2.6 after Dichotomization

Y \ P	0-1	2-4	Total
0-1	49	12	61
2-4	118	421	539
Total	167	433	600

The smaller of the two off-diagonal cells may serve as a rough index of scalability: the smaller the number, the better the scalability; a zero entry would indicate perfect scalability. However, these indices are comparable only for those pairs of subtests in which the scores were dichotomized in the same way. Accordingly, the 15 pairs of subtests fall into five groups as shown in Table 2.9. Within each of these groups there is an inverse relationship between Pearson's r and the Index of Scalability, the only exception being the two pairs of Group IV.

Table 2.9

The Relationship between Scalability and Correlation

Group	Subtests *	Index of Scalability	Pearson's r
I	Y(5), RT(4)	8	.487
	Y(5), YT(4)	8	.583
	R(5), RT(4)	13	.584
	P(5), RT(4)	35	.566
	R(5), YT(4)	38	.557
	P(5), TY(4)	64	.593
II	P(5), Y(5)	12	.552
	R(5), Y(5)	17	.627
	P(5), R(5)	30	.628
III	Y(5), PT(3)	21	.401
	R(5), PT(3)	46	.419
	P(5), PT(3)	72	.487
IV	RT(4), PT(3)	50	.510
	YT(4), PT(3)	79	.443
V	YT(4), RT(4)	32	.523

* The number of items in each subtest is given in brackets.

What "stages" of achievement in arithmetic are revealed by Table 2.9? If we accept an index of scalability of less than 25 (i.e. about 4 per cent of the subjects in our sample) as a rough-and-ready criterion, these stages can be described as in Figure 2.4.

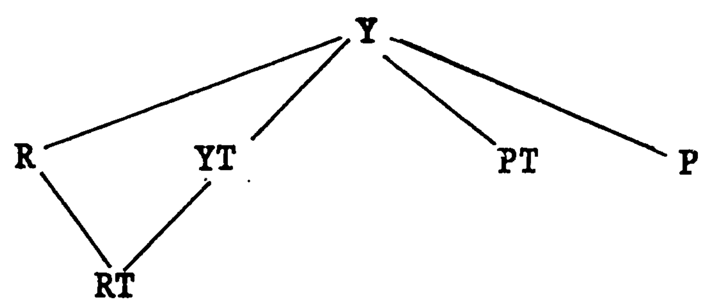


Figure 2.4 - Stages of Acquisition in the Arithmetic Achievement Test.

As shown in this figure, subtest Y is learnt before all other subtests and subtests R and YT before subtest RT. Subtests R, YT, PT, and P, though differing from each other in difficulty, do not scale by the above criterion; they do not form distinct stages. Likewise, the pairs PT - RT and P - RT don't form scales. The six subtests permit of fifteen pair-wise comparisons and, as shown in Table 2.5, in each of these fifteen pairs, the subtests are of unequal difficulty. By way of comparison, there are seven pairs of subtests (including the pair Y and RT) which form stages, as can be seen in Figure 2.4. This goes to show, then, that information about stages is not recoverable from information about degree of difficulty, or vice versa. cf Section 4.3, where stages of achievement will be discussed through analysis of profiles rendered by individual test items.

To conclude, information about scalability of different subtests in an achievement test will give extra information as to the stages of learning in the school subject in question. This information is not contained in the correlation matrix or in the mean scores of the subtest. For our arithmetic test, Table 2.10 gives the 15 joint frequency distributions, and Table 2.11 presents the same data after dichotomization.

Table 2.10. Joint Frequency Distribution of Number of Correct Responses in the Six Arithmetic Subtests

Subtest		P					Y					R					PT			YT				RT			
		0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	0	1	2	3	0	1	2	3
	0						8	19	18	10	7	22	17	12	5	6	43	15	4	26	19	12	5	31	22	9	0
	1						3	19	24	27	32	11	30	25	27	12	52	40	13	24	34	34	13	44	35	20	6
P	2						2	5	19	37	58	8	12	29	31	41	41	48	32	10	21	42	48	34	36	35	14
	3						1	2	16	43	82	1	8	18	35	82	37	50	57	5	18	41	80	13	37	58	36
	4						1	1	3	15	148	0	1	6	25	136	14	43	111	1	5	26	136	5	18	59	86
	0	8	3	2	1	1						9	3	2	0	1	10	5	0	8	5	1	1	10	4	1	0
	1	19	19	5	2	1						15	17	10	3	1	30	12	4	20	20	5	0	25	14	6	1
Y	2	18	24	19	16	3						8	26	14	22	10	39	33	8	18	22	22	18	29	33	17	1
	3	10	27	37	43	15						8	12	30	34	48	45	51	36	15	26	43	48	33	41	39	19
	4	7	32	58	82	148						2	10	34	64	217	63	95	169	5	24	83	215	30	58	118	121
	0	22	11	8	1	0	9	15	8	8	2						24	17	1	18	14	10	0	28	10	4	0
	1	17	30	12	8	1	3	17	26	12	10						40	21	7	17	23	20	8	30	29	9	0
R	2	12	25	29	18	6	2	10	14	30	34						42	28	20	15	22	29	24	34	34	14	8
	3	5	27	31	35	25	0	3	22	34	64						37	47	39	10	20	37	56	25	31	47	20
	4	6	12	41	82	136	1	1	10	48	217						44	83	150	6	18	59	194	10	46	107	114
	0	43	52	41	37	14	10	30	39	45	63	24	40	42	37	44				43	41	60	43	79	58	40	10
PT	1	15	40	48	50	43	5	12	33	51	95	17	21	28	47	83				22	43	50	81	38	63	61	34
	2	4	13	32	57	111	0	4	8	36	169	1	7	20	34	150				1	13	45	158	10	29	80	98
	0	26	24	10	5	1	8	20	18	15	5	18	17	15	10	6	43	22	1					37	20	8	1
	1	19	34	21	18	5	5	20	22	26	24	14	23	22	20	18	41	43	13					35	39	18	5
YT	2	12	34	42	41	26	1	6	22	43	83	10	20	29	37	59	60	50	45					38	42	55	20
	3	5	13	48	80	136	1	0	18	48	215	0	8	24	56	194	43	81	158					17	49	100	116
	0	31	44	34	13	5	10	25	29	33	30	28	30	34	25	10	79	38	10	37	35	38	17				
	1	22	35	38	37	18	4	14	33	41	58	10	29	34	31	46	58	63	29	20	39	42	49				
RT	2	9	20	35	58	59	1	6	17	39	118	4	9	14	47	107	40	61	80	8	18	55	100				
	3	0	6	14	36	86	0	1	1	19	121	0	0	8	20	114	10	34	98	1	5	20	116				

Table 2.11

Data of Table 2.10 after Dichotomization

		P		Y		R		PT		YT		RT	
		0-1	2-4	0-1	2-4	0-1	2-4	0	1-2	0-1	2-3	0-1	2-3
P	0-1			49	118	80	87	95	72	103	64	132	35
	2-4			12	421	30	403	92	341	60	373	145	288
Y	0-1	49	12			44	17	40	21	53	8	53	8
	2-4	118	421			66	473	147	392	110	429	224	315
R	0-1	80	30	44	66			64	46	72	38	97	13
	2-4	87	403	17	473			123	367	91	399	180	310
PT	0	95	92	40	147	64	123			84	103	137	50
	1-2	72	341	21	392	46	367			79	334	140	273
YT	0-1	103	60	53	110	72	91	84	79			131	32
	2-3	64	373	8	429	38	399	103	334			146	291
RT	0-1	132	145	53	224	97	180	137	140	131	146		
	2-3	35	288	8	315	13	310	50	273	32	291		

3. CORRELATIONAL ANALYSIS OF ARITHMETIC TEST ITEMS

In Chapter 2 an analysis was presented which included, inter alia, subtests of arithmetic ability. The present section presents a more fine-grain analysis of contingencies between individual items. Since answers to items are dichotomized as correct vs. incorrect, the usual correlation coefficients are not necessarily the optimal coefficients to be used. Instead, a similarity coefficient for dichotomies has been employed.

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

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

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

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.

Table 3.1 describes the items in our arithmetic test and also refers to examples in the appendix.

Table 3.1

Items of Arithmetic Subtests Included in the Smallest Space Analysis

Question is about	Item *not* involving "time" operation*			Item involving "time" operation*		
	Code	Example No.**	No.of Items	Code	Example No.	No.of Items
Principal	P	13	4	PT	23	2
Yield	Y	11	4	YT	24	3
Rate	R	12	4	RT	14	3
			12			8

* Items involving time operation - items about interest for a period other than one year, and thus requiring an additional arithmetical operation.

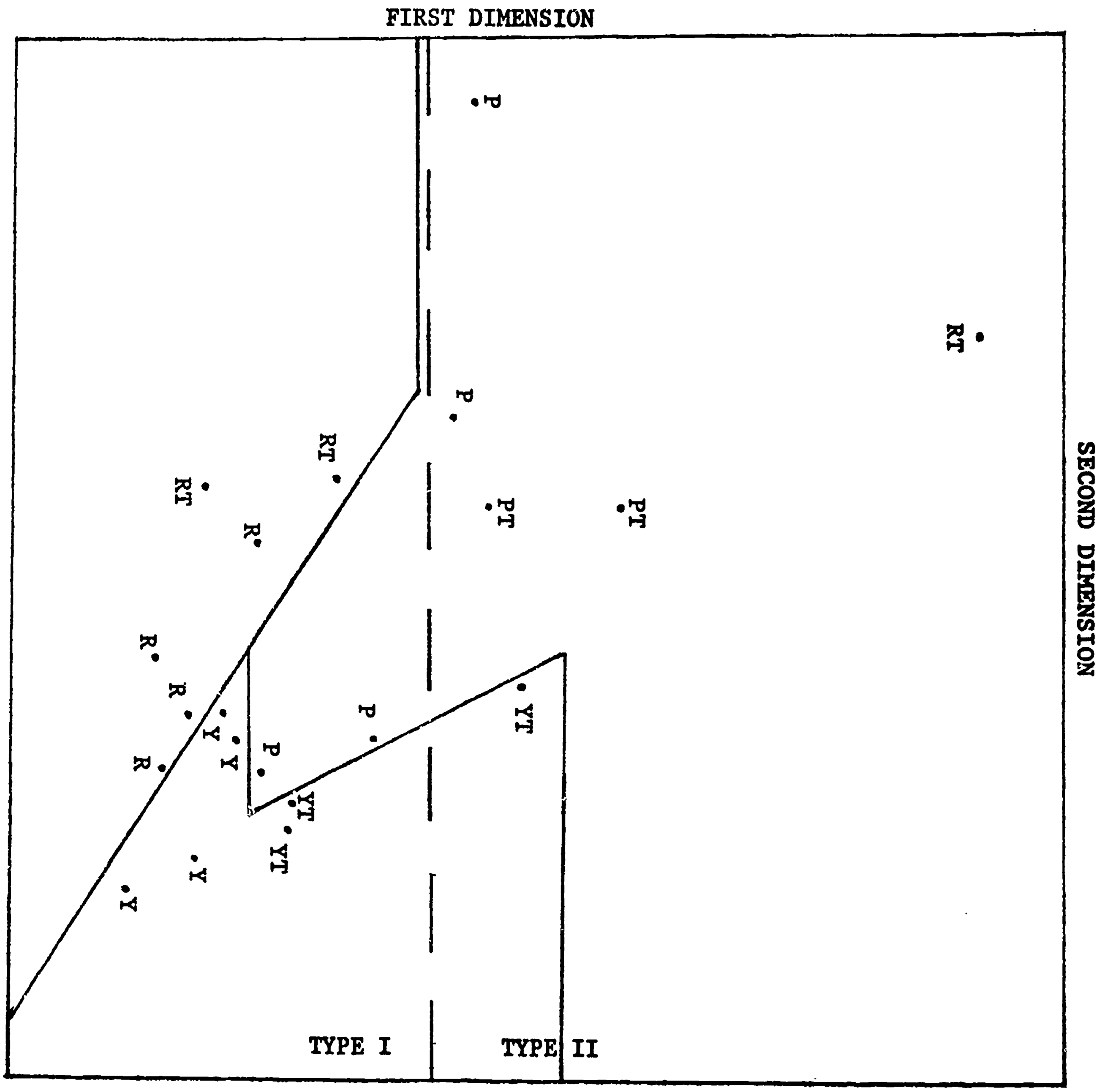
** Numbers refer to examples in appendix.

The matrix of similarity coefficients between these items is given in Table 3.2 (page 40). It was submitted to an SSA-I, which resulted in a good fit: coefficients of alienation .158 for the two-space, and .110 for the three-space. Contrary to our practice in other sections of this report, the three-space and not the two-space is presented here. Figure 3.1 shows the first two dimensions of the three-space (the third dimension accounts for much less of the variance than the first two dimensions).

Each point in the figure is labelled by a capital letter showing to which subject the item represented by the point belongs.

Figure 3.1 shows that the three main groups of items --- viz. those asking about principal, about yield, and about rate of interest --- fall within three clearly distinguishable regions without any overlap. These we have separated by boundary lines. The distinction between items not involving a "time" operation and T-items emerges less clearly in this figure: a line separating these two types of items from each other will have to make various turns, quite unlike the lines dividing between P, Y, and R.

