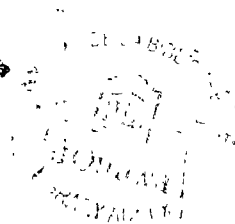
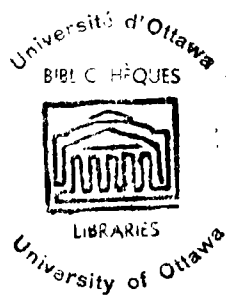


000635
C SP-2

~~A STUDY OF THE~~
RATIONALE OF THE WECHSLER-BELLEVUE
PICTURE ARRANGEMENT SUBTEST, FORM I

by Pascal J. Delli Colli

Thesis presented to the School of Psychology
and Education of the University of Ottawa as
partial fulfillment of the requirements for
the degree of Master of Arts in Psychology.



Ottawa, Canada, 1957

UMI Number EC55155

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform EC55155
Copyright 2011 by ProQuest LLC
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code

ProQuest LLC
789 East Eisenhower Parkway
P O Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGEMENTS

This thesis was prepared under the direction of Doctor Maurice Chagnon whose aid and encouragement have been invaluable.

The writer wishes to express his gratitude to Mr. Remi Lalonde, Principal of St. Lawrence High School, Cornwall, and Mr. Roland Major, Head of the Guidance Department, for their willing cooperation in furnishing the subjects and testing locale.

Sincerest thanks are also extended to Professor Raymond Vaillancourt and to all who in any way assisted in this project.

CURRICULUM STUDIORUM

Pascal Joseph Delli Colli was born on March 18th, 1930, at Montreal, Quebec. He studied at Loyola College in Montreal, and later attended St. Michael's College, Winooski Park, Vermont, where he obtained his Bachelor of Arts degree in the year 1954.

TABLE OF CONTENTS

Chapter	page
INTRODUCTION	vi
I.-A REVIEW OF THE LITERATURE	1
1. Survey of the literature	2
2. Hypotheses to be investigated	15
II.-THE EXPERIMENTAL DESIGN: SELECTION AND DESCRIPTION OF TOOLS	20
1. The procedure	20
2. Sub-hypotheses eliminated and selected	21
3. Sub-hypotheses and the instruments	23
III.-THE EXPERIMENTAL DESIGN: POPULATION AND STATISTICAL METHODS	45
1. The population	45
2. Statistical methods	50
IV.-PRESENTATION AND ANALYSIS OF DATA	55
1. The data	55
2. Analysis of data	64
SUMMARY AND CONCLUSIONS	68
BIBLIOGRAPHY	72
 Appendix	
1. THE RAW DATA	75
2. ABSTRACT OF A <u>Study of the Rationale of the</u> <u>Wechsler-Bellevue Picture Arrangement</u> <u>Subtest, Form I</u>	80

LIST OF TABLES

Table	page
I. - Age Distribution of Sample in Intervals of Four Months	48
II. - Range of Scores, Means and Standard Deviations for Each Test of the Battery Using a Sample of Eighty-Four Individuals .	56
III. - Correlations Between the Various Elements of the Test Battery and the Two Dependent Variables, Using a Sample of Eighty-Four Individuals	59
IV. - Test of Significance for the Obtained Multiple Correlation Coefficients for Each Dependent Variable	62
V. - Raw Data Obtained from the Test Battery . . .	76
VI. - Correlations Between and Among the Dependent and Independent Variables	79

INTRODUCTION

Much use has been made of the Wechsler-Bellevue Intelligence Scale, Form I, both as a measure of intelligence and as a diagnostic instrument. The literature contains a great many references in the latter field while there are relatively few in the former. However, in both instances, the lack of sufficient experimental studies concerning particular functions measured by the subtests has led some authors to urge this type of investigation.

This writer has undertaken to measure the element or elements which may presumably be found in one of the Wechsler subtests, namely, Picture Arrangement. The general problem, then, is concerned with the rationale or nature of this subtest. From the various sources, at least nine specific hypotheses could be distinguished, but not all of them could be investigated.

The first portion of the report presents the various hypotheses resulting from a general overview of the research on the Wechsler. The main theoretical and experimental rationales are also included. Chapter II furnishes one portion of the experimental design. The sub-hypotheses that have been eliminated and those that have been investigated are discussed with the accompanying reasons. The instruments used to examine each of the sub-hypotheses are described in relation to the elements which they could

purportedly measure in Picture Arrangement.

The third chapter completes the treatment of the experimental design and deals with the manner of selection of the sample. The statistical methods of analysis, including the necessary formulae, are also presented. Chapter IV consists of the data obtained with its analysis and discussion.

In conclusion, a summary of the findings is presented with suggestions for further research.

CHAPTER I

A REVIEW OF THE LITERATURE

The reader has garnered from the introduction that the theme with which this study concerns itself is the nature of the Picture Arrangement subtest of the Wechsler-Bellevue Intelligence Scale, Form I¹. In other words, exactly what function or functions are tapped by the Picture Arrangement subtest, and to what extent are these elements present should the subtest contain more than one? This, generally stated, is the problem.

To orient the present study in its proper setting, it is of some moment to consider what theoretical and experimental research has revealed in this field, both broadly and specifically. The review of the literature will necessarily involve a general overview of short forms of the Wechsler-Bellevue in which Picture Arrangement is used, of scatter and pattern analysis, and the nature of the subtest as described by various authors, together with specific and related studies on it. As a result, the related hypotheses will come to the fore, and the necessity and importance of the question will be considered in the light

¹ David Wechsler, The Wechsler-Bellevue Intelligence Scale for Adolescents and Adults, Record Form I, New York, Psychological Corporation, 1947.

of what has been accomplished up to the present time.

1. Survey of the Literature

Many studies on the Wechsler-Bellevue have been reported in the literature, some consisting of investigations of short forms, the majority dealing with scatter and pattern analysis, while one was directly concerned with discovering the nature of Picture Arrangement. All too few have taken up the question of the rationale of the subtests. Among those considered to be important, only two were theoretical studies while one was experimental.

a) The short forms. -- In bringing together the various sources of information on studies conducted with the Wechsler-Bellevue, it was noted that a good many consisted of reports on short forms. Though these are of no direct concern, they may help to give further reason for undertaking the present study. Among the many investigations, very few have included the Picture Arrangement subtest.

A study by Cohen and Steisel², using a population of thirty-two individuals of superior intelligence,

² Bertram Cohen and Ira M. Steisel, "An Evaluation of the Wechsler-Bellevue Test", in the Proceedings of the Iowa Academy of Science, Vol. 54, 1947, as reported by G.D. Lovell in the Psychological Abstracts, Vol. 23, No.11, item 5518, issue of November 1949, p. 689.

revealed that a correlation of .52 existed when Digit Span and Picture Arrangement were compared with the full scale of the Wechsler. It would seem that together, these two can predict the total score to some extent, or it could perhaps be assumed that they measure many of the same functions included in the remaining subtests. If this were actually the case, it would be very important to discover what is measured by these two subtests, to ascertain the overlap existing, and then possibly to eliminate one, or better still, to develop them further so that they might fully measure those elements which do not correlate with the rest of the Wechsler Scale. In connection with this, one author holds that

For justifiable inclusion in a battery a subtest should make a unique contribution, and where the battery is to be interpreted diagnostically, this requirement is quite compelling.³

In another report, using three subtests, Hilden and his associates reported that

The correlation of S [Similarities], PA [Picture Arrangement], BD [Block Design] with T Wtd. is .942, and that of S, V [Vocabulary], BD is .937. While the latter is slightly lower, it would be preferable for clinical usage since the substitution of V for PA reduces the correlation by a very small amount, but adds tremendously to the qualitative diagnostic value. This material may be suggestive of distorted thought processes,

³ Melvin R. Marks, "A Criticism of the Use of the Wechsler-Bellevue Scale as a Diagnostic Instrument", in the Journal of General Psychology, Vol. 49, First Half, issue of July 1953, p. 145.

which are not considered in the intelligence level as such, and which could not be obtained from PA.⁴

With such a high relationship, one might be tempted to be satisfied with using only these three subtests. On the other hand, it would seem to indicate that Picture Arrangement can be easily substituted for Vocabulary with no appreciable change in correlation, and thus that both Picture Arrangement and Vocabulary are tapping similar functions as related to the rest of the test, or can predict the total score.

Using two subtests, with a population of 523 individuals, Gurvitz⁵ showed that those which best predicted the total score were Picture Arrangement and Digit Span with a correlation of .90 with the full scale score. However, McNemar disagreed:

Digit Span plus Picture Arrangement, the pair proposed by Gurvitz, does not appear in the list because it correlates only .770, a value which is near .741, the correlation for the worst pair (Dig. Sp., Objects).⁶

4 A.H. Hilden, J.W. Taylor and P.H. Du Bois, "Empirical Evaluation of Short W-B Scales", in the Journal of Clinical Psychology, Vol. 8, No.4, issue of October 1952, p. 328.

5 Milton S. Gurvitz, "An Alternate Short Form of the Wechsler-Bellevue Test", in the American Journal of Orthopsychiatry, Vol. 15, No.4, issue of October 1945, p.728.

6 Quinn McNemar, "On Abbreviated Wechsler-Bellevue Scales", in the Journal of Consulting Psychology, Vol. 14, No.2, issue of April 1950, p. 80.

Even so, the relationship seems to be sufficiently high that further research into the nature of these subtests would be warranted.

b) Scatter and pattern analysis. -- Another aspect of the investigations on the Wechsler-Bellevue was that in which pattern and scatter analysis formed the bases for the diagnosing of mental dysfunction. Primarily, these studies were intended to differentiate between normals and abnormals, and among abnormals. This was based on the assumption that were any differences obtained either in the pattern or graph represented by performance on the subtests, or in the deviations of the quantitative data obtained between or among subtests, some classification or diagnosis of patients could be made.

One of these reports found that

Three tests reliably differentiate morons from schizophrenics: comprehension and picture arrangement, in which the schizophrenics are inferior to the morons; and arithmetic reasoning in which the schizophrenics are superior to the morons... Perhaps we have here a "sign" of mental deficiency as distinguished from schizophrenia.⁷

Although this was offered tentatively and as a hypothesis, a great many of the studies so offered have been widely

⁷ Ann Magaret and Clare Wright, "Limitations in the Use of Intelligence Test Performance to Detect Mental Disturbance", in the Journal of Applied Psychology, Vol. 27, No.5, issue of October 1943, p. 392.

accepted as factual. The reasons are clear. There was nothing definite enough on which to base diagnoses, and anything such as the above was quickly put to use as a possibly better indication than chance and clinical intuition. As Schofield remarked, "Attempts to use Wechsler-Bellevue subtest patterning as a clue to personality have stemmed more from a wish than from a rationale"⁸.

However, some attempt has been made to study the underlying assumptions experimentally. This was done by Cohen who published some of the material resulting from a factor analysis of the Wechsler⁹. Although the whole of the Wechsler was included, only the section on Picture Arrangement will be considered.