Figure 3.1. The first two dimensions of the three-space resulting from Smallest Space Analysis I representing interrelationships between arithmetic test items. (see next page)



FIRST DIMENSION
SECOND DIMENSION
TYPE I
TYPE II
P
P
P
P
RT
RT
RT
PT
PT
R
R
R
R
Y
Y
Y
Y
YT
YT
YT

Section 3

Table 3.2

Interrelationships between Arithmetic Test Items

Similarity Coefficients for Dichotomies

Item No.	Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	Y																				
2	R	73																			
3	Y	79	74																		
4	YT	76	74	77																	
5	P	78	76	77	78																
6	RT	62	73	67	69	69															
7	YT	63	67	67	69	70	67														
8	P	56	68	63	66	68	68	69													
9	YT	76	73	77	75	77	69	72	64												
10	P	69	70	68	70	74	69	68	69	73											
11	RT	45	53	51	55	52	59	59	59	56	58										
12	R	78	78	74	74	76	72	68	68	75	71	54									
13	Y	74	71	72	73	74	66	67	66	71	68	56	74								
14	RT	61	67	63	65	66	72	65	67	68	68	59	68	67							
15	Y	78	77	77	75	78	68	70	67	79	72	52	79	77	70						
16	PT	55	63	62	64	64	64	67	68	64	65	63	63	62	70	67					
17	R	75	73	74	75	77	69	67	66	73	70	54	76	73	67	77	63				
18	PT	57	59	59	62	59	63	62	59	62	66	61	63	59	69	62	67	60			
19	R	69	71	71	70	72	70	67	70	69	70	57	73	70	69	75	64	74	63		
20	P	49	59	56	58	57	63	62	66	56	62	60	58	59	66	59	64	58	65	63	
Item No.		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Instead, another boundary line has been drawn distinguishing between the following two types of items:

Type I All figures appearing in the question are either whole numbers or decimals.

Type II One or more of the figures in the question are fractions.

Items of Type II may be expected to be more difficult than those of Type I, and this was actually the case. Further, the two types of items form contiguous regions in the space diagram. As shown in Figure 3.1, the first dimension of the smallest space distinguishes between these two types. As stated above, when similarity coefficients for dichotomies are employed, the first dimension will usually represent item difficulty. In the present case, the order of items on the first dimension corresponds only very roughly to their order of difficulty; this can be seen by comparison of Figure 3.1 with Table 3.3 which shows the percentage of subjects answering an item correctly. While the fit is crude, the boundary line between Type I and Type II items divides between those items answered correctly by 65 per cent or more of the subjects and those answered correctly by a smaller percentage. It appears therefore that Types I and II form separate regions by virtue of their differential difficulty.

Table 3.3

Percentage of Subjects Answering Correctly Each Item of the Arithmetic Test (in Descending Order of Difficulty).

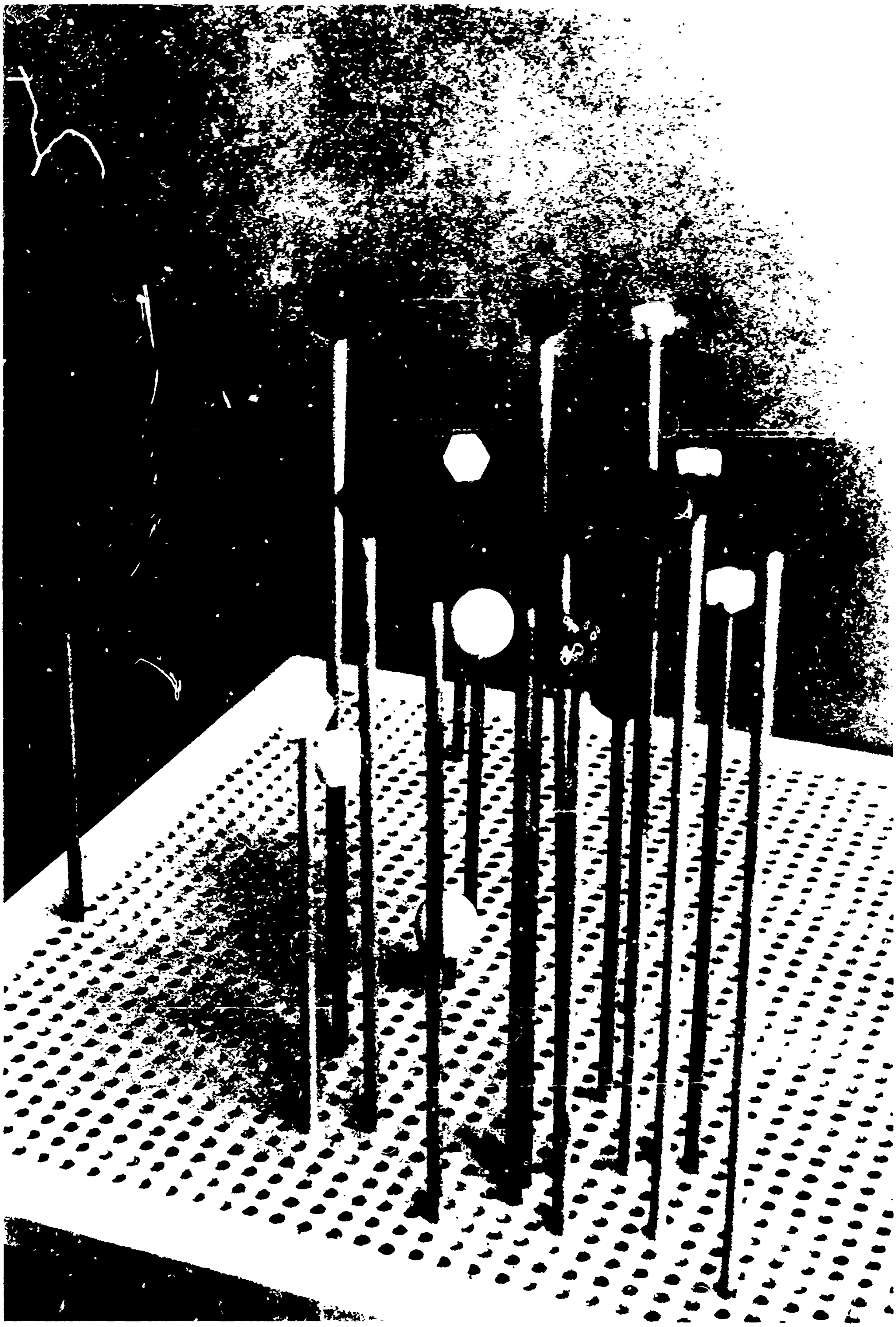
Item No.	Percentage Answering Correctly	Kind of Item	Type
11	41.3	RT	II
20	45.8	P	II
16	51.7	PT	II
18	53.3	PT	II
8	53.5	P	II
14	56.5	RT	I
6	58.5	RT	I
7	60.8	YT	II
10	65.3	P	I
19	67.0	R	I
2	70.7	R	I
4	72.7	YT	I
13	72.7	Y	I
17	74.0	R	I
9	75.3	YT	I
15	75.7	Y	I
12	75.8	R	I
5	77.2	P	I
3	77.5	Y	I
1	92.5	Y	I

The distinction between T and non-T items does not quite overlap with that between Type I and Type II items. Four of the Type I items are T, and two of the Type II items are non-T. That the time aspect does express itself in the correlational structure can best be seen in the full three-space. A gadget which has been devised to represent points in three dimensions is shown in the photograph in Figure 3.2.

In Figure 3.2, T-items are represented in white, other items in black. Shape differentiates between types R (hexagonals), Y (cubes), and P (balls). Except for one item, at the bottom of the figure, T-items tend to form a separate region in the three space: These items cluster around an obliquely inclined plane.

The Smallest Space Analysis, then, shows the content of the arithmetic items to be reflected in the correlational structure.

Figure 3.2. The three-space resulting from Smallest Space Analysis I representing interrelationships between arithmetic items. (see next page)



4. PROFILE ANALYSES OF TEST ITEMS

4.1. The MSA-I Computer Program

Smallest space analysis is designed to reveal the structure of interrelationships between variables. Data on individual subjects are taken into account only indirectly, via the coefficients of correlation that they generate. Now, it is possible to ask a different kind of question, namely, what are the profiles of individual subjects in respect of these variables, and how are these profiles related to each other. Here another type of analysis is required, which will be described below.

Until recently only one type of profile analysis has been in general use. Scalogram analysis[(1)] has frequently been employed to investigate whether the profiles of individual subjects form a particular kind of uni-dimensional structure. For data which do not render such a scale, an appropriate technique of analysis has been developed only recently by the principal investigator. The program, which is called Guttman-Lingoes Multidimensional Scalogram Analysis I (G-L MSA-I),[(14)] is now operational on the electronic computers of the University of Michigan and of the Hebrew University. One additional program in the MSA series has become operational at the time of writing, and still others are being developed. Since extensive use was made of the MSA-I in the present project, it is necessary to give here at least an intuitively comprehensible account of this technique.

Section 4.1

The MSA renders a space in which subjects are represented as <u>points,</u> variables as <u>partitions,</u> and categories of the variables as <u>regions</u> of the partitions. The program calculates coordinates for each point in a space with the smallest possible number of dimensions.

Consider the scale of Table 4.1.

<u>Table 4.1</u>

<u>A Perfect Scale for Four Dichotomous Variables</u>

	Variables			
Subject	I	II	III	IV
1	+	+	+	+
2	+	+	+	-
3	+	+	-	-
4	+	-	-	-
5	-	-	-	-

The five subjects may be represented as five points along a straight line, and the four variables may be regarded as four partitions, each of which divides this one dimensional space into two contiguous regions; see Figure 4.1.

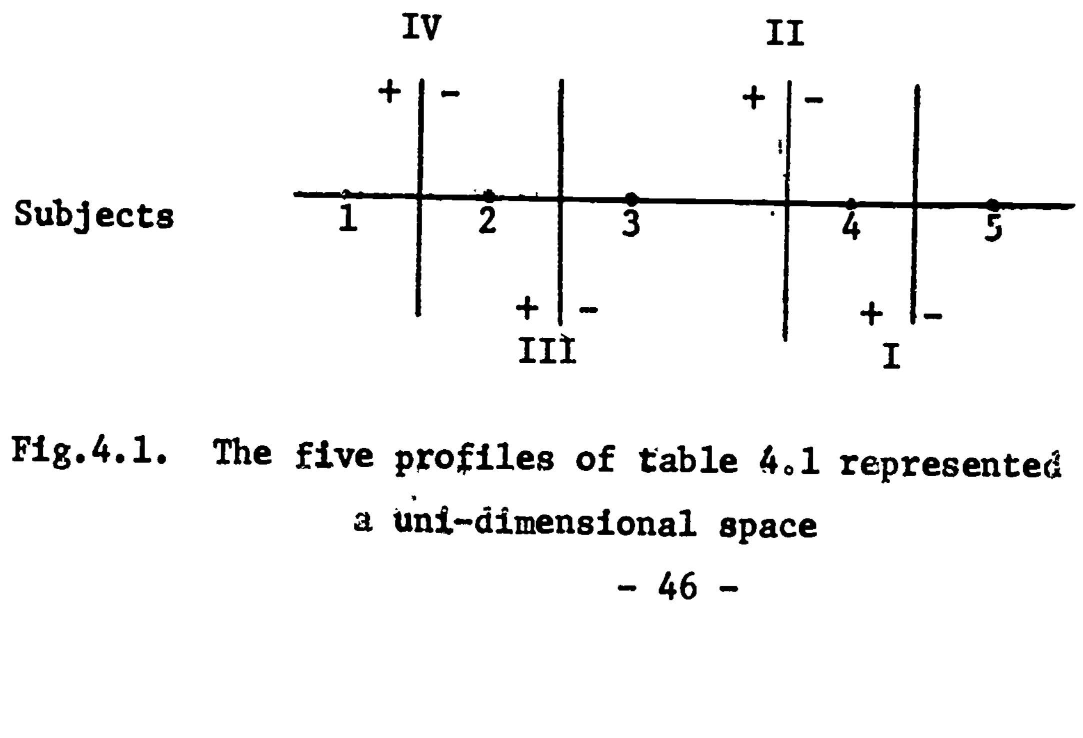


Fig.4.1. The five profiles of table 4.1 represented in a uni-dimensional space

Suppose now that still other profiles occur in the data. These will not form a scale with the profiles of Table 4.1. Hence more than one dimension will be required to represent all profiles together. A very simple example of a two-dimensional representation is given in Figure 4.2 which shows the profiles of Table 4.2.

Table 4.2

Eight Profiles of Four Dichotomous Variables, which do not form a Scale

	Variables			
Subject	I	II	III	IV
1	+	+	+	+
2	+	+	-	+
3	-	+	+	+
4	-	+	+	-
5	+	+	-	-
6	-	+	-	-
7	-	-	+	-
8	-	-	-	-

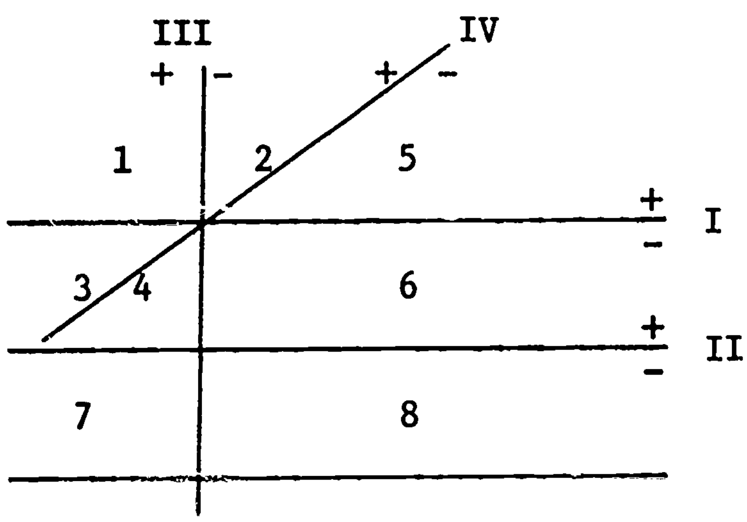


Fig.4.2 - The eight profiles of Table 4.2 represented in a two-dimensional space.

Inspection of Figure 4.2 shows the following types of relationships between partitions.

(a) Partitions with parallel lines indicate that the variables represented by them form a scale. (See also Figure 4.1). Thus, in Figure 4.2, I and II form a primitive scale; only the following subprofiles occur:

I+, II+ (Subject 5)

I-, II+ (Subject 6)

I-, II- (Subject 7)

The fourth combination (I+, II-) does not occur.

(b) Partitions with orthogonal lines indicate that the variables in question are unrelated in the sense that there is no constraint on the occurrence of profiles. In Figure 4.2, III is orthogonal to I and to II. As shown in this figure, each of the above three subprofiles of I and II occur both with III+ and III- (cf. also Table 4.2).

(c) Partitions related in a different manner. For instance, in Figure 4.2, IV cuts I and II at an oblique angle. This means that the three variables neither form a perfect scale nor are independent of each other. Thus, of the three subprofiles of I and II,

I+, II+ occurs with both IV+ and IV-

I-, II+ occurs with both IV+ and IV-

but I-, II- occurs only with IV-

It is easy to verify from Figure 4.2 and Table 4.2 that the relationship of IV to other subprofiles is a similar one.

When several lines cross each other at the same point in a space diagram, it will generally be the case that the relationship is closer between those variables whose lines are closer to each other. In Figure 4.3, for example, variable I will be closer to II than to III, and to III than to IV.

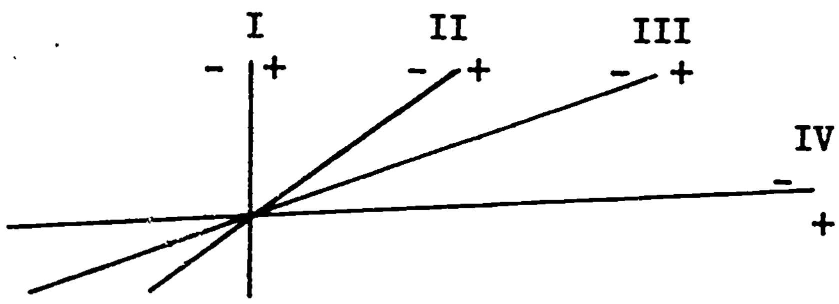


Figure 4.3. Schematized two-space diagram of five dichotomous variables.

The MSA-I will be a useful tool for describing typologies when there are a great number of variables and of profiles; the samples given above are indeed so simple that they can be worked out by hand. In practice,the data will usually reveal some deviation from the n-dimensional representation given by the MSA-I. The degree of deviation is indicated by the <u>coefficient of contiguity</u>, which may vary in principle from 0 (no fit) to 1 (perfect fit).

The investigator who is faced with a space diagram (which is printed out by the computer) is sometimes left with some freedom in deciding exactly where to draw the boundary lines, especially where there is no dense collection of points in the space. There is always the problem of interpreting the space, and for this an <u>a priori</u> facet theory of content is useful. Even in the absence of a sharp content theory, MSA-I is a powerful tool for testing certain kinds of hypotheses concerning typologies and their relationships to each other. When there is no theory on which such hypotheses can be based beforehand, the MSA is suggestive of new hypotheses and further kinds of analyses.

4.2. Analytical Ability Tests

Multidimensional scalogram analyses (MSA-I) were carried out for many of the analytical ability subtests and it was usually found that a great deal of constraint operated within the profiles. It was also found that only part of the possible profiles appeared with any appreciable frequency.

Since the items within each subtest were not constructed on the basis of an a priori facet design, we could not fully interpret the results. In this section, therefore, we just present the results as they appear, firstly, in order to illustrate our methods of work, and secondly, to suggest the outlines of further investigation with more systematically constructed tests.

The results of the tests reported in subsections 4.2.1 - 4.2.8 were obtained from the sample of 600 subjects. The items in the tables and figures are presented (from left to right) in the order they appear in the test. The plus sign (+) stands for correct answers, the minus sign (-) the incorrect ones.

4.2.1. Subtest Sv.

A partial order of the profile of five items of this test was obtained. The notion of partial order requires some words of explanation.

For a test of five items there are $2^5 = 32$ possible profiles when answers are dichotomized into correct vs. incorrect:

+ + + + +

+ + + + -

+ + + - +

.

.

.

- - - - +

- - - - -

In our test, four of these profiles are not obtained even for one of the subjects in the sample. This shows that at least a minimum of structure obtains between these items. When frequencies of these profiles are taken into account, it becomes apparent that further constraints operate. The seven most frequent profiles are given in Table 4.3 with their frequencies. (Frequency of 13 was chosen as cut off point; the next profile in the list occurred only 9 times).

Table 4.3

The Seven Most Frequent Profiles in the Sv Subtest

Profile	Frequency
+ + + + -	267
+ + + + +	137
- - - - -	31
+ + - + -	31
+ + + - -	22
+ + - - -	17
- + - - -	13
	518

These profiles, which account for about 86 per cent of our sample, may be arranged in partial order, as in Figure 4.4.

```
              + + + + +
                  |
              + + + + -
            /           \
  + + - + -               + + + - -
            \           /
              + + - - -
                  |
              - + - - -
                  |
              - - - - -
```

Figure 4.4. Partial order of the seven most frequent profiles.

Note that any two profiles connected by a line in the figure differ from each other in one variable only. If either one of the two profiles in the third row from the top were omitted, we would actually have a Guttman-scale. Such a scale is unidimensional: all profiles can be ordered along straight lines. In our case, however, a two-dimensional picture is obtained by retaining both profiles in the third row; these are not comparable or orderable between themselves.

The order of difficulty of the five items is given in Table 4.4.

Table 4.4

The Five Items of Subtest Sv in Descending Order of Difficulty

Item No.	No. of subjects solving correctly
5	157
3	470
4	481
1	510
2	521

Note that the same order of difficulty emerges in Figure 4.4. Starting from the top, the first item to turn from + to - is Item No.5. (i.e. this is the most difficult item) and the last, Item No.2, is the easiest. In the third row the third and fourth item turn to - ; this row reflects the almost equal degree of difficulty of these two items.

However, given the above degree of difficulty, the order in Figure 4.4 does not follow necessarily. Even in the case that no structure obtains -- i.e. that all profiles occur with appreciable frequency -- items may differ in their degree of difficulty. It is the usual procedure to report on relative difficulty of items, but the above should emphasize that the structure of profiles contains important additional information. In the present case, the partial order was worked out by hand from the list of profiles. In more complex cases it may be necessary to submit the data to the computer. Amongst our subtests there is an example showing that partial order does not invariably occur. In subtest Pd, which contains four items, all 16 possible profiles occur with a frequency of 10 or above. Other tests show partial order of different types involving varying numbers of profiles. This will be discussed in the following.

4.2.2. Subtest Ap

This subtest consists of four items. Of the $2^4 = 16$ profiles, only 6 appear with a frequency of more than 10. If two profiles which have a frequency of 9 are also included, a partial order obtains which holds for about 97 per cent of the subjects tested. This partial order is shown in Figure 4.5.

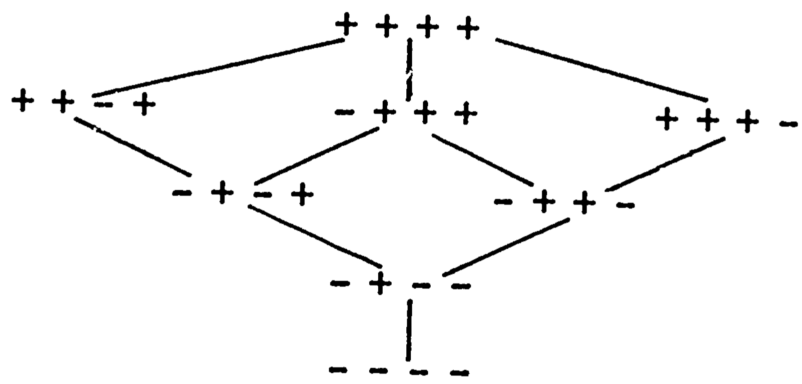


Figure 4.5. Profiles of subtest Ap

This figure should be compared with the previous one in which only seven profiles were included. (The previous figure, however, described a smaller proportion of the sample).

The relative difficulty of the four items is, in descending order: 1, 3, 4, 2.

4.2.3. Subtest Av_1

This subtest has also four items. Eleven of the 16 profiles appeared with frequencies of over 10, and accounted for 96.5 per cent of the subjects. These profiles are shown in Figure 4.6.

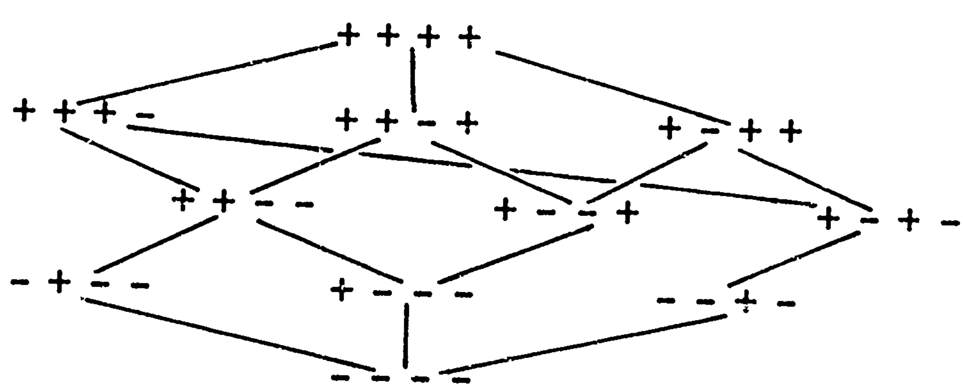


Figure 4.6. Profiles of subtest Av_1

In contrast to the partial orders obtained in the two subtests discussed above (see Figures 4.4 and 4.5), it is not possible to present the profiles of this subtest without crossing lines, which means that the configuration has more than two dimensions. Note that Figure 4.6 differs from the above two figures in still another respect. In the above two figures each profile of three incorrect answers (last line but one from bottom) can be derived from two of the profiles having two incorrect answers, by converting one of the pluses into a minus. In the present figure this is true only for one of the "three-minus" profiles, namely + - - -, and not for the other two.

The descending order of difficulty of the items in this subtest is: 4, 3, 2, 1.

The five profiles which appeared with a frequency of less than 10 are:

- + + +

- + + -

- - + +

- + - +

- - - +

The first four of these profiles have in common (a) that Item No. 1 -- the least difficult item -- is incorrect, and (b) that they are profiles of higher scoring subjects who get only one or two incorrect answers. This is to be expected: it is rare for a subject to score high and give an incorrect answer to the easiest item. The fifth profile is of low scoring subjects -- only one item is answered correctly, but this item is the most difficult one, No.4. Only two subjects of the 600 in our sample had this profile.

4.2.4. Subtests Pp and Pf

Each of these two subtests has two items. These four items will be treated as one subtest in the following discussion.

The 16 profiles present a similar picture to that of the Av_1 subtest discussed above. Again, 11 profiles have a frequency of above 10, and these can be presented only by crossing-over of lines as shown in Figure 4.7.

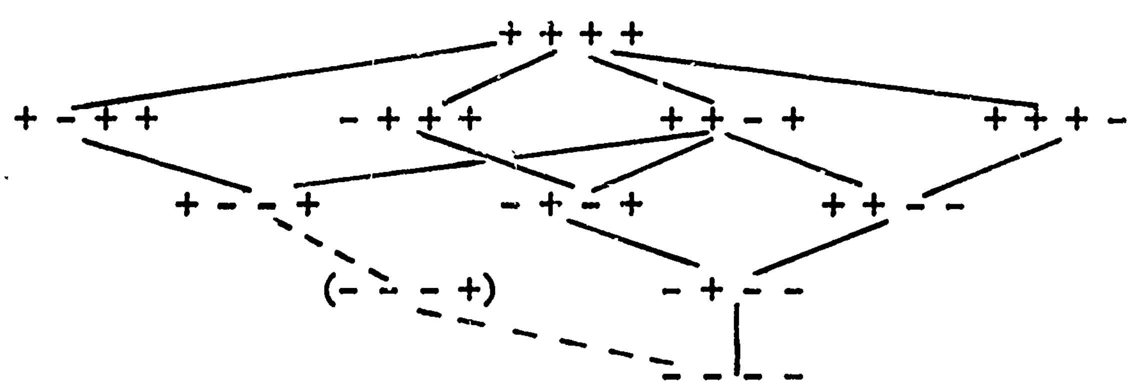


Figure 4.7. Profiles of subtests Pp and Pf (combined).

The two profiles shown at the left of the figure (+ - + + and + - - +) are responsible for the deviation from the two-space. Profiles obtained by turning one of the pluses of + - - + into a minus occurs with a frequency of less than 10. The above two profiles (+ - + + and + - - + are also deviant in that they do not reflect the order of difficulty of items. Item 2 was, in fact, the easiest to answer (with Items 1,4, and 3 following, in this order), but in these profiles Item 2 has a minus, although they belong to higher scoring subjects (1 and 2 items incorrect).

Such a deviation from an expected pattern should lead the investigators to a reexamination of the test items, so as to determine its cause. In the present case, the reason for the deviation is not far to seek. A technical mistake seems to have been made in the construction of Item 2. In the progression tests, the pictures (digits, etc.) are ordered from right to left, which is the direction of Hebrew script, (and according to the experimental evidence available at present, this is also the preferred direction of scanning, at least of those native Hebrew speakers who do not read a European language). When the item in question (see example No.9 in appendix) is read in this direction, the sequence of pictures shows a man entering a car and driving off. The correct answer is distractor c. But when the sequence is reversed, the "story" is reversed as well, and the man arrives from a drive, leaves his car, and walks off. A subject looking at the item in this manner will respond by choosing distractor b. It is plausible to assume, then, that the deviant profiles + - + + and + - - + are largely due to a group of subjects who merely made a mistake in the direction of reading the test item.

A Multidimensional Scalogram Analysis which included both the present progression tests and the digit test, Pd, reveals that these two subtests form a kind of primitive scale. It was rare for a subject who failed on the pictorial or figural progression test items, to answer correctly the digit series items, whereas the opposite occurred quite frequently -- namely, the subject might be able to solve a pictorial item and yet fail on an item involving digits. This finding seems to suggest that the digit progression test requires the ability to master the logical relations inherent in a series, and in addition, the ability to detect this relation within a series of digits. By contrast, the pictorial medium does not seem to pose any such special requirement.

4.2.5. Subtest Df

This subtest comprises six items, and there were therefore 64 possible profiles. Of these, only 31 profiles occurred in our sample of 600 subjects and only 7 with a frequency of above 10. There is, then, a great degree of constraint in this subtest.