This test measures the Nonverbal Organization factor in all the groups, relatively poorly in the psychoneurotics and schizophrenics and quite well in the brain-damaged group. Its measurement of present general functioning is generally poor in the three groups. The test has low communality with the battery in the psychoneurotics and schizophrenics, i.e., its variance is to a considerable degree "tied up" in measuring something which is not measured by the other tests

⁸ William Schofield, "Critique of Scatter and Profile Analysis of Psychometric Data", in the Journal of Clinical Psychology, Vol. 8, No.1, issue of January 1952, p. 20.

⁹ Jacob Cohen, Multiple Factor Analyses of the Wechsler-Bellevue Intelligence Scale Performance of Schizophrenic, Psychoneurotic, and Brain-Damaged Groups, microfilmed Doctor's thesis presented to the University of New York, New York, March 1950, vii-162p.

over and above the Nonverbal Organization factor. This specific factor may be as Wechsler suggests "social intelligence", or as Rapaport posits "planning ability and anticipation". The author would withhold judgement subject to further factor-analytic investigation, where this specific variance may be "captured" in an interpretable common factor.¹⁰

Although the above statement was offered as a factor-analytically based rationale, the conclusion suggests that the work is far from completed. Much more remains to be done; that is, further analysis is to be carried out before arriving at the underlying functions. As it stands, there has been a little progress towards establishing the basis for a rationale but nothing more. Indications were that further and more specific studies are to be carried out to be able to arrive at the underlying assumptions.

c) Theoretical and experimental rationales of Picture Arrangement. -- The two main theories concerning the nature of Picture Arrangement were advanced first by Wechsler and later by Rapaport. Others have contributed to the experimental phase of investigation.

The first hypothesis on what the Picture Arrangement subtest measures was offered by its author. His ideas on

¹⁰ Jacob Cohen, "A Factor-Analytically Based Rationale for the Wechsler-Bellevue", in the Journal of Consulting Psychology, Vol. 16, No.4, issue of August 1952, p. 275.

the functions tapped were that Picture Arrangement

... is the type of test which effectively measures a subject's ability to comprehend and size up a total situation. The subject must understand the whole, must get the "idea" of the story before he is able to set himself effectively to the task. There is ... some trial and error experimentation but the subject is called upon to attempt appraisal of the total situation... The subject matter of test nearly always involves some human or practical situation. The understanding of these situations more nearly corresponds to what other writers have referred to as "social intelligence". We ourselves do not believe in such an entity. ...social intelligence is just general intelligence applied to social situations.¹¹

It should be noted that this was really more of a theory based on observation than a rationale proper. Nothing has been done by the author to verify the statement experimentally, as far as is known.

Rapaport, however, assumed Picture Arrangement to be a test of "planning ability and anticipation", "judgment and attention":

On the Picture Arrangement subtest we submit that the subject's achievement is a reflection of his ability to anticipate the consequences of initial acts or situations, and hence is a reflection of his planning ability.¹²

¹¹ D. Wechsler, The Measurement of Adult Intelligence, Third Edition, Baltimore, Williams and Wilkins, 1944, p. 88-89.

¹² David Rapaport, Merton Gill and Roy Schafer, "The Bellevue Scale", in Diagnostic Psychological Testing: The Theory, Statistical Evaluation, and Diagnostic Application of a Battery of Tests, Vol. 1, Chicago, Year Book Publishers, 1945, p. 217.

The planning in the Picture Arrangement subtest implies both attention and judgement. Attention should effortlessly convey to the subject the essential features of the individual pictures, or the deviations of the pictures from each other.¹³

Here again it can be said that this is a theory based on clinical observation although some attempt was made at validation through qualitative analysis of the type of failures in individual records.

A factor analysis of the Wechsler, using a male veteran population of Neurotics, Psychotics and Organics between the ages of twenty and forty, brought out a more refined rationale:

The pictures on this test are visual stimuli and are each meaningful, but it requires their arrangement by the subject to make them "tell a story". This phrase is no accident, but points to the necessity for verbal conceptual manipulation in coping with the problem presented by this subtest. As the subject works on any item, he must reason through his placement, set up hypotheses about a part of the sequence and accept or reject them on the basis of the pictures which remain. This process continues until a hypothesis which includes all the pictures of a series is found satisfactory by the subject, that is, until he has the story sequence which makes sense to him, or, alternatively, fails to do so.¹⁴

In addition the comment is made that

¹³ Id., op. cit., p. 219.

¹⁴ Jacob Cohen, Multiple Factor Analyses of the Wechsler-Bellevue Intelligence Scale Performance of Schizophrenic, Psychoneurotic, and Brain-Damaged Groups, microfilmed Doctor's thesis presented to the University of New York, New York, March 1950, p. 78-79.

It [Picture Arrangement] measures the Non-Verbal Organization factor in all the groups, but in addition, may possibly measure verbal ability in the schizophrenics (.25), and distractibility in the psychoneurotics (.23).¹⁵

Although the first statement may give the impression that the findings were strictly experimental, the latter shows that this was not the case. The suggested rationale was not directly investigated; rather, it was but a logical interpretation or point of departure on which to base the general findings and by means of which the factors were to be identified or named.

Other studies were conducted along these same lines. Balinski¹⁶ found two factors in the fifteen-year-old group. He described the first element as non-verbal or performance, while he called the second the verbal or verbalization factor. Among all the age groups, the Picture Arrangement subtest was found to be part of the second factor three times out of six. In his study, Bolle¹⁷ found two elements, the first of which he named the

15 Id., ibid., p. 126.

16 Benjamin Balinski, "An Analysis of the Mental Factors of Various Age Groups from Nine to Sixty", in the Genetic Psychology Monographs, Vol. 23, First Half, issue of February 1941, p. 212-216 and 221-225, passim.

17 Arthur Bolle, A Factor Analysis of the Echelle D'Intelligence Ottawa-Wechsler, unpublished Master's thesis presented to the Institute of Psychology and Education of the University of Ottawa, 1954, p. 59.

ability to acquire and use knowledge of a verbal nature and the second, the ability to evaluate and organize visual material, where "to evaluate" meant the ability to recognize and establish relationships. The Picture Arrangement subtest was found to be part of the second factor five times out of eight. Both studies would indicate that Picture Arrangement is a complex test measuring several things at the same time.

There were, however, certain investigations which attempted to define the nature of the subtest experimentally. One of these was a study by Burnham in which she sought to find whether a relationship could be obtained between the subtest and the percentage of human movement responses on the Rorschach Psychodiagnostic Test¹⁸. This author

... assumed that if both Rorschach human movement responses and Wechsler-Bellevue Picture Arrangement scores are indicative of "social intelligence" or "interest in people" and are also related to general intelligence, some relationship should be demonstrable between them. A randomly selected group of mental hygiene patients was used with various diagnostic classifications included.¹⁹

¹⁸ Hermann Rorschach, Psychodiagnostics Plates, Bern, Hans Huber, 1921.

¹⁹ Catherine A. Burnham, "A Study of the Degree of Relationship Between Rorschach M% and Wechsler-Bellevue Picture Arrangement Scores", in the Rorschach Research Exchange and Journal of Projective Techniques, Vol. 13, No.2, issue of [June], 1949, p. 206.

The hypothesis seemed to be quite reasonable and logical. However, it is thought that had other groups been used, a more meaningful conclusion might have been reached. The very nature of the pictures in the Picture Arrangement subtest suggests human movement or humans in action. The general movement responses could have been tied up with the human responses. Possibly much more could have been gleaned from this study had more Rorschach categories been used together with a sample of normals as well as one of clinical groups. The results could perhaps have helped to lay some groundwork for the present research. Still one fact was noted-- the question of "social intelligence" was brought out. Though Wechsler²⁰ stated that it is only general intelligence, he claims that other authors call it "social intelligence" because the pictures deal with human social situations. Burnham's study stated that practically no correlation was found between her criterion and Picture Arrangement. Yet this is not to be taken to mean that this subtest does not measure "social intelligence". Rather it could be interpreted to mean that other criteria, if such could be had, might correlate significantly. As it was, Burnham found a correlation of only .09, which could be a

²⁰ David Wechsler, The Measurement of Adult Intelligence, Third Edition, Baltimore, Williams and Wilkins Co., 1944, p. 88-89.

chance relationship since her population included only ninety patients.

Another investigation, which compared the Rorschach with the Wechsler-Bellevue, held that for the Picture Arrangement subtest

None of the correlations which were predicted to reveal a relationship between this subtest and $W\%$, and $H\%$ is statistically significant. The significant relationship between picture arrangement and P was not predicted (.05 level of confidence). In addition, the relationships of picture arrangement with $F+\%$ and $CF+C$ approach significance at the .10 level of confidence. The latter, however, is negatively related.²¹

It can be noted that the above report is more extensive than that of Burnham. More scoring categories were included, thus yielding a greater opportunity for arriving at the possible functions tapped by Picture Arrangement. However, only a trend towards three probably significant relationships was found, although these Authors' conception of significance is not thought to be severe enough with the test they used.

Two studies which compared the Primary Mental Abilities subtests with those of the Wechsler had little to offer in the way of results. One did not include

²¹ Jules D. Holzberg and Lillian Belmont, "The Relationship Between Factors on the Wechsler-Bellevue and Rorschach Having Common Psychological Rationale", in the Journal of Consulting Psychology, Vol. 16, No.1, issue of February 1952, p. 27-28.

Picture Arrangement and Digit Span saying that they "... were omitted because of testing time and the fact that these two subtests contribute least to the total score"²². The other²³ included it, but the results were not found to be statistically significant except for the relationship between Picture Arrangement and the Number subtest of the Primary Mental Abilities which was found to be significant at the 5% level of confidence.

One other investigation²⁴ which has no direct bearing but which will help to explain the raison d'etre of the present research, concerns the Digit Symbol subtest of the Wechsler. This was the only experimental study, apart from that of Burnham mentioned above, which directly tried to discover the nature of one of the subtests, the others being concerned with the whole or parts of the Wechsler. As far as this author knows, no other attempts have been made to inquire into the rationale of any of the

22 George Becker, The Relationship Between Thurstone SRA Primary Mental Abilities Tests and the Wechsler-Bellevue Intelligence Test, unpublished Master's thesis, presented to Fordham University, New York, 1948, p. 14.

23 Arthur Smith, A Comparison of the SRA Primary Mental Abilities Test with the Wechsler-Bellevue Intelligence Scale, unpublished Master's thesis, presented to Illinois State Normal University, Normal, 1949, p. 17.

24 T. Burik, "Relative Roles of the Learning and Motor Factors Involved in the Digit Symbol Test", in the Journal of Psychology, Vol. 30, First Half, issue of July 1950, p. 41.

subtests.