Since the profiles having a frequency of above 10 accounted for less than 90 per cent of the sample, the next three profiles on the list (having frequencies of 9, 8 and 7) were also included in the following analysis. The 10 profiles shown in Figure 4.8 account for about 92 per cent of our sample.

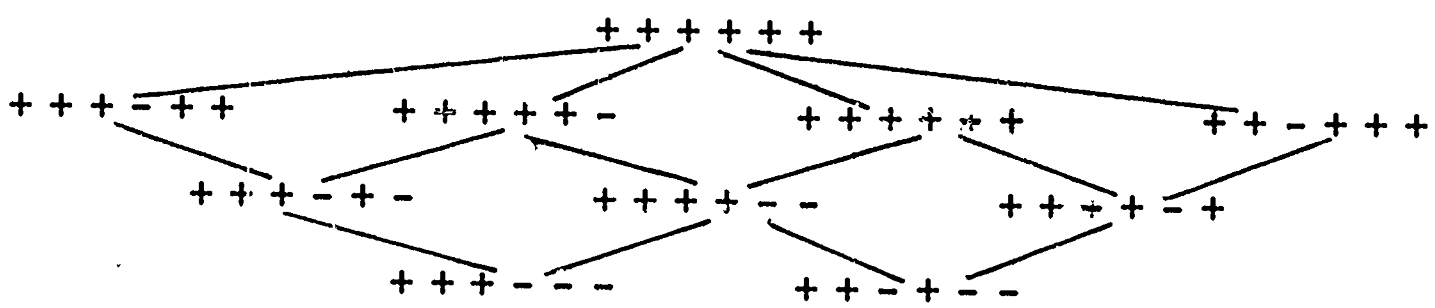


Figure 4.8. Profiles of subtest Df

4.2.6. Subtests Sp_1 and Sp_2

Several analyses were carried out on the two pictorial progressions tests of our battery, Sp_1 and Sp_2. On the basis of a preliminary analysis it was decided to combine the items of the two tests, discarding two of the items. Of the 32 possible profiles for the five remaining items, 13 profiles have a frequency of 10 or more and accounted for 86.3 per cent of the subjects. The partial order is shown in Figure 4.9.

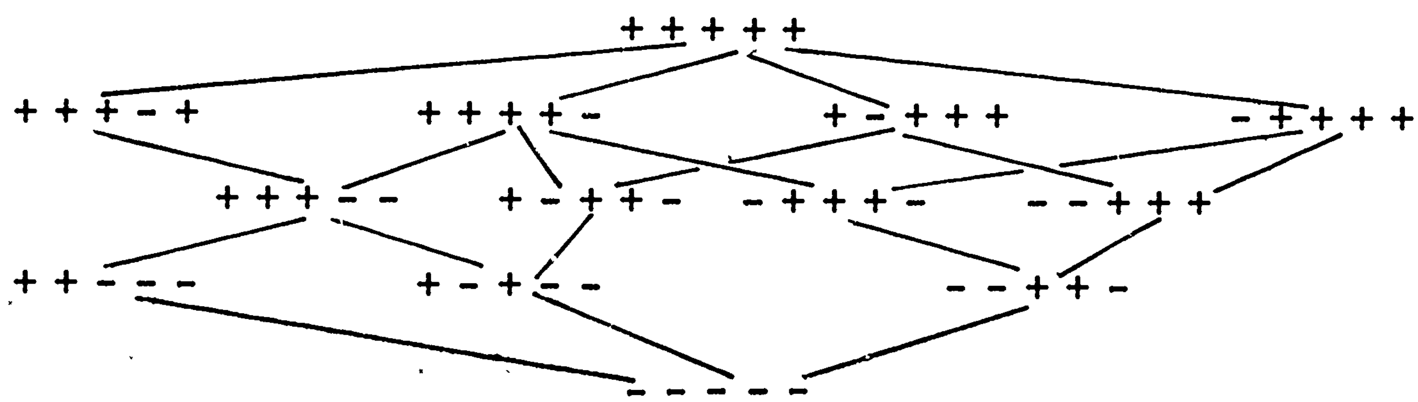


Figure 4.9. Profile of subtests Sp_1 and Sp_2

In the analyses of previous subtests it was tried to account for at least 90 per cent of the subjects. To do this for the Sp subtest, five additional profiles must be included:

- + + - -	frequency 9
+ + - + +	frequency 9
- - + - -	frequency 7
+ - - + -	frequency 7
+ + - - +	frequency 7

These profiles cannot be conveniently included in the above figure.

4.2.7. Subtest Cf

Of the 16 possible profiles of this four-item subtest, nine appeared with a frequency of 10 or more, and accounted for 92.7 per cent of the subjects. Figure 4.10 shows the partial order of these nine profiles.

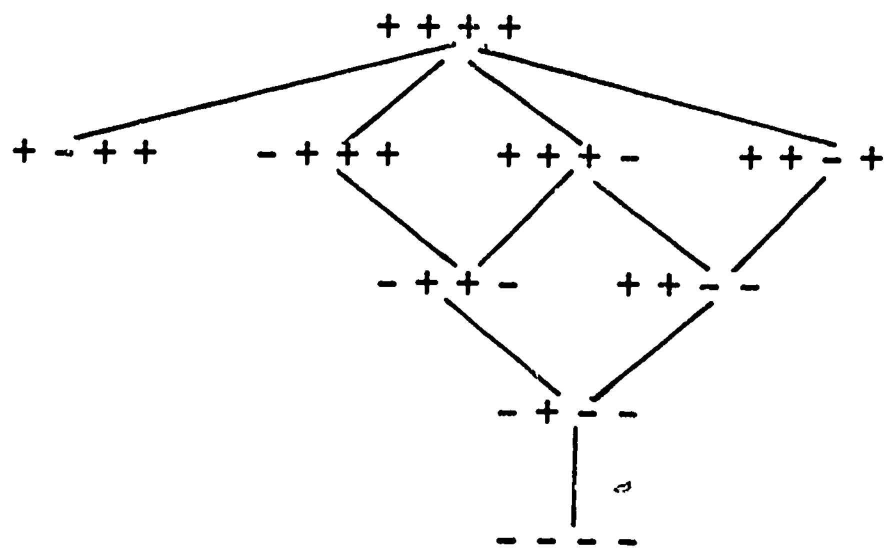


Figure 4.10. Profiles of subtest Cf

Note the "outsider" profile at the left; cf. section 4.2.4.

4.2.8. Subtest Cp

Table 4.5 shows the obtained profiles in descending order of frequency. All eight possible profiles appear with frequencies of 10 or above. Note, however, that the four most frequent profiles form a primitive scale. But there are 89 cases, or about 15 per cent of our sample, which have other profiles.

Table 4.5

Profiles of Subtest Cp

Profile	Frequency
+ + +	267
+ + -	69
- - -	75
+ - -	70
+ - +	35
- + +	29
- + -	13
- - +	12

4.3. Arithmetic Test

Considering each of the arithmetic subtests as a unit, the relative difficulty of items appears to be partly predictable. In general

(a) T-items are more difficult than non-T items

(b) Items with simple or decimal fractions are more difficult than those involving only whole numbers (see Section 3.1)

Tables 4.6, 4.7 and 4.8 show the order of difficulty of the items in the three subtests. The notation a, b, c, d, e, f, g will be used in the following to indicate the relative difficulty of items.

Table 4.6

Y Items in Ascending Order of Difficulty

Item No.	Notation	Percent Answering Correctly	T	Simple Fraction	Decimal Fraction
1	a	92.5	-	-	-
3	b	77.5	-	-	-
15	c	75.7	-	-	+
9	d	75.3	+	-	-
13	e	72.7	-	-	+
4	f	72.7	+	-	+
7	g	60.8	+	+	÷

Table 4.7

R Items in Ascending Order of Difficulty

Item No.	Notation	Percent Answering Correctly	T	Simple Fraction	Decimal Fraction
12	a	75.8	-	-	-
17	b	74.0	-	-	-
2	c	70.7	-	-	+
19	d	67.0	-	+	-
6	e	58.5	+	-	-
14	f	56.5	+	-	-
11	g	41.3	+	+	+

Table 4.8

P Items in Ascending Order of Difficulty

Item No.	Notation	Percent Answering Correctly	T	Simple Fraction	Decimal Fraction
5	a	77.2	-	-	-
10	b	65.3	-	-	-
8	c	53.5	-	+	+
18	d	53.3	+	+	-
16	e	51.7	+	+	-
20	f	45.8	-	+ *	-

* In this item there appear two simple fractions.

This order of difficulty is to be expected. An operation involving fractions, as well as an additional operation (time) present additional opportunity for the subject to make an error. It may be the case, however, that the differential difficulty of items also reflects stages of achieving mastery in the area in question. To determine whether this is so, a profile analysis was carried out. In Section 2.4 the problem of stages was approached through an analysis of the scores of various subtests, whereas in the present section the analysis is deepened through a profile analysis of individual items.

In one respect our approach to the profile analysis of the arithmetic test differed from the approach to the analysis of the analytical ability test (Section 4.2). In the former there was an hypothesis as to stages of acquisition, whereas in the latter we had no hypothesis to guide us. Specifically, we might predict that:

(a) Within each subtest, non-T items must be mastered before the same formula can be correctly applied by the subject in solving T-items that require an operation involving time.

(b) Subjects must first master items without fractions before they can correctly solve items that involve such fractions.

The profile analysis confirms this hypothesis. It showed, further, that the above two factors (fractions and T-items) are independent of each other as far as occurrence of profiles is concerned. Thus, there are subjects who first master the T-item, and are still unable to solve correctly items involving fractions; and, conversely, there are some subjects who have learned to solve items in which fractions appear, whereas they are still unable to solve T-items. The stages of acquisition, then, may be described as in Figure 4.11.

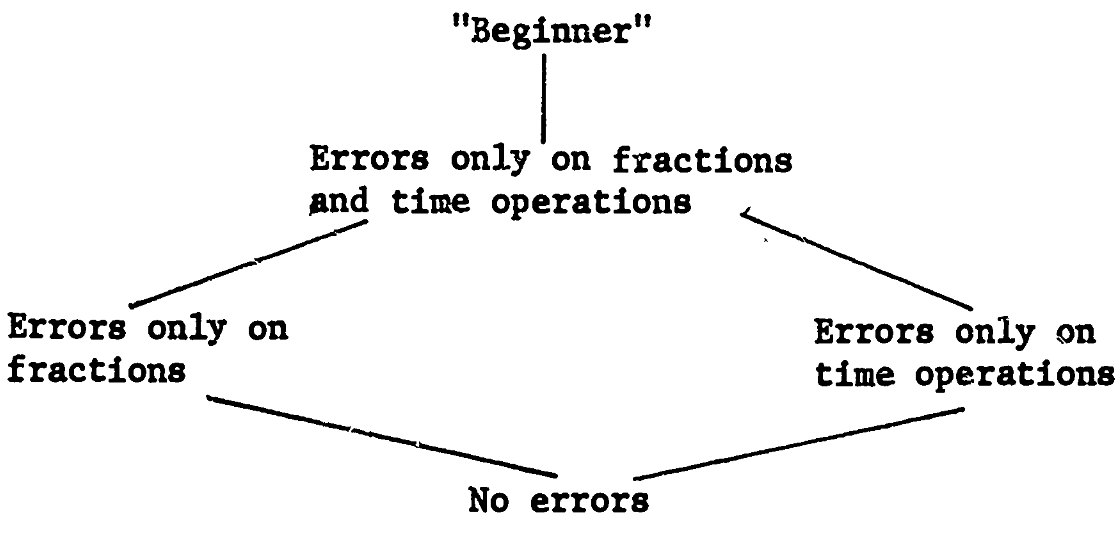


Figure 4.11. Schematized stages of acquisition in arithmetic subtests

Section 4.3

Figure 4.12 shows the more frequent profiles of Y items (including YT items). All profiles having a frequency of 10 or more are included in the figure, and in addition, some less frequent profiles are shown which fit into the general scheme of Figure 4.11. The frequency of each profile is given in the right hand column of the figure. The 15 profiles of Figure 4.12 describe the results of 418 subjects out of the sample of 600.

While a considerable number of subjects remain unaccounted for, the small frequency of their profiles (less than 10) suggest that the latter may be regarded as chance deviations from the profiles in Figure 4.12.

Figure 4.13 shows the stages for R items. Again, all profiles of frequency 10 and above are included, as well as some of the less frequent ones that fit the general scheme. This figure also shows that decimal fractions are mastered before simple fractions, at least as far as the evidence of items contained in this subtest goes.

For Figures 4.12 and 4.13 see the next two pages.

In Figure 4.12, one deviant profile - b - occurred with a frequency of 13. This was not included in the figure.