The foregoing sections give rise to one question which apparently remains unanswered, namely, the choice of the topic. Why was Picture Arrangement to be investigated when any of the other subtests could have been examined? As was seen above, one of the reasons for its choice stemmed from studies that have already been done on it, both theoretical and experimental. Another factor is the seeming importance of the subtest in pattern analysis. Again, its inclusion in some of the short forms would seem to indicate that it is of some importance. These factors have led the writer to investigate the nature of Picture Arrangement since it is being used in various situations and especially because performance on it is being interpreted according to the existing hypotheses.

2. Hypotheses to be Investigated

Bringing together all the material that in any way suggested possible hypotheses, it is seen that there were many. Looking over the suggested rationales, one can easily detect the mention of factors such as non-verbal organization, comprehending and sizing up a total situation, social intelligence, anticipation, planning ability, reasoning, distractibility, evaluation or ability to recognize and establish relationships, and those that have

been experimentally verified and found to have no relationship such as human movement, whole responses, color and form, pure color, and good form level or quality on the Rorschach. These are hypotheses that should be verified, including those that have already been investigated.

All the studies that have been mentioned point to the necessity for further and thorough examination. Watson in 1946 held that

The differences of opinion in regard to rationale between Wechsler and Rapaport are both encouraging and disconcerting. Encouraging, because they show that attention is being given to the meaning of the subtest scores obtained; and disconcerting, because they show that much remains to be done before a thoroughgoing rationale for the various subtest scores is established... Further experience, clinical, experimental and statistical, will eventually allow more final judgements to be drawn. Studies showing whether scores on the subtests are correlated with successful performance in other situations which require the same functions seem to be indicated. Factor studies are also applicable.²⁵

The above statement shows the importance of the present undertaking and also suggests a method by which the investigation is to be conducted. It may be argued that it was written too long ago to have any application at the present date, but in 1952 the same cry was heard from Holzberg and Belmont:

25 Robert I. Watson, "The Use of the Wechsler-Bellevue Scales: A Supplement", in the Psychological Bulletin, Vol. 43, No.1, issue of January 1946, p. 66.

While the rationale for many clinical instruments has been slowly accumulating, this has been more in the nature of assumptions rather than experimentally demonstrated facts. While Wechsler offers some rationale for the subtests of his intelligence scale, these are more incidental comments than a systematic development of test rationale in terms of a psychological theory which "deals with the rationale of how, by what processes, the organizing principles of the personality and their pathological alterations are carried over into test performance [3, p. 248].²⁶

It is evident, then, that it is of some concern that something be done to discover what lies behind each of the subtests. Although some of the short forms which included Picture Arrangement seemed capable of substituting for the whole test or of predicting the total score, something should be known about what each subtest measures as related to the rest of the test. Scatter and pattern analysis can be useful, but if the matter is to be comprehensive, the rationale must needs be verified. Unless it is known exactly what psychological functions or processes have been affected in any particular case under study, and whether the same or different factors have been influenced in a number of individuals on a particular subtest, a diagnosis cannot be precise. In other words, should a subtest measure more than one or two elements, intratest patterning could perhaps be more revealing in differential diagnosis.

²⁶ J.D. Holzberg and L. Belmont, op. cit., p. 23.

Some attempts, however, have been made to clarify the situation, and although the results do not seem too encouraging, many hypotheses still remain to be studied. The present research, then, was undertaken to show something of the rationale of the Picture Arrangement subtest and, in a limited sphere, to gain further knowledge of what is involved in the process of arranging a specific series of pictures in proper sequence.

From the literature the following general hypothesis was derived: the Picture Arrangement subtest of the Wechsler-Bellevue Intelligence Scale, Form I, does not measure one unitary function. To this general proposition, a number of specific hypotheses are related, namely:

1. Picture Arrangement measures "social intelligence", (Wechsler).
2. Picture Arrangement measures "planning ability", (Rapaport).
3. Picture Arrangement measures "foresight or anticipation", (Rapaport).
4. Picture Arrangement measures "judgement", (Rapaport).
5. Picture Arrangement measures "attention", (Rapaport and Cohen).
6. Picture Arrangement measures "reasoning", (Cohen).
7. Picture Arrangement measures "distractibility", (Cohen).
8. Picture Arrangement measures "a verbal or verbalization factor", (Cohen, Balinski and Bolle).

9. Picture Arrangement measures "non-verbal organization, or ability to recognize and establish relationships, or logical sequence and relationships", (Cohen, Balinski and Bolle).

The definitions of the above terms are not included at this point since an explanation is forthcoming in the next portion of this report.

Having established the suggested elements which in any way contribute to the general hypothesis, the method of investigation, the possibility of studying each factor and the instruments used will be elaborated upon in the following chapter.

CHAPTER II

THE EXPERIMENTAL DESIGN: SELECTION AND DESCRIPTION OF TOOLS

The general and specific hypotheses obtained from the review of the literature give rise to the type of approach necessary to investigate them. From the method of study will result the possibility or inability of examining the particular assumptions set forth in the preceding chapter. Finally, the selection and description of tools, as related to the sub-hypotheses to be verified, will be considered.

Accordingly, this chapter reports on the three steps taken:

1. The procedure necessary to study the suggested rationales;
2. The elimination and selection of sub-hypotheses and the reasons therefor;
3. The selection of instruments to be used to test the assumptions chosen for verification.

The development of these topics form the content of the sections which follow.

1. The Procedure

Before arriving at a final formulation of the specific problem, it is necessary to outline the method by which this study is to be conducted. As was suggested in the

literature, a correlational approach¹ is to be followed in the investigation of the suggested sub-hypotheses. This, therefore, will involve the use of test scores which are to be compared with those obtained on Picture Arrangement. Since the latter consists of three separate scores (accuracy, time bonus and total), each of them is to be compared with each of the tests used to verify the particular assumptions. It is hoped that with this refinement the results may possibly be more extensive and applicable.

From the number of elements which are to be examined, the use of multiple correlation has also been indicated in order to determine the relative contribution of each factor which may be present in Picture Arrangement.

2. Sub-Hypotheses Eliminated and Selected.

Due to the method of investigation, some of the specific hypotheses were necessarily eliminated while others were studied.

a) Sub-hypotheses eliminated. -- A problem that poses itself is whether all of the specific hypotheses can be examined, as well as why they can or cannot be. In answer to this it can be stated that although it had been

¹ Robert I. Watson, "The Use of the Wechsler-Bellevue Scales: A Supplement", in the Psychological Bulletin, Vol. 43, No.1, issue of January 1946, p. 66.

hoped to test for "social intelligence", "planning ability and anticipation", no adequate measuring instruments could be found for any of these. Though there exists a test which presumably measures "social intelligence", it is thought to tap only general intelligence as applied to social situations² and is generally considered an invalid measure. With respect to "planning and anticipation", the existing tests are still in a very early stage of development and are under investigation at the present. Yet, be that as it may, Thurstone's Primary Mental Abilities, Reasoning subtest³, supposedly contains these two factors. More will be said in the discussion of this test in connection with the "reasoning" sub-hypothesis. Nor was "judgement" to be tested for; rather, the "reasoning" element is to be studied because the former has been found to contain the latter element⁴, and conversely, that the "judgement" factor "...had weight in the Reasoning test of

2 Harold G. Seashore, Director of the Test Division, Psychological Corporation, letter to the Author, November 29, 1955.

3 L.L. Thurstone and T.G. Thurstone, Examiner Manual for the SRA Primary Mental Abilities, Preliminary Edition, Chicago, Science Research Associates, 1947, p. 1.

4 J.P. Guilford, Editor, and J.I. Lacey, Research Associate, Army Air Forces Aviation Psychology Program Research Reports, Printed Classification Tests, Report No. 5, Washington, Superintendent of Documents, U.S. Government Printing Office, 1947, p. 129.

Thurstone"⁵.

b) Sub-hypotheses investigated. -- With the elimination of the "judgement", "social intelligence" and "planning ability and anticipation" elements, those which remain for investigation give rise to the specific problem, namely, whether Picture Arrangement measures reasoning or logical sequence and relationships, verbal conceptualization, attention and distractibility, and to what extent each of these are to be found in the subtest. Only those factors which could come under scrutiny will be included primarily because the necessary instruments are available and also because they are thought to be valid and reliable measures.

3. Sub-Hypotheses and the Instruments

In considering the tests, their choice will be explained with special reference to the sub-hypotheses which they are supposed to examine. Because of the difficulties involved in choosing one definition in preference to another and in the rapprochement between the one chosen and the one given to a test by its author, it was decided to accept those given to tests which are thought to

⁵ Harold M. Corter, "A Factor Analysis of Some Reasoning Tests", in the Psychological Monographs, Vol. 66, No.8, issue of March 1952, p. 28.

compare best with the elements of the hypothesis which were selected for study. Since for each test a definition is usually given as to what it is supposed to measure, that meaning which best described what apparently occurs in the arrangement of the pictures in Picture Arrangement was chosen to verify a particular element of the hypothesis.

Besides the question of definitions, the problem of "pure" tests also arises. Although it cannot be claimed that such tests exist, some do approximate this requirement. Those that have been subjected to factor analysis have in most cases shown to be closest. Others have not been so analysed but they are also thought to be useful here because of the singularity of content. Thus, the tools selected to study the various aspects of the Picture Arrangement subtest are: the Primary Mental Abilities Test⁶, Verbal-Meaning and Reasoning subtests; the Differential Aptitude Tests⁷, Abstract Reasoning portion; the Minnesota Clerical Test⁸, Number and Name Checking.

⁶ L.L. Thurstone and T.G. Thurstone, SRA Primary Mental Abilities, Chicago, Science Research Associates, 1947, p. 1-2 and 5-6.

⁷ G.K. Bennett, H.G. Seashore, A.G. Wesman, Differential Aptitude Tests, Abstract Reasoning, Form A, New York, Psychological Corporation, 1947.

⁸ D.M. Andrew, D.G. Paterson and H.P. Longstaff, Minnesota Clerical Test, New York, Psychological Corporation, 1933.

Various reasons presented themselves for this choice of tests. These will be considered in the following subsections together with other relevant material.

Since the main concern is the Picture Arrangement subtest, it would be well to begin with its description and to follow through with the relation of the other tests to their specific hypotheses. The characteristics of the tools, methods of scoring, a general spercu of the literature, and a word on their validities and reliabilities will also be put forth.

a) Picture Arrangement. -- The Picture Arrangement subtest consists of seven items, the first of which is a sample or demonstration series of three separate picture cards-- much like cartoon drawings. And just like cartoons, these series can be arranged in such a way that they will tell a story. The first series (House) contains three cards-- as in the sample set. The second and third items (Holdup and Elevator) have four cards each, while the fourth (Flirt) contains five. In the last two series (Taxi and Fish) six cards are included in each. A certain amount of manipulation of the cards is necessary and the test is timed. The timing is not similar to that of a speed test since the time allotted is quite generous-- one minute being allowed for each of the first three series and two minutes for each of the last three. The timing is

necessary especially for the last two series where a time bonus of three points is given if an accuracy score of two or three is obtained within twenty seconds, or a bonus of two points within thirty seconds, or one of one point within forty seconds.