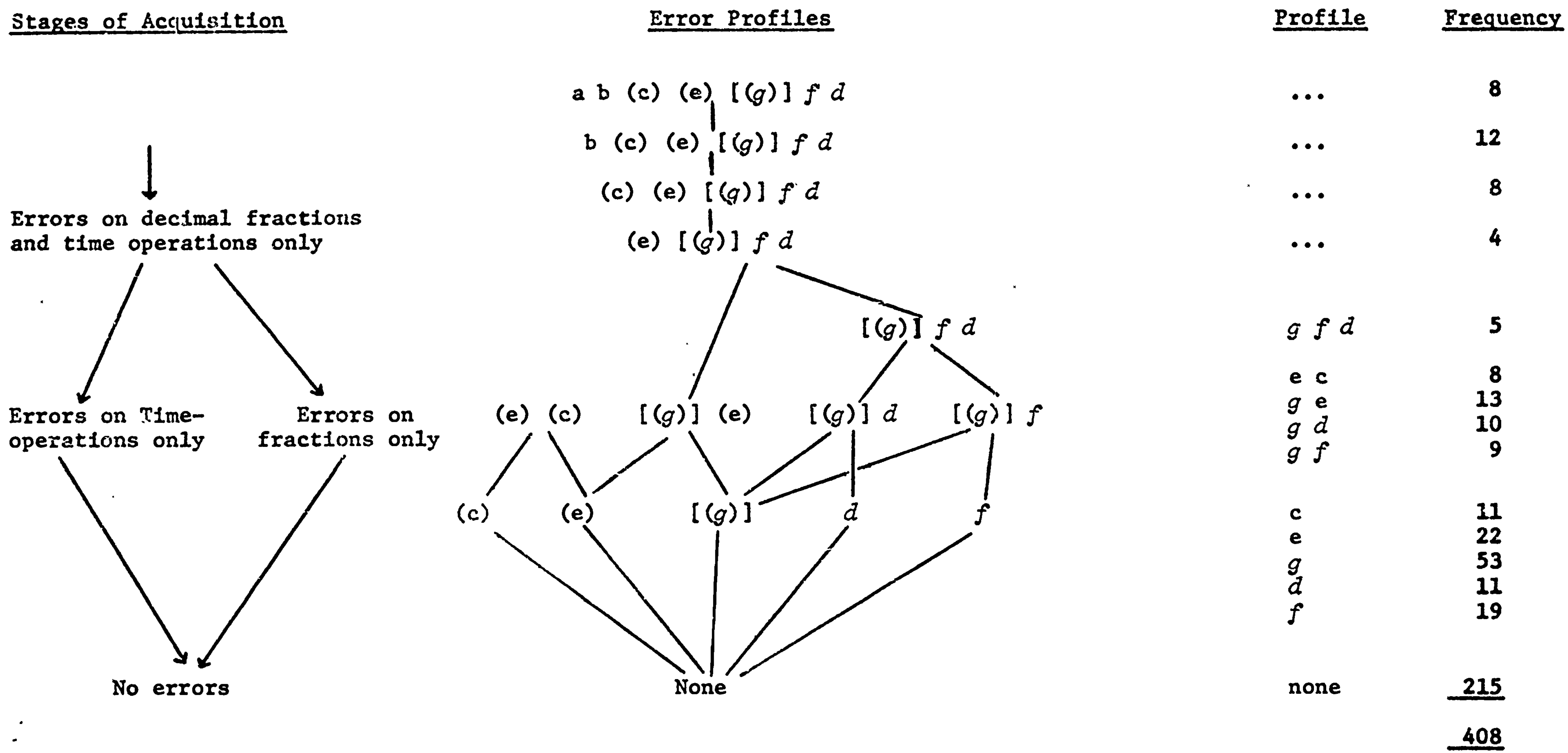


Figure 4.12. Stages of acquisition of Y and YT items

Key: Decimal fraction - ( )
Simple fraction - [ ]
T-item - *Italics*

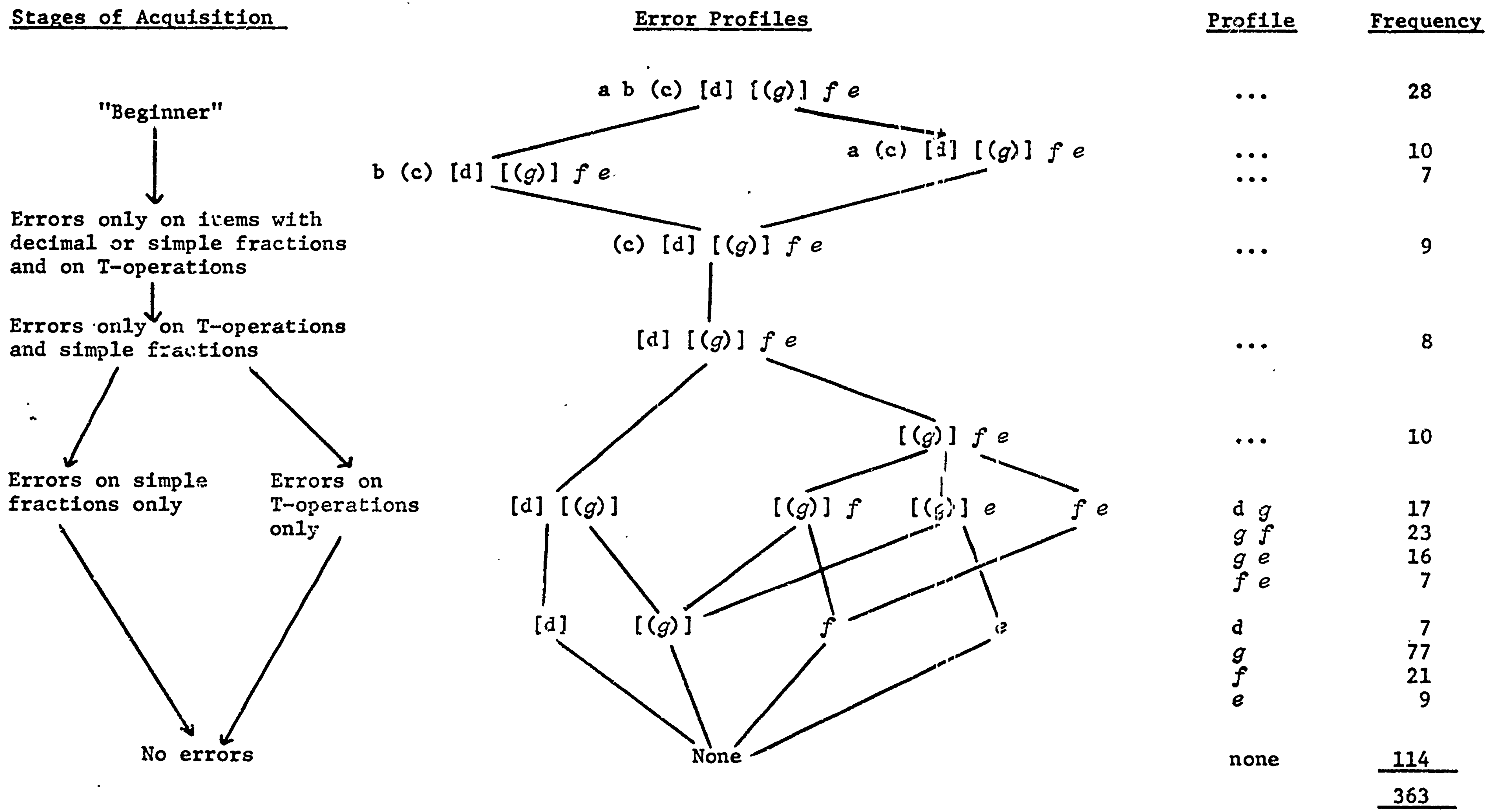


Figure 4.13. Stages of acquisition of R and RT items

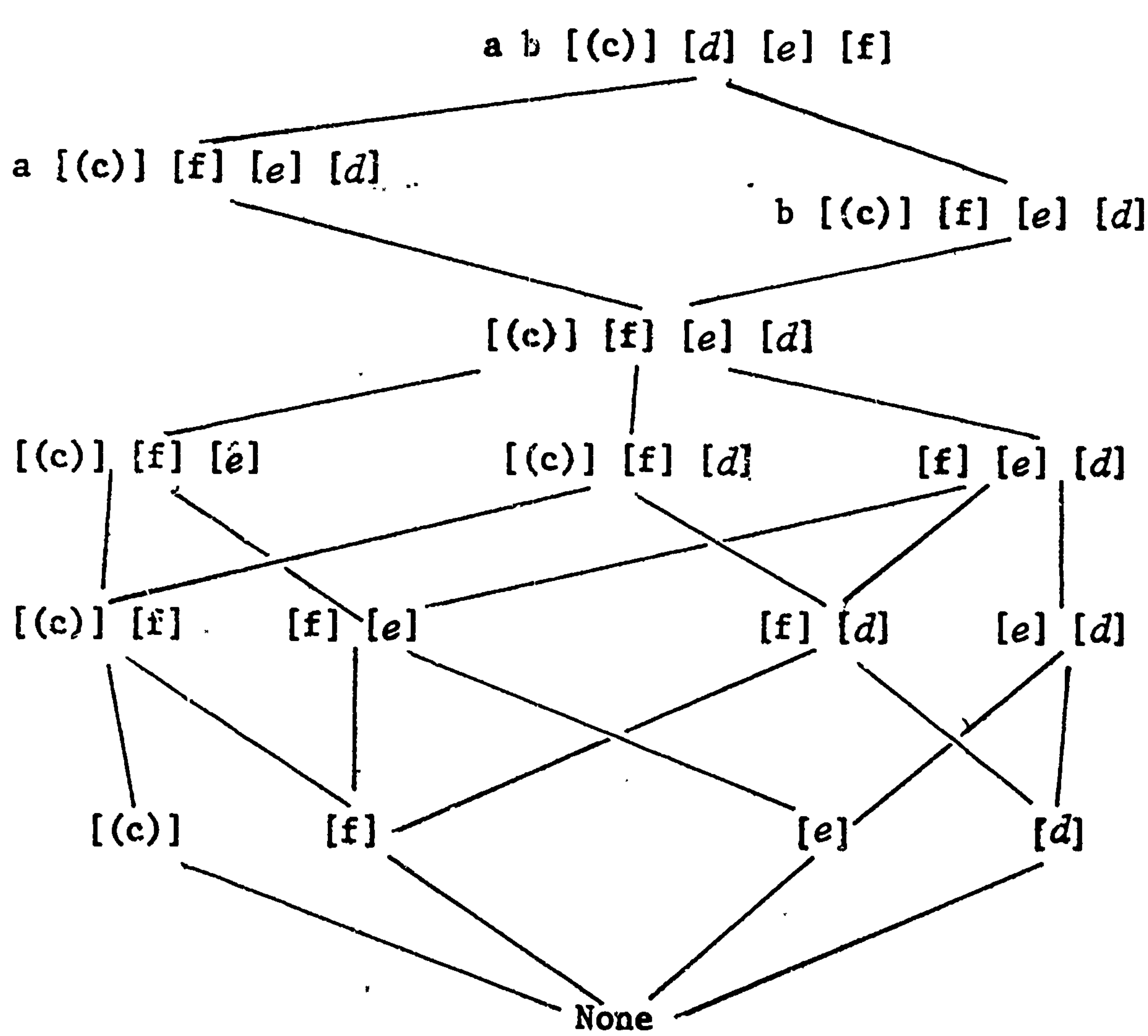
Key: Decimal fraction - ( )
Simple fraction - [ ]
T-item - *Italics*

In subtest P, the two T-items (e and d) also involve simple fractions. Hence this subtest does not lend itself to testing the prediction that some subjects make errors in T-items only and not on fractions. The profiles having a frequency of 10 or above are shown in Figure 4.14. As shown in this figure, T-items and items involving fractions will not be mastered before more simple items (a and b) have been learned. There is one exception to this, however. One profile appearing with a frequency of 11 not shown in the figure was

a b [(c)] [f] [*e*]

There are, thus, 11 subjects out of the 600 who answered correctly item d, which involved both a simple fraction and a time operation, while they made errors on both of the simpler items, a and b. Perhaps this is also a chance deviation from the expected pattern of responses.

Error Profiles



Profile	Frequency
...	43
...	10
...	32
...	19
c f *e*	15
c f *d*	10
f *e d*	25
c f	15
f *e*	10
f *d*	18
e d	14
c	22
f	32
e	21
d	22
none	111
	419

Figure 4.14. Error profiles of P and PT items.
See Key of Figure 4.13.

Section 4.3

In conclusion it may be said that our hypothesis regarding the stages of acquisition (Figure 4.11) is in general substantiated. It should be remembered that these data reflect the state of affairs in a nation-wide sample of eighth-graders at a particular point of time. Our results are thus influenced by the teaching method as practiced in a particular year in the eighth grade. By themselves, the data do not provide a basis for deciding whether the above stages are either necessary or desirable in teaching percentages. It may be the case, for instance, that fractions should be taught before T-items or vice versa. Our analysis merely describes the existing situation.

5. ANALYZING RELATIONSHIPS BETWEEN DISTRACTORS

5.1. The SSA-II Computer Program

For the purpose of these analyses we employed the Guttman-Lingoes SSA-II program[11]. This is an asymmetric analysis rendering the smallest possible Euclidean space which preserves ordinal information about distance. An information function, $P_{ab|ij}$, about distances was employed in this analysis, defined as follows:

Let $N_{ab|ij}$ = number of respondents who chose distractor a on item i and distractor b on item j

$N_{a|ij}$ = number of respondents who chose distractor a on item i and who answered incorrectly on item j

then $$P_{ab|ij} = \frac{N_{ab|ij}}{N_{a|ij}}$$

This is asymmetric in a and b.

An example of values of $P_{ab|ij}$ for three items is given in Table 5.1. Note that the columns in each submatrix add up to 100 per cent and that the matrix is asymmetric: for any two items, the submatrix above the diagonal generally differs from that below the diagonal.

$P_{ab|ij}$ is the probability of choosing distractor b on item j, given that distractor a was chosen on item i (and given that item j was answered incorrectly). The sum of these probabilities over all distractors (wrong answers) in item j is, of course, 1.00.

Table 5.1

Example of Joint Distributions of Answers to Four Distractors in Several Items, in Percentages

Item		i				j					k			
	Distractor	a	b	c	d	a	b	c	d	....	a	b	c	d
i	a	100				40	27	34	16		34	11	32	22
	b		100			26	41	31	16		31	38	22	16
	c			100		14	17	10	15		07	10	19	10
	d				100	20	15	25	53		28	41	27	52
	Total	100	100	100	100	100	100	100	100		100	100	100	100
j	a	23	15	16	09	100					19	10	12	06
	b	18	28	23	08		100				11	21	13	04
	c	36	32	19	21			100			34	26	35	20
	d	23	25	42	62				100		36	43	40	70
	Total	100	100	100	100	100	100	100	100		100	100	100	100
	.													
	.													
	.													
	.													
k	a	30	29	13	19	39	25	29	20		100			
	b	06	21	13	17	11	26	13	14			100		
	c	49	37	58	33	39	42	43	32				100	
	d	15	13	16	31	11	07	15	34					100
	Total	100	100	100	100	100	100	100	100		100	100	100	100

The distance between distractors may be specified to be inversely related to the $P_{ab|ij}$ values, comparing categories only within each j separately. Thus, if

$$P_{ab|ij} > P_{ac|ij},$$

then

$$D_{ab|ij} < D_{ac|ij},$$

where $D_{ab|ij}$ is the distance between distractor a in item i and distractor b in item j.