For the last two items the accuracy score varies from three points to one point according to a set arrangement while a score of three or two points is given for the fourth series. A score of zero is given when the arrangement does not conform to the list⁹ given by Wechsler, or when an individual goes beyond the allotted time for any one series. The first three items have a score of two each if accurate, with no time bonus, or a score of zero if inaccurate.

An important point to be considered centers on the reliability of the subtest. Nothing definite has been mentioned about it by Wechsler. However, Derner and his associates¹⁰, using a sample of 158 individuals, found a test-retest reliability to be .64 and concluded that it

⁹ David Wechsler, The Measurement of Adult Intelligence, Third Edition, Baltimore, Williams and Wilkins Co., 1944, p. 179.

¹⁰ G.F. Derner, M. Aborn, and A.H. Canter, "The Reliability of the Wechsler-Bellevue Subtests and Scales", in the Journal of Consulting Psychology, Vol. 14, No.3, issue of June 1950, p. 177.

appears to be one of the least stable subtests. Yet the procedure of re-testing is worthy of note. Out of the 158 subjects, sixty were re-tested after one week, sixty after four weeks, and thirty-eight after six months. They noted that scores increased by 2.8 weighted score points after one week using sixty individuals. This tends to show that practice effect may have had quite an influence on the results and was a factor to be controlled. Once the content is known by the subject, it would be easy for him to think over the arrangement until such time as he is to take the retest, and possibly have an inkling that he had it wrong. Another possibility concerns the technicians who administered the test. If any indication was given that the arrangement was not correct, the subject would be more careful in the next testing, or again there may have been discussion among the subjects. All kinds of controls may have been lacking since none of them were explicitly mentioned. However, apart from these comments, the procedure used is rather questionable. Due to its very nature, Picture Arrangement is not amenable to a test-retest method. The content already being known, it would be difficult to draw any conclusions for purposes of assessing its reliability. The method used would be more appropriate in discovering the effect of practice in this subtest.

Another consideration, brought out by Wechsler, might affect to some extent the reliability of the subtest:

It is of some moment whether the story told by the pictures is that of a bird building its nest, or a policeman pursuing a thief in a radio car. The former is a situation a country boy may grasp at once, the latter may puzzle him a good deal; and what holds for such simple situations plays an even greater role when the story told by the pictures is more complicated. ...we found that subjects of a particular cultural level were rather favored by one or another of the series. For example, some of the Soglow pictures, which are not only among the best, but also among the most difficult, were surprisingly easy for the sophisticated type of individual.¹¹

This would indicate that cultural factors have a definite influence on the Picture Arrangement scores. On the other hand, depending on the interpretation given the statement, "Educational and social factors made very little difference"¹². If "educational and social" factors are taken to mean cultural factors, then there is disagreement between Wechsler and Gurvitz. However, it will depend on the type of groups chosen as to how much of an influence these factors will have on performance. If the groups are homogeneous, intra-group comparisons of the results should not vary to any great extent as far as these elements are

¹¹ David Wechsler, op., cit., p. 88.

¹² Milton S. Gurvitz, "An Alternate Short Form of the Wechsler-Bellevue Test", in the American Journal of Orthopsychiatry, Vol. 15, No.4, issue of October 1945, p. 728.

concerned.

b) Picture Arrangement measures a verbal and a reasoning element. -- It was said that a test which had been factor analyzed would approach a "pure" test. The one which suggests itself as a measure of both the verbal and reasoning factors is Thurstone's SRA Primary Mental Abilities Test. The literature in general has tried to show whether or not there are "primary abilities" and how many there actually are, while other studies dealt with the prediction of academic achievement and success at the college level. The former investigations were conducted at a time when factor analysis was still in its early stage. In connection with this McNemar, in 1939, stated that

So far as the factorial methods are concerned it would seem that the weakest links in an otherwise strong chain are those which bridge the gap between mathematical factors and psychological meaning and the chasm between sample and universe values. In fact the latter link is practically missing, and from the reviewer's viewpoint, this constitutes a serious limitation. This not only has to do with how many centroid factors are of non-chance significance, but also with the stability of test projection.¹³

Spearman argued somewhat along the same line, though he went further:

¹³ L.L. Thurstone, "Primary Mental Abilities", in Psychometric Monographs No.1, as reviewed by Quinn McNemar, in the Journal of Educational Psychology, Vol. 30, No.1, issue of January 1939, p. 76-77.

The results are the opposite to anything ever obtained in previous researches ... there has been found a complete absence of any general factor. ...the new method ... is liable to go completely astray in a certain application of questionable principles to inadequate data.¹⁴

These comments were made when little was known about the application of factor analysis and when it was difficult to accept the theory of multiple abilities. To-day, however, this theory has become widely accepted and many more factors have been found other than the seven found by Thurstone, although the problem of identification of factors is ever present. In 1951, Michael and his co-workers found that "with the exception of S. [spatial relations] and Vs [visualization] the factors identified are common with those of Thurstone"¹⁵. Guilford and his associates¹⁶ found many more "primary abilities" though as yet they have not been fully validated. Much work is still being done in ascertaining their presence.

14 C. Spearman, "Thurstone's Work Re-worked", in the Journal of Educational Psychology, Vol. 30, No.1, issue of January 1939, p. 16.

15 W.B. Michael, W.S. Zimmerman, and J.P. Guilford, "An Investigation of the Spatial Relations and Visualization Factors in Two High School Samples", in Educational and Psychological Measurement, Vol. 11, No.4, issue of Winter 1951, p. 572.

16 J.P. Guilford, Editor, and J.I. Lacey, op. cit., p. 823-838, passim.

The above was given by way of introduction to the Primary Mental Abilities Test. Only a part of the test has been used in this project, namely, the Verbal-Meaning and Reasoning subtests.

1) Verbal-Meaning. -- The choice of this subtest is based on the definition given to it by Thurstone. He defined Verbal-Meaning as the "... ability to understand ideas expressed in words"¹⁷. But what relationship could this have to Picture Arrangement which is a non-language test? Following Cohen's¹⁸ line of thought and findings, the pictures are visual stimuli requiring their arrangement by the subject to make them "tell a story", necessitating verbal conceptual manipulation in coping with the problem presented by the subtest. Paraphrasing this statement, the subject must first have an idea of what he is to do and actually does, then tell himself the story to see if all the parts fit together. Although he is not asked the story orally, he must produce and manipulate that story mentally. To produce that story, the individual must use

¹⁷ L.L. Thurstone and T.G. Thurstone, Examiner Manual for the SRA Primary Mental Abilities, Preliminary Edition, Chicago, Science Research Associates, 1947, p. 1.

¹⁸ Jacob Cohen, Multiple Factor Analyses of the Wechsler-Bellevue Intelligence Scale Performance of Schizophrenic, Psychoneurotic, and Brain-Damaged Groups, micro-filmed Doctor's thesis presented to the University of New York, New York, March 1950, p. 78.

words or give "mental voice" to his ideas. It will also be necessary for him to know the meaning of the words he uses for the proper expression of those ideas.

Reasoning thus, a test is needed to fulfill the above requirement. Since the definition ascribed to the Primary Mental Abilities Verbal-Meaning subtest is such that it coincides with the hypothesis, it was chosen as one of the tests to be included in the battery.

The subtest consists of fifty items, each of which requires the subject to choose one of four words which has the same meaning as the first or given word. The testee has four minutes in which to complete the test. Each individual answers on a self-scoring pad and the score on this test is the sum of the number of correctly marked items, with no correction for guessing.

Before considering the reliability of the subtest, a critical note should be added. Wesman and Seashore¹⁹ caution that there may be a difference between so-called verbal ability and vocabulary. These may be two entirely different concepts. In line with this it is noted that Thurstone included a Verbal-Meaning test in his battery which may be simply recognition of words having the same

19 A.G. Wesman and H.G. Seashore, "Frequency Vs. Complexity of Words in Verbal Measurement", in the Journal of Educational Psychology, Vol. 40, No.7, issue of November 1949, p. 395.

meaning, or, possibly "verbal ability" or facility with words and their correct usage, and a Word-Fluency test which may possibly be what is called the "ability to verbalize".

Although the Verbal-Meaning subtest is supposedly loaded with a verbal factor, one study revealed that it

...has significant weights on the "academic" factor, the "judgement" factor and the "productiveness" factor. In terms of the parameters used in this test it can be described as a measure of learned activity that requires discrimination and evaluation in its solution.²⁰

The first factor, A, was called an "academic" or learned activity factor. It found its expression in vocabulary tests ... a reasoning test, and others. This factor was believed to be similar to a "restrictive task" factor of Thurstone and a "central integrative field factor" found in Halstead's work.²¹

The third factor, J, was called a "judgement" factor. It involves discrimination, ability to weigh and evaluate, to exercise control in one's thinking, and to function wisely as in a mature person.

The factor, P, called "productiveness" was primarily concerned with speed of output, although ideational richness played some role.²²

This information tends to clarify what the subtest measures.

Though the verbal factor was present, three additional factors were found to be included. It will be noted that

20 Harold M. Corter, op.cit., p. 22.

21 Id., ibid., p. 23.

22 Id., ibid., p. 28.

this statement of what is contained in the subtest coincides quite well with the hypothesis of the present study although the third factor was not included as such.

The reliability of the test is quoted at .92 by Thurstone²³, using a sample of five hundred tenth grade students. This was presumably based on a test-retest situation although it is not clearly stated in the Manual whether this was the actual procedure. However, Traxler²⁴, using a sample of 104 girls, found a test-retest reliability coefficient of .81 for grade eleven, and .83 for grade twelve. He went on to say that from these results the test seems to be reliable.

11) Reasoning. -- Thurstone's definition of "reasoning" was a primary factor in the choice of this subtest. The definition states that the Reasoning test measures "the ability to solve logical problems,-- to foresee and plan"²⁵. It was said earlier that "foresight and planning" would not be investigated because of the lack of

23 L.L. Thurstone and T.G. Thurstone, Examiner Manual for the SRA Primary Mental Abilities, Preliminary Edition, Chicago, Science Research Associates, 1947, p. 7.

24 Arthur Traxler, "Stability of Scores on the Primary Mental Abilities Tests", in School and Society, Vol. 53, No.1365, issue of February 1941, p. 256.

25 L.L. Thurstone and T.G. Thurstone, Examiner Manual for the SRA Primary Mental Abilities, Preliminary Edition, Chicago, Science Research Associates, 1947, p. 1.

a measuring instrument. It might be argued that the Reasoning test could have been used for this purpose, but studies conducted by Army Air Force Psychologists²⁶ apparently show that there is more to "planning and foresight" than the definition would seem to indicate. However, the first part of the definition, the solving of logical problems, seemed to fit in with the specific hypothesis in which there is a mention of logical sequences.

The Reasoning portion of the Primary Mental Abilities is made up of thirty items which are to be solved within a six minute period. Each item consists of a series of letters which follow each other according to a certain rule. The subject must study the line, discover the rule and mark the next letter that would logically follow if the line were to be continued. For example, if a line read "abababab", the next letter would be "a"; or if it read "abghcdijef", the letter that would logically follow would be "k". The score is the total number of correct answers with no correction formula for guessing.

Carter's²⁷ study, mentioned in connection with the Verbal-Meaning subtest, found the same three factors

²⁶ J.P. Guilford, Editor, and J.I. Lacey, op. cit., p. 823-838, passim.

²⁷ Harold M. Carter, op. cit., p. 22 and 28.

applicable to the Reasoning subtest except that these were highly loaded, with moderate loadings in an unnamed factor, and thus would presumably be measuring the same things as the Verbal-Meaning subtest. However, Thurstone²⁸ found a correlation of only .178 between the two subtests.

A noteworthy item was mentioned by Thurstone in connection with the "reasoning" factor:

Recent research has shown that R is really two separate abilities: inductive reasoning, the ability to reason from specific cases to a general rule; and deductive reasoning, the ability to reason from stated premises to a logical conclusion. The present test is a composite measure of both abilities.²⁹

Although the test is a composite measure of both abilities, it might have been better had they been separated to yield two scores so that the influence of each could have been advantageously explored in relation to the problem of this research.

The test-retest reliability of the Reasoning test is quoted at .93 by Thurstone³⁰ using the same population as for the Verbal-Meaning test.

28 L.L. Thurstone and T.G. Thurstone, Examiner Manual for the SRA Primary Mental Abilities, Preliminary Edition, Chicago, Science Research Associates, 1947, p. 7.

29 Id., ibid., p. 2.

30 Id., ibid., p. 7.

c) Picture Arrangement measures logical sequence and relationships. -- Although one test had already been included to study the "reasoning" factor, another was chosen as possibly measuring what the other did not and also as a probable check on the other. Because of the similarity with Picture Arrangement, of the reasoning involved, and also that it is a non-verbal test, the Abstract Reasoning portion of the Differential Aptitude Tests will be used to investigate this aspect of the hypothesis. The test's Authors comment as follows:

The Abstract Reasoning test is intended as a non-verbal measure of the student's reasoning ability. The series presented in each problem requires the perception of an operating principle in the changing diagrams. In each instance, the student must discover the principle or principles governing the change of the figures and give evidence of his understanding by designating the diagram which should logically follow. ³¹

From this statement it can readily be seen why it was thought that the Abstract Reasoning subtest would be a likely measure of "logical sequence and relationships".

Since the battery is still at an early stage of development, though being widely used, not enough time has elapsed for the completion of any great number of investigations. Much of the work that has been accomplished

³¹ G.K. Bennett, H.G. Seashore and A.G. Wesman, A Manual for the Differential Aptitude Tests, New York, Psychological Corporation, 1952, p. 6.

was concerned with the question of reliability as well as with the prediction of achievement in certain school subjects. With reference to the Abstract Reasoning test, no study was found which had any direct bearing here.

The Abstract Reasoning subtest consists of fifty items which are to be solved within a twenty-five minute period. It is a non-verbal or non-language type of test in which the subject is required to reason through a series of four diagrams to see what is going on, and then to select one out of five choices which would logically follow the given sequence of designs. The test resembles Picture Arrangement to some extent in that the procedure is somewhat similar. The resemblance carries through as far as noticing details quickly and accurately, seeing the differences between them, and making a choice which would logically follow. The differences arise from the fact that Picture Arrangement deals mostly with human situations, with the individual cards requiring some manipulation, while Abstract Reasoning deals with symbols, the first four diagrams being already arranged and necessitating the choice of a fifth.

The test is so constructed that it makes all diagrams clear, and the differences between one diagram and another self-evident. However, the progressive differences between designs in one series are clues to the rule which

will influence the choice of the fifth diagram.

The method of scoring is such that an individual is punished for guessing. The formula is the sum of the number of items correct less one quarter the sum of errors.

The reliability of the test has been reported³² to be .62. This was obtained on test-retest data gathered from grade nine boys whose scores were compared with those they received when they reached grade twelve. However, test-retest coefficients of .90 and of .92 were reported³³ for grades eleven and twelve respectively.

Comparing the Reasoning subtest of the Primary Mental Abilities with the Abstract Reasoning test, a correlation of .60 has been quoted, while one of .44 was found between the latter and the Verbal-Meaning of the former³⁴. Nonetheless, the coefficients are thought to be sufficiently low for the inclusion of both reasoning tests, and also because one purportedly taps both inductive and deductive reasoning while the other attempts to measure abstract reasoning or "The ability to perceive relationships in abstract figure patterns, generalization, and eduction of

32 Id., ibid., p. 68.

33 Id., ibid., p. 66.

34 Id., ibid., p. 72.

principles from nonlanguage designs"³⁵.

d) Picture Arrangement measures attention and distractibility. -- The verification of both of these elements must be carried out by a test or tests which will investigate each factor either directly or indirectly. It is thought that the attention factor could best be measured by the Minnesota Clerical Test, using both the Number and Name Checking portions. However, the distractibility factor is also to be verified by means of the same test. The choice of this instrument is based on the definition ascribed to it, namely, that it measures "The ability to observe and compare, to discriminate small differences rapidly, to adjust to a new situation, and to give attention to a problem"³⁶. Each element in this definition apparently applies to Picture Arrangement as is evident from observation.

The attention factor is to be studied directly, but distractibility is to be investigated indirectly. A critical note on scoring seems to be in order at this point. An article by Candee and Blum suggests a revised scoring system so as to have both a speed and an accuracy score:

35 Id., ibid., p. 7.

36 D.M. Andrew, "An Analysis of the Minnesota Vocational Test for Clerical Workers", in the Journal of Applied Psychology, Vol. 21, No.2, issue of April 1937, p. 169.

The scoring technique of right minus wrong does not apply to the test because an answer is not determined by chance, but is a result of a comparison of two series of items. The answer is based upon an individual's ability to pay attention to all the parts of the numbers or names to be compared.³⁷

Although the study had been undertaken from a practical standpoint, that is, for the hiring of clerical staff, it also has practical applications here. The present writer has refined the statement further and has assumed that the error score would effectively represent a measure of distractibility or lack of attention. Of course, this assumption is also based on the definition given to the Clerical Test.

The test consists of two parts-- one being Number Checking, the other Name Checking. The first part has a time limit of eight minutes while the latter is limited to seven. Each portion consists of two hundred pairs of items. The subject is instructed to check the left hand items with those on the right, and if they are the same, to place a check mark in the space provided between each pair. If they differ, he is to make no mark in the space. The Number Checking portion consists of series of numbers, some short, some rather long. The idea is to note whether one

³⁷ B. Candee and M. Blum, "A New Scoring System for the Minnesota Clerical Test", in the Psychological Bulletin, Vol. 34, No.8, issue of October 1937, p. 545.

or more of the numbers in a series does or does not coincide with the other of the pair. Similarly, in Name Checking it is a question of noting whether one or more of the letters of one of the pair does or does not correspond exactly with the other. The score on each test is the sum of the correct items minus the sum of errors. The scores remain separate for the two tests.

A great amount of work was put into the Clerical Test, and much of what has been accomplished has consisted mainly in the establishment of norms, in the prediction of success in clerical occupations, and thus became an instrument of selection in that field. Another portion of the work has centered on the establishment of its validity and reliability.

The use of both tests, Number and Name Checking, was dictated by one of the authors:

Further reason for the retention of both the clerical checking tests in the battery is that the validity of the test is considerably increased when both of the tests are included.³⁸

The reliability of the test is given as .76 for Number Checking, .77 for Name Checking, and .86 for both parts when they were taken together in a test-retest type of

³⁸ D.M. Andrew, "An Analysis of the Minnesota Vocational Test for Clerical Workers", in the Journal of Applied Psychology, Vol. 21, No.2, issue of April, p. 170.

situation³⁹. The validity of the test was found to be .65 when correlated with the personal history of employed clerical workers. The degree of relationship between the two tests is said to be .66⁴⁰. In a study by Pond⁴¹, the reliability of the instrument was given as .70, while the validity was in the order of .65. The correlation between Number and Name Checking was cited as .92 when corrected for attenuation.

In concluding this section, it would be well to note that the tests which were chosen as measures of each element of the hypothesis under study, are to be accepted at face value, so to speak. It is not the writer's contention to either prove or disprove what is measured by the tests, nor to debate their relative validities and reliabilities, although the latter have been taken into consideration. This may be taken to mean that the definitions offered by the authors of the tests were accepted as possibly only descriptions of what each one measures, and also that it is left to other researchers

39 D.M. Andrew and D.G. Paterson, Minnesota Clerical Test, Manual Revised in 1946 for the Minnesota Vocational Test for Clerical Workers, New York, Psychological Corporation, 1946, p. 3.

40 Id., ibid., p. 6.

41 M. Pond, "What is New in Employment Testing", in the Personnel Journal, Vol. 11, No.1, issue of June 1932, p. 11.

to examine and determine their validities and reliabilities.

The material which has been discussed thus far, then, consisted in a statement of the general procedure to be followed in testing the hypothesis, a word on the elimination of some of the sub-hypotheses, the selection of those elements which would be investigated, and finally, a discussion of the tests which were to be employed in this research to study those specific assumptions. To this end, a description of the Picture Arrangement subtest, the Primary Mental Abilities Verbal-Meaning and Reasoning tests, the Differential Aptitude Tests Abstract Reasoning, and the Minnesota Clerical Test was given and related to the sub-hypothesis each was to investigate, together with a brief mention of the literature for each one. Reasons for the choice of each of the tests were presented, and their reliabilities were reported as well as their validities when available.

The above comprises only part of the experimental design. The population and the statistical methods yet remain to be discussed and are treated at length in the following chapter.

CHAPTER III

THE EXPERIMENTAL DESIGN: POPULATION AND STATISTICAL METHODS

After the selection of instruments for the verification of the hypothesis had been made, the next step was to be the administration of the tests on a given population. The choice of individuals must be justified, and once selected, the methods of statistical analysis which are to be employed on the basis of the sample and hypothesis, are to be set down. These points will now be considered.

1. The Population

In focusing attention on the population, the method by which selection was carried out and the reasons for choosing a selected rather than a random sample, should be explained. The elements which were to be held constant were education, language, sex and the age range. A certain number of individuals was to be administered all of the tests. The description of the actual group which was used is also to be accounted for.

a) Education. -- The effects of the levels of education on Picture Arrangement is not definitely known, though there has been some suggestion that schooling has an influence on the results, as was reported in one of

the researches¹ in the field of Diagnostic Testing. Educational background was therefore controlled by limiting the choice of the sample to a senior High School population. All the pupils from grades eleven and twelve of one High School were to be used so as to obtain greater homogeneity.

b) Language. -- Since one of the tests used to examine the "verbal" element required a knowledge of English, only English-speaking subjects were to be used.

c) Sex. -- Jastak found that "The contents of the last three items are definitely more congenial to the thinking of women than men"². Because this conclusion was based on an item analysis of the records of 586 men and 586 women, it was deemed important to eliminate the influence of the sex factor by choosing a homogeneous group. An all male sample was to be used.

d) Age.-- Bolle³ found that the performance factor, which he named "B" and described as the ability to organize

1 Sol L. Garfield, "An Evaluation of the Wechsler-Bellevue Patterns in Schizophrenia", in the Journal of Consulting Psychology, Vol. 13, No.4, issue of August 1949, p. 286.