Entering the first submatrix of the second row of submatrices in Table 5.1 we see that

$$P_{aa|ij} = .23$$

$$P_{ab|ij} = .18$$

$$P_{ac|ij} = .36$$

$$P_{ad|ij} = \underline{.23}$$

$$\underline{1.00}$$

An alternative specification is to require distances to be inversely related to <u>likelihoods</u>, or to compare the $P_{ab|ij}$ within each i rather than within j. To obtain the values for likelihood, Table 5.1 must be read by rows instead of by columns. Entering the second submatrix of the first row of matrices, we obtain the following likelihood values:

$$P_{aa|ji} = .40$$

$$P_{ba|ji} = .27$$

$$P_{ca|ji} = .34$$

$$P_{da|ji} = \underline{.16}$$

$$\underline{1.17}$$

Note that these likelihood values do not have to add up to 1.00 and that they differ from the corresponding probability values.

The likelihood and probability measures may result in a different rank order of distances. In the above example

$$P_{ba|ji} > P_{da|ji}$$

but

$$P_{ab|ij} < P_{ad|ij}$$

In other words, according to the likelihood measure, distractor a in item i is closer to distractor b in item j than to distractor d in item j, whereas according to the probability measure, the reverse is the case. An analysis by probability may therefore render a different picture than one by likelihoods. In our study, analyses were carried out for both distance measures.

An example of the use of probabilities in an SSA-II can be found in Laumann and Guttman,[10] and of likelihood in Guttman.[4]

Likelihoods are usually less dependent than conditional probabilities on the marginal (or unconditional) distributions of the items, and may often be expected to give a smaller space. For our present data, similar two-dimensional spaces were derived from the two kinds of specifications, the likelihoods generally giving a more interpretable picture.

5.2. Distances Between Arithmetic Test Distractors

In achievement tests, classification according to type of error may be utilized for diagnostic purposes; cf. also (6), where some of the following results are presented.

The following is an example taken from our test.

The arithmetic test included 16 questions pertaining to percentages, the following being a typical item:

475 kgs. of sugar were delivered to a grocery store; 48 per cent of the sugar was sold on the first day. How many kgs. of sugar were sold on that day?

(1) 475

(2) 218

(3) 989

(4) 228

(5) other

The correct answer is (4). Distractor (3) will be chosen by respondents who employ the wrong formula in solving the question: $\frac{100\ (475)}{48}$. Choice of distractor (1) obviously results from copying one of the numbers appearing in the question itself. Distractor (2) displays a number which is close to the correct answer. The design of distractors in this test employed the following types:

(a) application of wrong formula;

(b) copying a number appearing in the question;

(c) a number close to the correct answer;

(d) "other".

(Some items of the test actually included a "miscellaneous" category; but ideally all distractors would conform to the same pattern).

Table 5.2

Arithmetic Achievement Distractors

Coefficients of Alienation

Items	Distractors	No. of dimensions Two	Three
Probability			
all items	a, c, b	.318	.237
T-items	a, b, c, d and miscellaneous	.306	.225
non-T-items	a, b, c, d and miscellaneous	.298	.221
non-T-items	a, b, c, d	.301	.225
non-T-items	a, b, c	.278	.205
Likelihoods			
all items	a, b, c	.321	.245
T-items	a, b, c, d and miscellaneous	.262	.208
non-T-items	a, b, c, d, and miscellaneous	-	.205
non-T-items	a, b, c, d	.303	.223
non-T-items	a, b, c	.276	.204

Owing to the limitation of capacity, the SSA-II program operative on the Hebrew University 7040, can only handle up to 55 variables. Since each of the 20 items of our arithmetic test has four incorrect answers, it was not possible to submit all items to the analysis. Separate analyses were therefore carried out as follows:

1) T-items (see Chapter 3) were analyzed separately from non-T-items.
2) All 20 items were submitted to the SSA-II, but only three types of distractors were included: a, b, and c (the miscellaneous and "other" types were disregarded in this analysis).
3) These three types of distractors were also analyzed for T-items and non-T-items separately.
4) Another subset of distractors, which excluded only the "other" category was submitted to analysis separately for the T-items and the non-T-items.

All analyses were carried out for both probabilities and likelihoods. While there was no appreciable difference in goodness-of-fit, the space diagrams for likelihoods were in general more easily interpretable. All in all, 10 smallest space analyses have been carried out, each for two and for three dimensions. These are shown in Table 5.2.

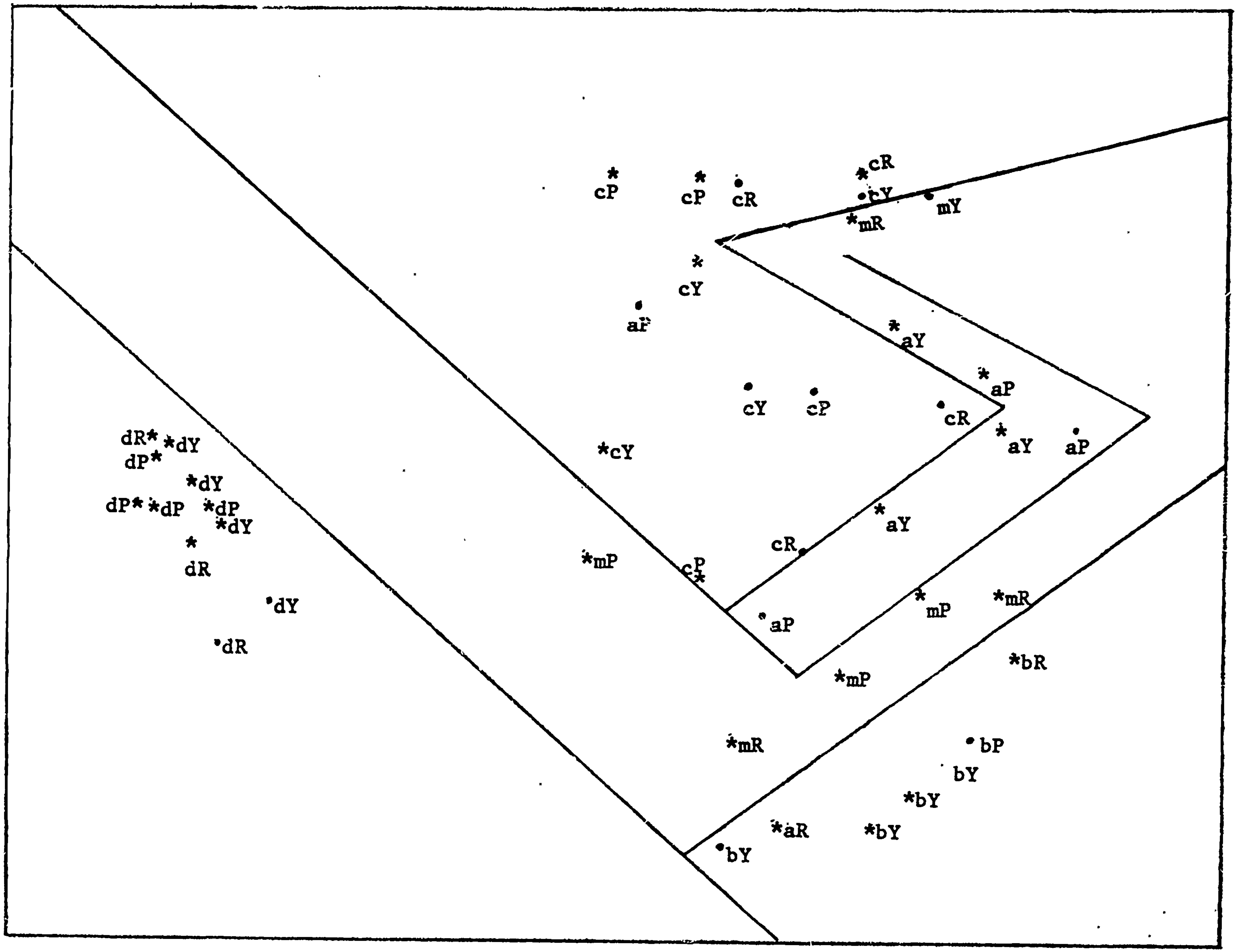
As shown in the table, coefficients of alienation were rather high. For two dimensions, the range was from .262 to .321; and for three dimensions, from .204 to .245. In order to find out whether they require further dimensions one set of data (all items, distractors a, b, and c) was also run on four dimensions. However, the coefficient of alienation decreased only slightly: from .237 for three dimensions to .192 for four. This contrasts with the appreciable decrease of .318 for two dimensions to .237 for three. It appears, therefore, that further dimensionality does little to increase the goodness-of-fit. The discussion of results which follows pertains to the three-dimensional space diagram.

Figure 5.1 shows the first two dimensions of the three-dimensional space obtained by the SSA-II. The eleven items represented here fall into three types according to whether the subject is required to find the principal (P), the rate (R), or the yield (Y); compare the above example. The figure shows five regions of distractors: distractors of types a, b, c and d, and miscellaneous distractors (m). Points which are the higher ones on the third dimension are indicated by asterisks. This information clarifies the fact that the regions are indeed distinct in the total three-space, since neighboring points in two regions are usually distant from each other on the third dimension. There are only a few distractors which do not fall within the appropriate region.

It is possible, therefore, to assign a score for each type of error. A student's profile of error scores will tell not only how much he has been achieving in a given area but also what are the typical types of errors he makes. Thus the diagnostic effectiveness of such a test is increased.

Distractor analysis of two of the analytical ability subtests, Cf and Cp, were also carried out. These subtests differed from the arithmetic tests in that they were not systematically constructed. The analysis again showed trends towards regions, but the results were not strong enough to warrant further discussion in the present report.

Figure 5.1. The first two dimensions of the three-dimensional space obtained by SSA-II for distractors of eleven arithmetic-test questions. Questions are on principal (P), rate (R), or yield (Y). Points relatively high on the third dimension are indicated by an asterisk. (see next page)


cP
cP
cR
cR
cY
mY
mR
cY
aP
aY
aP
cY
cP
cR
dR
dY
dP
cY
aY
aP
dY
dP
dP
dP
dY
aY
dR
cR
mP
cP
dY
aP
mP
mR
dR
bR
mP
mR
bP
bY
bY
aR
bY
bY

5.3. Profiles of Arithmetic Test Distractors

In the previous section it has been shown that the types of distractors which were built into the test items are psychologically effective. Our SSA-II has revealed the relationships between distractors which suggested a typology of subjects. A direct analysis of such a typology was carried out next, employing the MSA-I.

The results of this analysis show that profiles of the frequencies of distractors of types a, b, and c, and of correct answers form a quasi-scale, whereas no clear configuration emerges (one-dimensional or otherwise) when distractors of type d and miscellaneous distractors are also included. In the following, the analysis is described in more detail.

Each subject was assigned a score on each type of distractor, representing the frequency with which this type was chosen in the 20 test items of the arithmetic test. It was found necessary to group the number of choices into a smaller number of scores; otherwise the number of profiles would have grown formidably large. Table 5.3 shows how scores were assigned to the correct answer as well as to all the distractors according to the number of times each was chosen. In the final analysis type d distractors and miscellaneous distractors were disregarded. In assigning these scores we took into account the frequency distribution of number of choices.

Table 5.3

Categorization of No. of Choices for

Each Distractor and Correct Answer

No. of Choices (in 20 items)	Categories for MSA-I					
	Correct Answer	Distractor a	Distractor b	Distractor c	Distractor d	Miscellaneous Distractors
0	1	1	1	1	1	1
1	1	2	2	2	2	2
2	1	2	2	2	2	2
3	1	3	3	3	3	3
4	2	3	3	3	3	3
5	2	3	3	3	3	3
6	2	3	4	3	4	4
7	3	3	4	3	4	4
8	3	3	4	3	4	4
9	3	3	4	3	5	5
10	4	3	4	3	5	5
11	4	3	4	3	5	5
12	4	3	4	3	5	5
13	5	3	4	3	5	5
14	5	3	4	3	5	5
15	5	3	4	3	5	5
16	6	3	4	3	5	5
17	6	3	4	3	5	5
18	6	3	4	3	5	5
19	7	3	4	3	5	5
20	7	3	4	3	5	5

Table 5.4

Scale of Arithmetic Distractors Resulting from MSA-I

Point on Coordinate	Profile of Categories: Correct	Distractors: a	b	c	Frequency
183	7	1	1	1	30
173	7	2	1	1	17
148	4	2	1	1	6
137	6	1	1	1	41
120	5	1	1	1	10
117	6	2	1	1	6
101	4	1	1	1	4
097	5	2	1	1	5
084	5	2	1	2	3
053	6	2	1	2	7
051	5	1	1	2	3
039	6	2	1	1	16
033	5	1	2	1	6
027	5	2	2	1	4
018	4	1	2	1	3
008	6	1	1	2	9
006	3	1	2	1	4
003	6	1	2	1	6
001	4	2	2	1	8
008	3	2	1	2	4
010	3	2	2	1	5
015	5	1	3	1	3
017	5	3	1	1	3
023	5	1	2	2	7
033	5	2	3	1	3
042	4	1	3	1	4
050	4	1	2	2	3
060	4	2	3	1	3
068	3	2	3	1	3
069	4	2	2	2	7
078	3	2	2	2	8
116	4	3	2	1	8
129	4	2	3	2	4
131	2	2	2	1	3
137	3	2	3	2	4
155	5	3	2	2	3
174	4	3	3	1	3
183	4	3	2	2	5

015	5	1	3	1	3
017	5	3	1	1	3
023	5	1	2	2	7
033	5	2	3	1	3
042	4	1	3	1	4
050	4	1	2	2	3
060	4	2	3	1	3
068	3	2	3	1	3
069	4	2	2	2	7
078	3	2	2	2	8
116	4	3	2	1	8
129	4	2	3	2	4
131	2	2	2	1	3
137	3	2	3	2	4
155	5	3	2	2	3
174	4	3	3	1	3
183	4	3	2	2	5
200	2	2	2	2	2
241	2	1	3	2	3
252	2	3	2	1	2
255	3	3	3	2	5
256	2	2	3	2	5
306	2	3	3	1	2
314	2	3	2	2	2
368	2	3	3	2	2
442	1	3	2	2	4
488	2	2	3	3	2
493	1	3	3	2	3
599	2	3	4	2	4
601	2	3	3	3	6
707	2	2	4	3	3
818	1	1	4	3	2
950	1	3	4	3	3

The MSA-I resulted in a one-dimensional space with coefficient of contiguity .966 (by only 9 iterations). Table 5.4 presents the resulting order of profiles in this space, the frequency of these profiles in the sample of 600 subjects, and the point on the coordinate on which each profile appeared. In assigning points on the coordinates, the MSA-I attempts to minimize the average size of deviations, taking into account the frequency of each profile. Spaces between lines in the table indicate roughly the spaces between profiles along the coordinate.