2 J. Jastak, "An Item Analysis of the Wechsler-Bellevue Tests", in the Journal of Consulting Psychology, Vol. 14, No.2, issue of April 1950, p. 92.

3 Arthur Bolle, A Factor Analysis of the Echelle D'Intelligence Ottawa-Wechsler, unpublished Master's thesis presented to the Institute of Psychology and Education of the University of Ottawa, 1954, p. 59.

and evaluate visual material, existed in Picture Arrangement between the ages of from fifteen to nineteen inclusively. From the findings, no major differences in factor loadings were found in this age range. Because of this, the sample in the present study was to be limited in age range and to consist of a group of fifteen to nineteen-year-olds.

e) The number of individuals. -- In order to reduce the standard error of the correlations which were to be computed, it was planned to include at least one hundred subjects in the present study.

Through the use of the above-mentioned principles of selection, the resulting group which was tested consisted of eighty-four third and fourth year High School English-speaking male students, ranging in age from fifteen years and four months to eighteen years and six months. Table I presents the age distribution of the group using an interval of four months per class. The mean age of the group was found to be 17.03 years with a standard deviation of 9.24 months or .77years.

A quick glance at the table will reveal two points. The first is that the distribution is bimodal. This could perhaps be explained in terms of the two grades, more individuals being older in grade eleven and more being younger in grade twelve, although the effect of bimodality

Table I.- Age Distribution of Sample in Intervals of Four Months.

Age		Frequency
18 yrs. 4 mos.	to 18 yrs. 7 mos.	4
18 yrs. 0 mos.	to 18 yrs. 3 mos.	5
17 yrs. 8 mos.	to 17 yrs. 11 mos.	15
17 yrs. 4 mos.	to 17 yrs. 7 mos.	14
17 yrs. 0 mos.	to 17 yrs. 3 mos.	9
16 yrs. 8 mos.	to 16 yrs. 11 mos.	10
16 yrs. 4 mos.	to 16 yrs. 7 mos.	11
16 yrs. 0 mos.	to 16 yrs. 3 mos.	7
15 yrs. 8 mos.	to 15 yrs. 11 mos.	5
15 yrs. 4 mos.	to 15 yrs. 7 mos.	4
Total Number		84
Arithmetic Mean		17.03 yrs.
Standard Deviation		9.24 mos.

may be due to the size of the interval used. Secondly, if a curve were to be fitted, the result would be a negatively skewed one, indicating that the sample was on the older side.

Though some of the students were bilingual, all the subjects had a good knowledge of English. The distribution of scores on the Primary Mental Abilities Verbal-Meaning subtest justified the description of the group as English-speaking.

It had been hoped that one hundred students would be included, but since the group tests were administered over a period of approximately six weeks, some of the students were automatically disqualified by missing a day of school when one or more of the group tests was to be given. The testing sessions were so organized that they coincided with the school's testing and guidance programme, thus making the testing appear to be part of the regular school routine.

Sixteen individuals were eliminated for failing to take one or more of the group tests. However, it does not seem that selective factors were operating, judging from the results obtained on the tests which these individuals did take. It would appear that the usual absenteeism was present rather than direct refusal to submit to the battery. In all but five cases, only one test was missed; but

unfortunately two of the tests consisted of two parts, each of which was given at one sitting. This means that actually two tests were missed in one period when one of these double tests was administered.

From the method of selection of the sample according to education, sex, age, language and school, it can be readily seen that the aim was to obtain as homogeneous a group as possible, the purpose being the elimination of any possible effects that might be caused by these factors. As a result, were any differences to be evidenced, it could be inferred that greater differences would be found in a heterogeneous group when the results were to be interpreted. The sample which was finally used approximated the requirements quite closely.

2. Statistical Methods

The statistical methods of analysis took into account the general purpose of this study, namely, the discovery of the degree of relationship existing between Picture Arrangement and each of the elements hypothesized. The specific aims were the discovery of:

1. The degree of relationship between each of the tests in the battery and each of the Picture Arrangement scores;

2. The degree of relationship between all the tests of the battery taken together and each of the Picture Arrangement scores;

3. The contribution of each variable to the total predicted variance in each of the Picture Arrangement scores.

a) The degree of relationship between Picture Arrangement accuracy, time bonus and total scores individually and the scores on each of the tests selected to study the sub-hypotheses was to be investigated. A correlational approach was therefore indicated. The Pearson product-moment coefficient was to be used on the condition that linearity of regression and homoscedasticity were present. The raw-score formula was to be used:

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{N\sum X^2 - (\sum X)^2} \sqrt{N\sum Y^2 - (\sum Y)^2}} .$$

The standard error of an assumed simple correlation of zero was to be found by

$$\sigma_{r_0} = \frac{1}{\sqrt{N - 1}} ,$$

while for the obtained significant relationships and those approaching significance, the variability was to be found by

$$\sigma_r = \frac{1 - r^2}{\sqrt{N - 1}} .$$

For the multiple correlations, the standard error was to be found by using the formula:

$$\sigma_R = \frac{1 - R^2}{\sqrt{N - m}}$$

which is used with relatively small groups.

Each simple and multiple correlation was to be tested for significance according to the formulae:

$$F_r = \frac{r_{12}^2}{(1 - r_{12}^2)/(N - 2)} \quad \text{or} \quad F_R = \frac{R^2/m}{(1 - R^2)/(N - m - 1)}$$

respectively. One other formula was to be employed for the multiple correlations. Since these are usually considered inflated values, a correction for each was to be made using the formula:

$$cR^2 = 1 - (1 - R^2) \left(\frac{N - 1}{N - m} \right)$$

thus shrinking them to a more probable population value.

b) The degree of relationship of all the tests taken together with each of the three Picture Arrangement scores was to be found, thus requiring the computation of two multiple correlation coefficients. The Doollittle method was to be employed.

c) Finally, the contribution of each variable to the total predicted variance in Picture Arrangement accuracy, time bonus and total scores was to be discovered through the analysis of the beta weights in the multiple correlation formula. The formula states, in terms of the coefficient

of multiple determination, that

$$R^2 = \cancel{\beta}_{12} r_{12} \cancel{\beta}_{13} r_{13} \cancel{\beta}_{14} r_{14} \cancel{\beta}_{15} r_{15} \cancel{\beta}_{16} r_{16} \cancel{\beta}_{17} r_{17}$$

In testing the significance of the results, the accepted levels of confidence were described as

i) significant if the simple or multiple coefficient is 2.58 times as great as the standard error of the correlation, placing it at the 1% level of probability, and

ii) approaching significance if the simple or multiple correlation is 1.96 times as great as the standard error, placing it at the 5% level of confidence.

In the former instance, a simple coefficient of .2830 would be necessary to be statistically significant, while in the latter, one of .2146 would be required to be considered as approaching significance. For the multiple correlations, the necessary coefficient would vary according to what was obtained because of the slightly different formula used.

The discussion of significance which has been presented dealt with the number of standard deviations away from the mean, that is, 2.58 or 1.96 times the deviations obtained. This method of evaluating the significance has been extensively used and was included for this reason. However, another method for determining significance will also be employed, namely, the analysis of variance or what

is commonly called the "F" test. The use of a table of "F's" will determine the significance of any obtained relationship, simple or multiple, at the same levels of confidence.

The points which have been discussed, then, have covered both the population and the statistical methods. With reference to the former, the questions of education, language, sex, age, school, the final group that resulted and the number of drop-outs have been examined. In the treatment of the statistical methods, the choice of the formulae and procedures were explained in relation to the hypothesis and the population. A simple correlational approach, followed by multiple correlation was needed. Finally, the criteria for statistical significance were presented, taking into account the probable inflated value of the multiple coefficients.

Now that all the preliminaries have been considered, the first step towards the realization of the purpose of this research will centre around the examination and discussion of the results obtained.

CHAPTER IV

PRESENTATION AND ANALYSIS OF DATA

The foregoing discussion of the population and statistical methods necessary in the treatment of the data now ushers in an examination of the results obtained. The primary aim of this section is to discover the relationship of each test to the three separate scores of the Picture Arrangement subtest, namely, accuracy, time bonus and total. This will be followed by a study of the extent of the relationship of all the tests taken together to each of the Picture Arrangement subscores. Finally, a discussion of the contribution of the independent variables to the total variance in Picture Arrangement ensues.

1. The Data

The raw data was tabulated in order to calculate the relationships existing between and among the tests. Since the raw scores as such have no direct bearing in this discussion, they will be found in the appendix for reference. However, a resume of the raw scores will be found in Table II. The range of scores, the means and standard deviations for each test administered to the group were included.

Table II.- Range of Scores, Means and Standard Deviations for Each Test of the Battery Using a Sample of Eighty-Four Individuals.

Test	Range	Mean	Sigma
P. A.			
Accuracy	2 to 15	9.37	2.73
Time Bonus	0 to 5	.94	.65
Total Score	2 to 19	10.31	3.68
P. M. A.			
Verbal-Meaning	7 to 48	26.80	9.53
Reasoning	3 to 26	15.36	5.61
D. A. T.			
Abstract Reasoning	8 to 45	28.61	11.42
Minnesota Clerical			
Number Checking	67 to 140	101.48	16.43
Name Checking	44 to 167	107.13	21.49
Minnesota Clerical			
Error Scores in			
Number Checking	0 to 10	1.88	1.98
Name Checking	0 to 20	6.27	4.29

a) Relationship of each test to Picture Arrangement accuracy, time bonus and total scores. -- From the raw scores, the necessary correlations were computed according to the formula cited in the above description of the statistical methods.

Each of the coefficients was then tested for significance and linearity. Basing the regression function for each of the three dependent variables (the three separate scores of Picture Arrangement) respectively on each of the independent variables (the other tests), no evidence was found to reject the hypothesis of linearity although in one case (the relationship between P.M.A. Reasoning and Picture Arrangement accuracy) there was a tendency for departure from linearity. On the other hand, basing the regression function for each of the independent variables on the same dependent variables respectively, evidence for rejecting the hypothesis of linearity was found only in one case with only a tendency towards curvilinearity in two cases. However, the interest here centers on the former statement.

One further consideration concerns the concept of homoscedasticity. From the scattergrams it was evident that heteroscedasticity was present when the Error Scores in Number and Name Checking of the Minnesota were compared to Picture Arrangement accuracy and total scores, and more

so when Picture Arrangement time bonus score was compared to each of the independent variables. In the former case this was of little importance because the obtained coefficients were not statistically significant, but in the latter, a correlation of .2340 (between Picture Arrangement time bonus and Verbal-Meaning) was found to approach significance (P.05). Nevertheless, it may still be used to indicate the degree of relationship. However, because of heteroscedasticity in all the relationships obtained between the time bonus score and each of the other variables, and also because of the non-significant coefficients, this part of the experimental design was abandoned. Thus, only two dependent variables remain, that is, Picture Arrangement accuracy and total scores.

After these considerations, the intercorrelations were then obtained and tabulated. Since the interest now centers around the two divisions of Picture Arrangement, only the relationships between each of these and each of the tests in the battery are included. These will be found in Table III. The complete tabulation has been included in Appendix I. The standard error of a coefficient of zero for the particular sample used is .1097.

The significance of each relationship found in the Table was checked by means of the F test. Four correlations were found to be statistically significant at the 1% level

Table III. - Correlations Between the Various Elements of the Test Battery and the Two Dependent Variables, Using a Sample of Eighty-Four Individuals.^a

Test	P.M.A. Verbal- Meaning	P.M.A. Reason- ing	Min. Numb. Check.	Min. Name Check.	D.A.T. Abst. Reas.	Error in Numb. Check.	Min. Name Check.
P.A. Acc. Score	.1700	.3189 ^b	.1171	.0456	.3510 ^b	-.1260	-.1730
P.A. Total Score	.2130 ^c	.2750 ^b	.1432	.0770	.3190 ^b	-.0920	-.1820

a The Standard Error for a correlation coefficient of zero for an N of 84 is .1097.

b Significant at the 1% level of confidence.

c Significant at the 5% level of confidence.

of probability and one at the 5% level. A definite relationship was found to exist between the Picture Arrangement accuracy score and the P.M.A. Reasoning and D.A.T. Abstract Reasoning tests. The relationship between Picture Arrangement total score and P.M.A. Reasoning and D.A.T. Abstract Reasoning was found to be significant at 1% while that between the former and P.M.A. Verbal-Meaning was found to be approaching significance at 5%.

Where the standard error of a coefficient of zero is used, the majority of the statements in the above paragraph remain true. The one exception is that of the relation between Picture Arrangement total score and P.M.A. Reasoning which was significant beyond the 5% probability level and was very close to the 1% level of confidence.

b) Relationship of all tests together to Picture Arrangement. -- Once all of the interrelationships were found, the Doolittle method of multiple correlation was used to find the two multiple coefficients, that is, one for each of the Picture Arrangement subscores.

A multiple coefficient (R) of .4065 was obtained when Picture Arrangement accuracy was compared to all of the other tests, while one of .3780 was had when the total score was compared with the same test battery.

Each of the multiple coefficients was checked for significance by means of the F test. As can be seen from

Table IV, neither was found to be significant at the 1% nor at the 5% level of confidence. However, if the standard error of a multiple coefficient is used as the criterion of significance, both are found to be significantly different from zero. When the "t" test is applied, a multiple correlation of .502 would be necessary for significance at 1% for both variables.

c) Contribution of variables to variance of Picture Arrangement. -- As was seen earlier, the multiple correlation formula would yield the amount of variance loading of each test in Picture Arrangement. Since two multiple coefficients were found, two such equations will be considered, one for the accuracy score and one for the total score.

1) If Picture Arrangement accuracy score is considered,

$$R^2 = .0008 + .0628 + .0195 - .0084 + .0804 + .0054 + .0048 = .1653$$

Taking each relationship which was found to be significant at 1% by means of the F test, it was seen that the P.M.A. Reasoning test explained 6.3% of the variance in the Picture Arrangement accuracy score. The D.A.T. Abstract Reasoning test controlled 8% of the variance. It was also found that the total variance loading of 16.5% was the result of all the relationships taken together-- those which were found to be significant as well as those which were not. Since only 14.3% can be claimed as being accounted for, the other

Table IV. - Test of Significance for the Obtained Multiple Correlation Coefficients for Each Dependent Variable.

Dependent Variable	Multiple Correlation	Standard Error of R	Variance Ratio
P.A. Acc. Score	.4065	.0957	1.8 ^a
P.A. Total Score	.3780	.0983	1.6 ^a

a Not statistically significant.

2.2% remained doubtful. In the final analysis it can be said that while 14.3% of the variance loading in the accuracy score was accounted for, 85.7% was still to be explained.

ii) With Picture Arrangement total score as the dependent variable,

$$R^2 = .0184 + .0331 + .0270 - .0125 + .0624 + .0085 + .0060 = .1429 .$$

The F test of significance was again used and relationships were found at the 1% and 5% levels of confidence. In the first instance, the D.A.T. Abstract Reasoning controlled 6.2% of the variance while P.M.A. Reasoning accounted for 3.3% of the variance loading. In the second case, P.M.A. Verbal-Meaning explained 1.8% of the variability. The sum total of the variance accounted for by the three tests was 11.3%, leaving 3.6% doubtful. This would mean that 88.7% of the variance loading in Picture Arrangement total score still remained to be controlled.

Up to this point, the results have been considered under various aspects. The raw data was summarized in tabular form in which the range of scores, means and standard deviations for each test were included. A second table, which comprised the various correlations between each of the two dependent variables and each of the tests, was also introduced. The last table presented the obtained R's with their respective standard errors and variance

ratios, with the size of relationship needed for significance noted in the text. The significant correlations and those approaching significance were singled out. Finally, consideration was given to the variance controlled by each of the tests found to contribute significantly in each of the dependent variables. The amount of variance which remained unexplained for each of the two variables was also computed.

Now that all of the data has been presented, there still remains the question of giving meaning to the results which were obtained.

2. Analysis of Data

Carrying the discussion of the obtained results one step further, the significant correlations and those which approached significance need to be given a more precise explanation as to the relation of each of the Picture Arrangement subdivisions to the tests involved.

a) Picture Arrangement accuracy score. -- Two relationships were found to be significant at 1%, the first being P.M.A. Reasoning with a correlation of $.3189 \pm .0986$, the other the D.A.T. Abstract Reasoning with a correlation of $.3510 \pm .0962$ when compared to the accuracy score. The factors which are supposedly tapped by those two tests can be considered as also being contained in the accuracy score, that is, the accuracy score measures "The ability to solve

logical problems-- to foresee and plan"¹ as well as "The ability to perceive relationships, ...generalization, and education of principles from nonlanguage designs"². The reasoning factor appears to be definitely present. There is also a question of arranging the pictures in a logical order, perception of relationships in and among the pictures, foresight and planning in the sense that an individual must look ahead in placing the pictures in an order which will make sense, and the discovery of a rule or principle involved in arranging them in the proper order.

b) Picture Arrangement total score. -- Two significant relationships were obtained at the 1% level of confidence while one approached significance at 5%. The D.A.T. Abstract Reasoning correlated $.3190 \pm .0986$ with Picture Arrangement total score, while P.M.A. Reasoning correlated $.2750 \pm .1015$. At 5%, P.M.A. Verbal-Meaning correlated $.2130 \pm .1048$. Since the total score includes the accuracy score, the discussion in the preceding subsection also applies here. Additionally, since the relation with Verbal-Meaning approached significance, it is probable that the

¹ L.L. Thurstone and T.G. Thurstone, Examiner Manual for the SRA Primary Mental Abilities, Preliminary Edition, Chicago, Science Research Associates, 1947, p.1.

² G.K. Bennett, H.G. Seashore and A.G. Wesman, A Manual for the Differential Aptitude Tests, New York, Psychological Corporation, 1952, p. 72.

"...ability to understand ideas expressed in words"³ is also present.

Before concluding this discussion, one fact remains to be explained. It was noted that though significant simple correlations were obtained, they were relatively low. It will be recalled that the sample chosen was a homogeneous one so that were any differences evidenced, the differences would have been much greater in a heterogeneous population. In other words, the fact that a homogeneous group was used tended to restrict the variability and, as a result, to lower the relationships. But with a heterogeneous or random sample there would probably have been greater variability, and thus, higher correlations would have been obtained.

Nothing has been said of the multiple relationships because they were not found to be significant when the F test was used. When the obtained multiple coefficients were compared to their standard errors, the correlations between the test battery and Picture Arrangement accuracy and total scores respectively were found to be significantly different from zero. However, the "t" test was not found to be significant. These correlations too would very likely have been higher with a heterogeneous group. It may

3 L.L. Thurstone and T.G. Thurstone, op. cit., p. 1.

also be that the F test was too stringent under the circumstances since the homogeneous group already tended to decrease the errors. Thus it may be stated that if they were considered as being significant, much of the variance still would remain unexplained.

Concerning negative findings, none were present as such. However, because of the type of distribution of scores obtained in the Picture Arrangement time bonus variable, this aspect could not be investigated as to its possible contribution. On the other hand, the findings were inconclusive with respect to the attention and distractibility elements, thus calling for further study in ascertaining their influence on the subtest.

In summing up, the results obtained, the significance of the simple and multiple correlations, the variance controlled, and the meaning of the results were presented and discussed. In addition, a word was said on the size of the relationships obtained and how they might apply to an unselected or random population. Finally, the inconclusive findings were noted.

SUMMARY AND CONCLUSIONS

This report has attempted to show something of the rationale of the Picture Arrangement subtest of the Wechsler-Bellevue Intelligence Scale, Form I. As was seen, the writer undertook this study at the instigation of those Authors who saw the shortcomings of interpreting the Wechsler subtests diagnostically according to the existing hypotheses.

Thus, a survey of the literature was made to unearth as many of the suggested assumptions as possible which were in any way related to the subtest. As a result, an examination of some of the short forms, of scatter and pattern analysis, and of the theoretical and experimentally investigated rationales was made. Nine hypotheses were proposed.

The second chapter was concerned with one part of the procedure necessary for the investigation of the specific hypotheses. However, because of the lack of valid and reliable measuring devices, not all of the suggested rationales could be studied. For those which could be examined, the necessary instruments were introduced, related to each of the specific hypotheses, and described. Chapter III considered the remaining portion of the experimental design. A description of the population which would be necessary for this particular study was described,

taking the factors of education, language, sex and age into consideration. The group which was actually used was found to compare favorably with the theoretical requirements. The statistical methods of analysis were also introduced, basing them on the hypothesis and sample used.

In the last chapter, the obtained data was presented. The relationship of each test to Picture Arrangement accuracy, time bonus, and total scores was discussed. A test for linearity was made and was generally found to hold, but heteroscedasticity was found to prevail with the time bonus variable, thus eliminating it from further consideration. The statistically significant relationships were singled out and identified. The overall relationship of all the tests taken together to Picture Arrangement accuracy and total scores were also presented and checked for significance. The contribution made by each of the independent variables which correlated significantly was also examined. Finally, the discussion of the results was carried one step further by giving meaning to the obtained data. A definite, though low, relationship was found to exist between the Picture Arrangement accuracy and total scores and both of the reasoning tests used, thus indicating that the reasoning element was present in the arrangement of the pictures. Only a trend towards significance was present when Picture Arrangement total score was compared to the Verbal-Meaning

test. Thus there was apparently insufficient evidence for rejecting the hypothesis that verbal ability is present in Picture Arrangement. A final note pointed out that the findings were inconclusive with reference to the study of the attention and distractibility elements, and that further investigation of their possible influence is therefore indicated.