The MSA results indicated that a further grouping of some of the scores would yield an even clearer picture. Indeed, a scale is arrived at with only a small number of deviations, as shown in Table 5.5.

Table 5.4 shows that all three types of distractors decrease in frequency as the "correct" score increases. As the subjects achieve mastery in this test, they first improve on distractors of types c, next on type b, and last on type a. Table 5.4, then, suggests certain stages of learning in this area. Table 5.5 presents the ideal types and their suggested interpretation.

In one respect these findings do not confirm our expectations. What we expected to find was a typology of errors which would be independent of the student's level of achievement. In other words, it was surmised that certain learning difficulties would be revealed that would serve as the basis for classifying the students and that this classification would cut across the classification by number of correct answers. The data, however, do not reveal such a picture: there are various error profiles, but, as shown in Table 5.5 a subject's profile is largely dependent on his "correct" score. What we found is that there are systematic patterns of errors, but these are dependent on the progress made by the subject.

Table 5.5

Stages of Achievement in Arithmetic

Correct	Categories of Ideal Type Distractor a	b	c	Interpretation
1	3	4	3	"Beginner"; very low total score
2	3	4	3	Rise in total score
2	3	4	2	Some improvement with regard to type c errors
2	3	2 or 3	2	Some improvement with regard to type b errors
3 or 4	1 or 2	2 or 3	1	Rise in total score; type c errors almost disappear; some improvement with regard to type a errors
5 or 6	1 or 2	2 or 3	1	Additional rise in total score
5 or 6	1 or 2	1	1	Type b errors almost disappear
7	1 or 2	1	1	

6. SUMMARY AND CONCLUSIONS

A relatively unexplored approach to achievement and ability testing has been taken in this study. Previously, not much attention has been paid to the kind of questions with which we set out to deal. It is suggested here that answers to these questions will be instrumental in increasing the diagnostic efficiency of educational tests.

The types of problems dealt with here required new methods of data analysis. Such methods have recently become available with the development of techniques of nonmetric multivariate data analysis. The computer programs employed in the present study -- the SSA-I, SSA-II, and MSA-I -- have been described in various places in this report (Sections 2.1, 4.1, 5.1) and, more formally, elsewhere.[8,11,12,13] These methods have made it possible to pose questions for which researchers hitherto did not possess efficient techniques of analysis. The efficiency of these methods has been tried out, probably for the first time, in the present project, and the result obtained with our data should serve to illustrate what types of questions a researcher may ask, what types of techniques of analysis he might employ to answer them, and what kind of results he may expect to get.

The structure of the interrelationships holding between a large battery of tests, as well as that holding for items of a subtest was investigated by a nonmetric Smallest Space Analysis: the Guttman-Lingoes SSA-I. In both cases, a space was obtained that was easily interpretable in terms of an _a priori_ definition by means of facets (Sections 2.2, 3.1). The results obtained for the test battery corroborated earlier findings on the relationships between different subtests of analytical ability and aptitude-achievement (Section 2.2). An additional analysis of data from a previously published study further strengthened our conclusions, and also served to compare the present method of analysis with the more conventional method of factor analysis. It was found that the SSA gave a more easily interpretable picture than factor analysis, and further obviated the need of looking for "meaningful coordinate axes". (Section 2.3).

But important as it is, correlational structure does not tell the whole story. In the planning of school curricula and of teaching methods, the question may be raised, which, if any, stages of development exist in a certain area. Until recently, techniques were available only for the investigation of unidimensional structures. In this study the G-L MSA-I was employed, which is suitable for the study of typologies and multi-dimensional structures. For the arithmetic test, stages of achievement were revealed, both by an analysis of subtests (Section 2.4) and by the analysis of individual test items (Section 4.3). For illustrative purposes, MSA was also applied to analytical ability subtests where no interpretation of the resulting structures is yet available (Section 4.2).

Yet another kind of analysis pertained to the distractors of test items. Distractors have so far been treated as step-children in test construction and analysis. In the analyses carried out previously, the students' answers were dichotomized into either correct or incorrect. By contrast, the present approach was based on the assumption that diagnostically important information will be revealed by asking not only _whether_ the subject gave an incorrect answer, but also _which_ incorrect answer he chose. In the arithmetic test, in which the distractors were systematically designed, our analyses have borne this out. A Smallest Space Analysis -- the G-L SSA-II -- showed that subjects who tend to make certain types of errors in one item tend to make the same kind of error on other items as well. (Section 5.2). Further, a multi-dimensional scalogram analysis resulted in a scale of profiles consisting of error scores on three types of distractors and on the score of correct answers. This suggests that the student learns to avoid certain types of errors in a sequence which coincides with achieving mastery within the area tested (Section 5.3).

It is hoped that this report will be suggestive of further research into problems of diagnostic effectiveness of school tests along the lines described here.

BIBLIOGRAPHY

1. Guttman L. 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.

2. Guttman, L. A faceted definition of intelligence. In: R. Eifermann (Ed.), Scripta Hierosolymitana, Vol.14, Studies in Psychology. Jerusalem, Israel: The Hebrew University, 1965. Pp.166-181.

3. Guttman, L. The structure of interrelations among intelligence tests. In: Proceedings of the 1964 Invitational Conference on Testing Problems. Princeton, New Jersey: Educational Testing Service, 1965.

4. Guttman, L. The Non-metric Breakthrough for the Behavioral Sciences. Invited address to the Automatic Data Processing Conference of the Information Processing Association of Israel, January, 1966.

5. Guttman, L., and Schlesinger, I.M. Development of diagnostic analytical and mechanical ability tests through facet design and analysis. Research Project No. OE-4-21-014. The Israel Institute of Applied Social Research, Jerusalem, Israel, 1966.

6. Guttman, L. and Schlesinger, I.M. Systematic construction of distractors for ability and achievement test items. Educational and Psychological Measurement, (in press).

7. Guttman, L. A general nonmetric technique for finding the smallest Euclidean space for a configuration of points. Psychometrika, (in press).

8. Guttman, L. Order Analysis of correlation matrices. In: R.B. Cattel (Ed.) Handbook of Multivariate Experimental Psychology, Rand McNally, (in press)

9. Hoger, D. Analyse der Intelligenzstruktur bei mannlichen Gymnasiasten der Klassen 6-9 (Untersekunda-Oberprima), Psychologische Forschung, 1964, 27, 419-474.

10. Laumann, E.O. and Guttman, L. The relative associational contiguity of occupations in an urban setting. American Sociological Review, 1966, 31, Pp.169-178.

11. Lingoes, J.C. An IBM 7090 program for Guttman-Lingoes Smallest Space Analysis II, Behavioral Science, 1965, 10, P.487.

12. Lingoes, J.C. An IBM 7090 program for Guttman-Lingoes Smallest Space Analysis - I. Behavioral Science, 1965, 10, Pp.183-184.

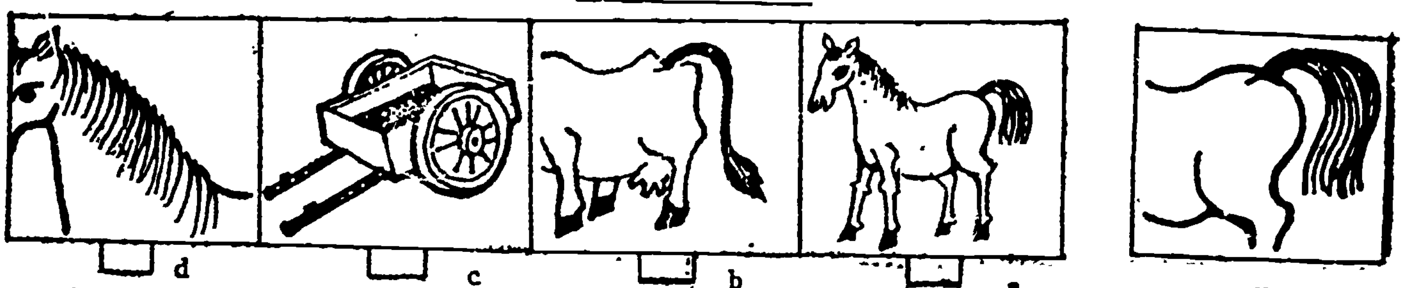
13. Lingoes, J.C. New computer developments in pattern analysis and non-metric techniques. I.B.M. Journal, (in press)

14. Lingoes, J.C. An IBM 7090 program for Guttman-Lingoes Multidimensional Scalogram Analysis - I. Behavioral Science, (in press).

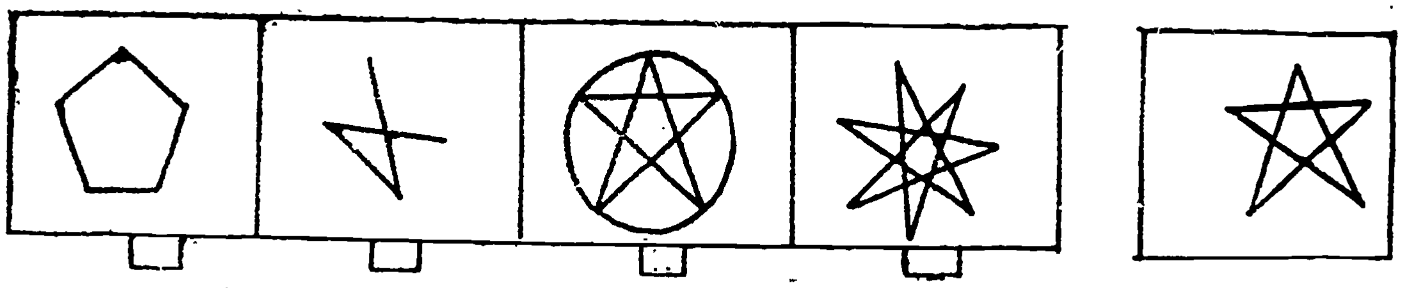
APPENDIX

LIST OF TEST ITEMS

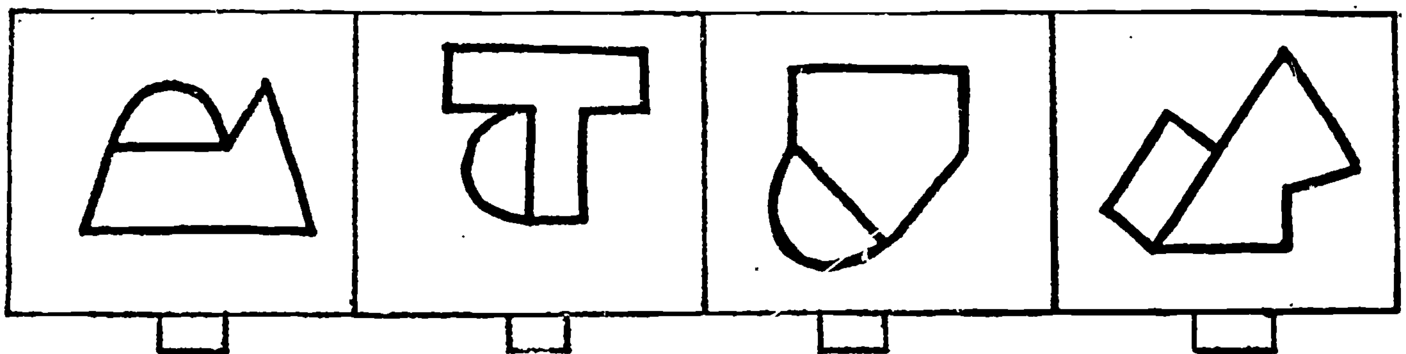
Fxamples 1 - 5



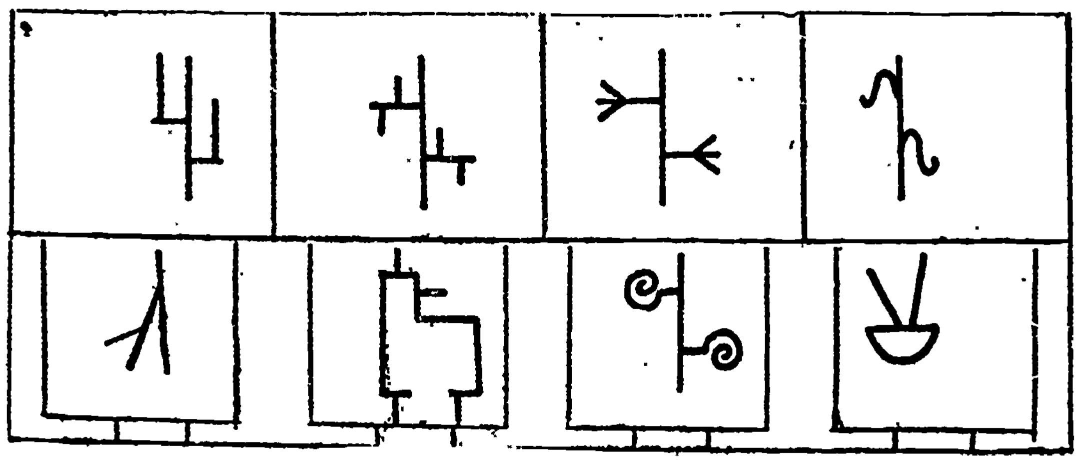
1. Completion: part is given (x), subject is to find whole in which part is embedded.



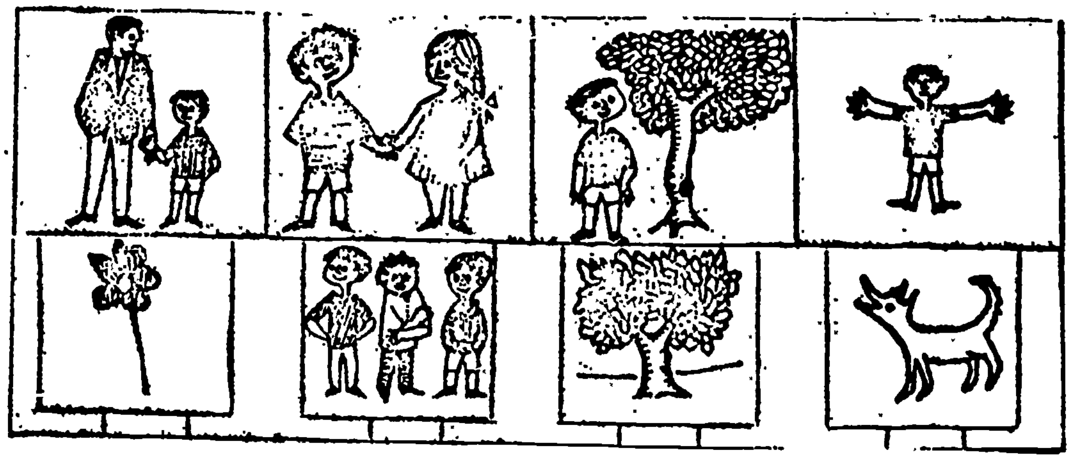
2. Completion: part is given (x), subject is to find whole in which part is embedded.



3. Differences: find which figure is different from given set.



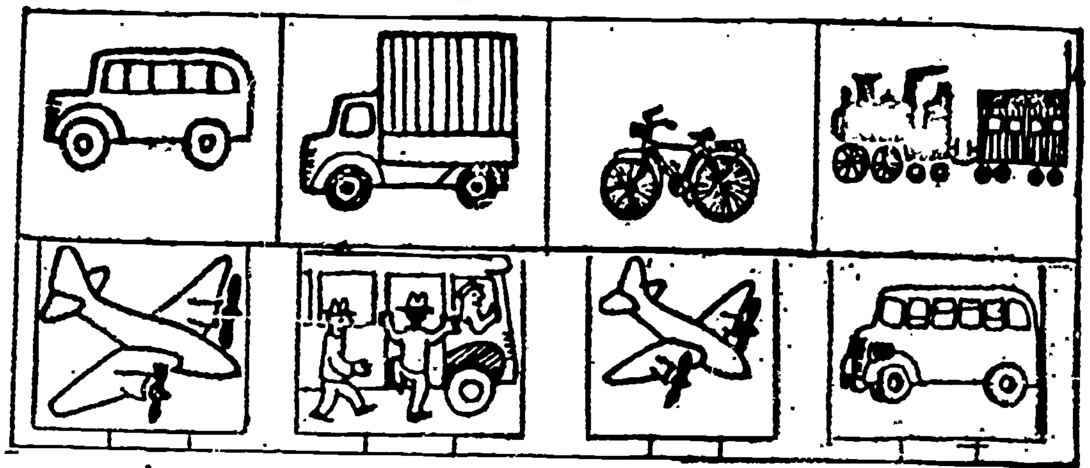
4. Similarities: find which is similar to given set.



5. Similarities: find which is similar to given set.

Examples 6 - 10

6. Similarities: find which is similar to given set.

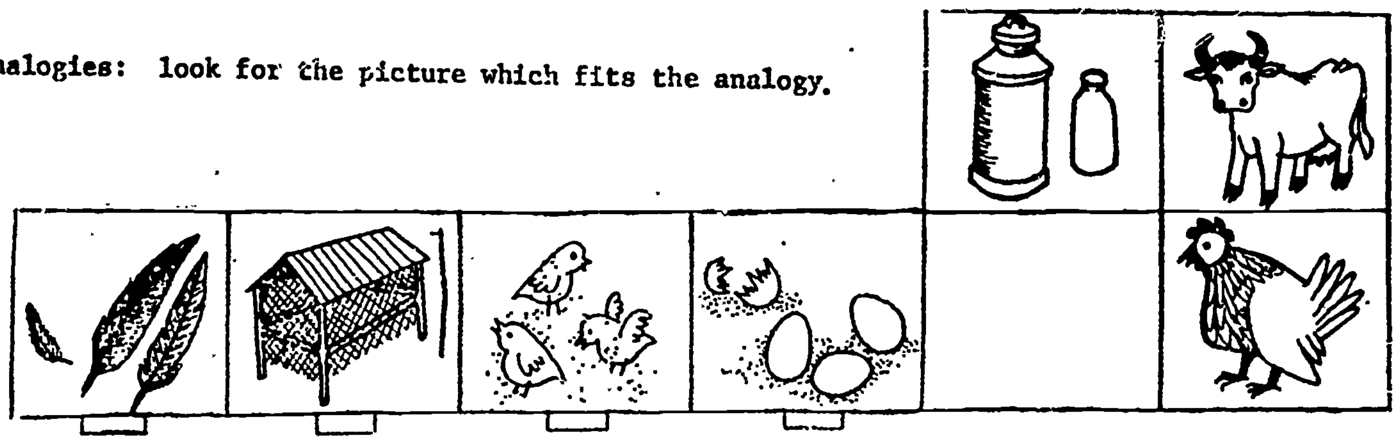


d c b a

7. Similarities: the pictures in the given set describe an absurd situation; the correct answer must resemble the absurd situation

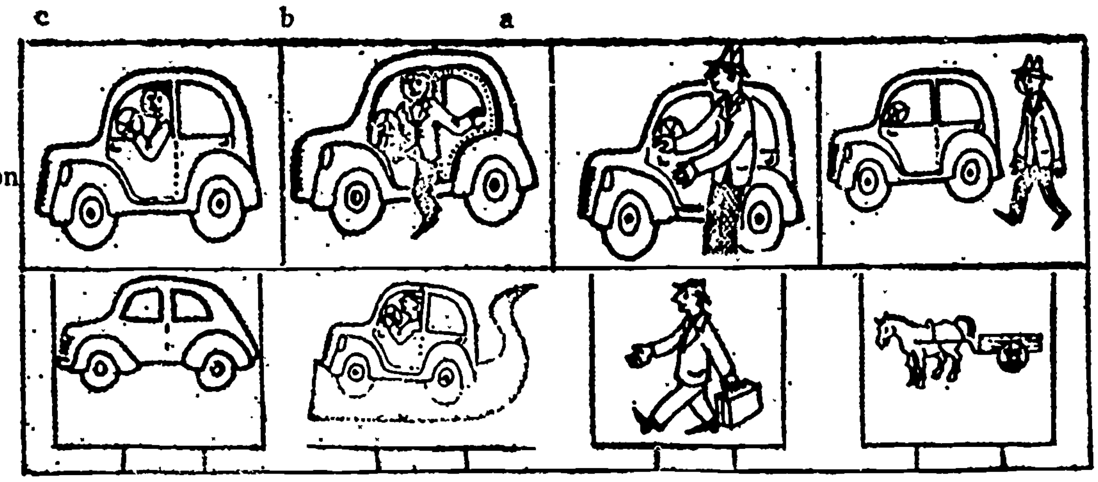
d c b a

8. Analogies: look for the picture which fits the analogy.



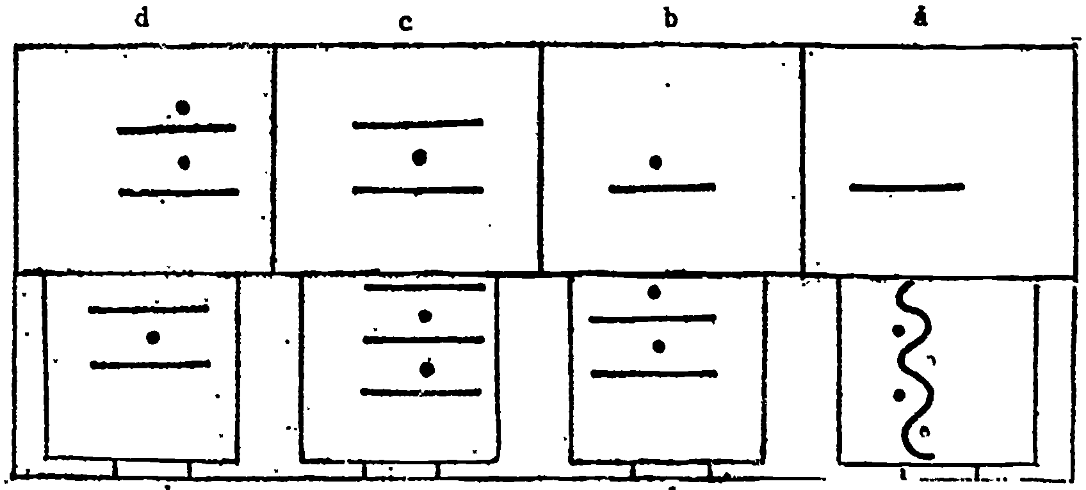
d c b a

9. Progressions: find the next picture in the logical progression (note that Hebrew speaking children would tend to see the movement from right to left, not from left to right.



d c b a

10. Progressions: find the next picture in the logical progression.



d c b a

11. In a school there are 125 pupils, 44% of whom are girls. How many girls are there altogether?

a	b	c	d	e
45	55	44	35	Other
[ ]	[ ]	[ ]	[ ]	[ ]

12. A man takes out a loan from the bank for a year of 500 Israeli Pounds and he pays 57.5 pounds interest. Calculate the percentage rate of interest he pays.

a	b	c	d	e
11.5	86.9	57.5	12.5	17.8
[ ]	[ ]	[ ]	[ ]	[ ]

13. On a cold rainy winter day 58 pupils, 5% of the total number of pupils at the school, did not come to school. How many pupils are there at this school?

a	b	c	d	e
80	1000	1160	290	Other
[ ]	[ ]	[ ]	[ ]	[ ]

14. A merchant received a loan of 800 Israeli Pounds for 3 years and paid 192 pounds interest over the 3 years. Calculate the percentage interest for one year.

a	b	c	d	e
8	51.2	.125	11	Other
[ ]	[ ]	[ ]	[ ]	[ ]

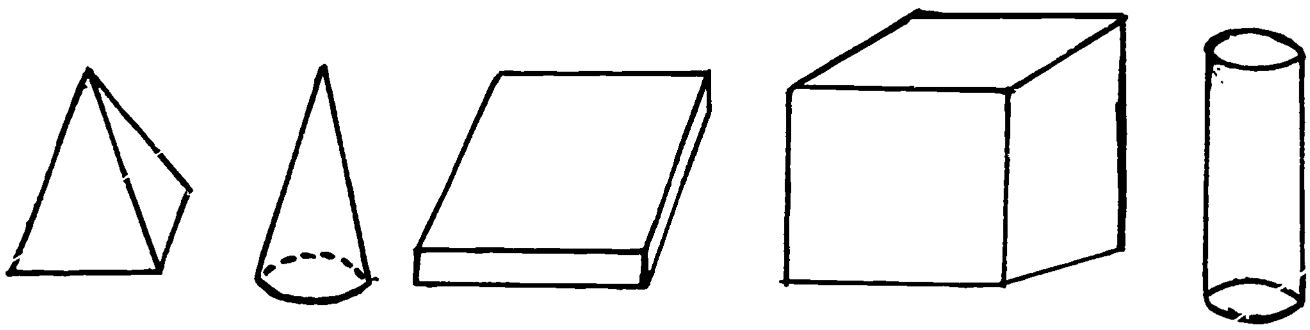
15. What is the sum of the degrees in two adjacent angles?

a	b	c	d	e
90	180	45	120	Other
[ ]	[ ]	[ ]	[ ]	[ ]

16. In what geographical form is every point on the perimeter equidistant from its centre?

a	b	c	d	e
equilateral triangle	circle	paralle-logram	square	other
[ ]	[ ]	[ ]	[ ]	[ ]

17.



A B C D E

The letter A is written under the

a	b	c	d	e
prism	cone	pyramid	cylinder	other
[ ]	[ ]	[ ]	[ ]	[ ]

18. Similarities: Child required to find which is similar to given set.

Dove Raven Eagle Canary

a	b	c	d
boy	butterfly	swallow	dog
[ ]	[ ]	[ ]	[ ]

19. Analogies: Child is required to complete the analogy.

Airplane - air

Car - ?

a	b	c	d
wheel	path	road	steering wheel
[ ]	[ ]	[ ]	[ ]

20. Analogies: Child is required to complete the analogy.

Bus - wheel
Man - ?

a	b	c	d
mouth	foot	hand	head
[ ]	[ ]	[ ]	[ ]

21. Progressions: Child is required to complete the progression.

16 11 13 8

a	b	c	d
20	10	16	3
[ ]	[ ]	[ ]	[ ]

22. Apple Bottle Cat Dart

a	b	c	d
Newspaper	Horse	Ship	Egg
[ ]	[ ]	[ ]	[ ]

23. A merchant receives a loan at 10 percent interest per year, and paid 45 IsraeliPounds interest for 1/2 year. What was the amount of the loan he received?

a	b	c	d	e
1000	450	900	45	other
[ ]	[ ]	[ ]	[ ]	[ ]

24. A farmer takes a loan of 960 Israeli Pounds, at a rate of interest of 8.5 percent per year, for a duration of $1^1/4$ years. The interest was subtracted at the time he received the loan. How many pounds were subracted?

a	b	c	d	e
8.5	102	130	110	other
[ ]	[ ]	[ ]	[ ]	[ ]