From all that has been said, three conclusions can apparently be drawn. Firstly, Picture Arrangement includes a reasoning element, secondly it measures the ability to recognize and establish relationships, and finally, it is probable that it also includes a verbal factor. It should be remembered, however, that the relationships, though significant, were rather low and explained only from ten to fourteen percent of the variance. The above conclusions could be generally applicable to a male population in the fifteen to nineteen-year-old age range.

At this point some suggestions for further research on the Picture Arrangement subtest seem to be in order. Many of the hypothetical rationales still remain unexplored, including those which were investigated here and found not to hold. However, a first suggestion would entail the construction of another set of pictures which would eliminate the "human social situations". The two series could then be compared to yield an indirect measure of the

presence or absence of "social intelligence" in Picture Arrangement. A second idea would require a change of directions in the administration of the subtest. It might prove more profitable to have each individual verbalize the arrangement he has chosen, thus giving rise to a thematic approach. This may be more fruitful than the mere arrangement of the pictures in definitely establishing the presence of the verbal factor. Finally, intratest patterning of the component elements together with pattern analysis of the items may possibly furnish what intertest pattern and scatter analysis have failed to yield.

BIBLIOGRAPHY

Balinski, Benjamin, "An Analysis of the Mental Factors of Various Age Groups from Nine to Sixty", in the Genetic Psychology Monographs, Vol. 23, First Half, issue of February 1941, p. 191-234.

A study of the Wechsler-Bellevue consisting of an analysis of the various factors present in different age groups, their appearance, disappearance, and replacement at various periods of time. It is suggested that Picture Arrangement measures a verbal and a non-verbal or performance component at age fifteen.

Bolle, Arthur M., A Factor Analysis of the Echelle D'Intelligence Ottawa-Wechsler, unpublished Master's thesis presented to the Institute of Psychology and Education of the University of Ottawa, 1954, xi-60 p.

An analysis of the factors present in the subtests revealed that different components were present at various age levels for the same subtests. It is thought that Picture Arrangement measures recognition and establishment of relationships, and verbalization between the ages of fifteen to nineteen inclusively.

Candee, Beatrice, and Milton Blum, "A New Scoring System for the Minnesota Clerical Test", in the Psychological Bulletin, Vol. 34, No.8, issue of October 1937, p. 545.

A critical note on the scoring technique of the test, suggesting that the error score is not due to guessing but to a lack of attention, thus furnishing a clue to a method for measuring distractibility.

Cohen, Jacob, Multiple Factor Analyses of the Wechsler-Bellevue Intelligence Scale Performance of Schizophrenic, Psychoneurotic, and Brain-Damaged Groups, microfilmed Doctor's thesis presented to the University of New York, New York, March 1950, vii-162 p.

Using a neuropsychiatric population, the various subtests were found to contain a variety of factors. Picture Arrangement was thought to include attention, distractibility, reasoning, non-verbal organization, and a verbal element.

Garfield, Sol, "An Evaluation of Wechsler-Bellevue Patterns in Schizophrenia", in the Journal of Consulting Psychology, Vol. 13, No.4, issue of August 1949, p. 279-287.

The neglect of sampling problems, of criteria for diagnostic classification of subjects, and the use of various methods of study which lead to incomparable results in a great number of investigations were criticized. Factors such as age, education I.Q., and proper classification were found important influences to be controlled.

Jastak, Joseph, "An Item Analysis of the Wechsler-Bellevue Tests", in the Journal of Consulting Psychology, Vol. 14, No.2, issue of April 1950, p. 88-94.

An item analysis of all the subtests revealed that for Picture Arrangement the order of difficulty did not differ from the standard presentation but that the influence of sex was in evidence for the last three series.

Rapaport, David, Merton Gill and Roy Schafer, "The Bellevue Scale", in Diagnostic Psychological Testing: The Theory, Statistical Evaluation, and Diagnostic Application of a Battery of Tests, Vol. 1, Chicago, Year Book Publishers, 1945, p. 44-318.

A description of the structure, material, and administration of the Wechsler, statistical demonstration of significance of scatter, the rationales of the subtests, analysis of the types of failures in individual records, and statistical analysis of data on control and clinical populations are presented. Pages 214-222 are concerned with the Picture Arrangement subtest which is thought to measure attention, judgement, planning and anticipation.

Schofield, William, "Critique of Scatter and Profile Analysis of Psychometric Data", in the Journal of Clinical Psychology, Vol. 8, No.1, issue of January 1952, p. 16-22.

The rationales, purposes of pattern study, and deficiencies of research designs are given critical consideration. Age, education and I.Q. are factors to be controlled.

Watson, Robert I., "The Use of the Wechsler-Bellevue Scales: A Supplement", in the Psychological Bulletin, Vol. 43, No.1, issue of January 1946, p. 61-68. Education and cultural background are thought to influence findings in research even if the subtests were pure functional entities. Further study of the rationales is urged, however, and two methods of investigation are suggested.

Wechsler, David, The Measurement of Adult Intelligence, Third Edition, Baltimore, Williams and Wilkins Company, 1944, vii-258 p.

A manual for the administration, scoring and interpretation of the Wechsler. The rationale suggested for Picture Arrangement is that it measures general intelligence as applied to social situations or what other writers have referred to as social intelligence.

APPENDIX I

THE RAW DATA

The raw data which has been tabulated on the following pages was obtained from tests whose scores were derived according to the methods devised by their respective Authors. The one category which is not normally employed refers to the error scores on the Minnesota Number and Name Checking test. However, the scores for Number and Name Checking are the result of the right-minus-wrong formula.

Since the total score on the Picture Arrangement subtest is a composite of accuracy and time bonus, this latter was omitted but may be obtained by subtracting the accuracy from the total score.

The symbols used as column headings in Table V are explained as follows:

Code: subject's identification number;
T1 : Picture Arrangement accuracy;
T2 : Picture Arrangement total;
T3 : P.M.A. Verbal-Meaning;
T4 : P.M.A. Reasoning;
T5 : Minnesota Number Checking;
T6 : Minnesota Name Checking;
T7 : D.A.T. Abstract Reasoning;
T8 : Number Checking Error Score;
T9 : Name Checking Error Score.

The correlations which were computed between and among all of the variables have been included in Table VI.

Table V. - Raw Data obtained from the Test Battery.

Code	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9
1	7	7	19	10	114	123	1	9	15
2	7	7	19	23	116	129			43
3	13	15	28	6	113	112	1	20	15
4	12	13	15	10	104	99	4	3	25
5	5	5	15	3	78	64	10	20	11
6	6	6	37	18	94	118		1	34
7	9	9	26	14	105	124	3	9	23
8	10	10	19	17	71	69	3	10	20
9	12	14	19	11	100	132	5	3	24
10	8	8	19	20	83	99	4	4	34
11	9	9	21	5	106	107	5	10	26
12	11	14	18	10	97	110	4	9	33
13	10	10	15	14	108	121	2	5	20
14	11	11	21	16	110	91	4	15	36
15	8	11	20	5	99	120		3	0
16	9	12	33	19	116	126	2	7	20
17	3	3	21	18	98	114	3	7	26
18	6	6	7	10	84	106	4	11	2
19	11	14	34	24	111	116	2	7	43
20	6	6	42	14	90	128	2	4	24
21	12	15	38	19	110	121		3	35
22	9	9	33	22	92	119		4	33
23	9	10	37	10	93	84	2	8	18
24	11	14	34	11	90	106	5	4	35
25	8	8	39	15	89	110	2	8	33
26	10	10	32	26	139	142	1	3	24
27	10	10	17	9	122	138	4	8	11
28	7	7	17	9	106	100	6	13	19
29	6	6	22	9	134	137	1	10	19
30	6	6	16	15	106	111	6	13	25
31	12	15	28	21	103	128	1	1	38
32	13	15	33	26	91	111	2	8	44
33	8	8	43	20	102	120	1		45
34	14	19	32	23	105	119	1	5	28
35	12	15	25	12	119	109		2	29

Table V. - Raw Data obtained from the Test Battery.
(Continued)

Code	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9
36	13	14	23	20	140	132		8	40
37	7	10	25	10	98	116		3	27
38	10	12	12	13	103	85	2	8	25
39	9	9	30	14	107	129	1	7	33
40	12	15	31	19	120	131		2	41
41	10	10	18	11	78	65		2	-8
42	9	9	14	14	89	93	6	4	22
43	4	4	19	3	97	102	1	15	7
44	13	15	21	10	83	62		12	30
45	6	6	26	16	71	75	2	11	33
46	9	9	13	15	121	99	1	5	41
47	9	12	29	16	80	66	2	14	36
48	10	10	26	14	95	69	1	2	30
49	15	18	25	15	121	84		4	27
50	2	2	18	3	102	80		6	7
51	4	4	13	3	83	44		12	2
52	9	9	23	13	102	104	2	5	24
53	12	12	20	14	76	74		3	29
54	10	10	17	12	125	105		2	30
55	12	14	36	25	112	126		1	29
56	10	10	46	23	117	123	4	8	31
57	7	7	26	10	106	108	4	6	21
58	12	12	20	23	81	78	3	8	43
59	13	14	28	14	106	103	2	6	37
60	12	12	30	22	102	101	4	3	40
61	10	10	23	17	67	110	1	15	28
62	10	13	45	20	99	97	1	6	41
63	7	7	31	14	82	95		5	35
64	6	6	12	13	96	81	4	11	33
65	13	15	35	20	85	116		7	28
66	14	19	33	16	111	128	5	5	42
67	10	10	15	14	126	119	2	5	39
68	8	11	40	18	106	106		1	36
69	8	8	39	17	113	116		6	34
70	10	10	44	21	89	88	1	5	43

Table V. - Raw Data obtained from the Test Battery.
(Continued)

Code	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9
71	6	6	43	16	128	137	4	7	29
72	6	6	32	20	116	167	1	2	22
73	13	16	37	19	121	120	1	7	36
74	8	8	30	21	78	120	1	4	38
75	8	8	48	10	90	96	1	2	38
76	9	9	36	25	120	120		3	40
77	10	10	20	20	81	83		8	28
78	8	8	30	21	95	109	5	7	41
79	13	15	30	13	128	121		4	45
80	12	14	47	16	86	100			39
81	9	9	31	20	99	128	1	2	29
82	5	5	23	21	120	126		1	35
83	12	14	18	17	102	100	1	3	1
84	13	13	26	15	71	99	3	7	36

Table VI.- Correlations Between and Among the Dependent and Independent Variables.

	P. M. A. Verbal Meaning	P. M. A. Reasoning	Minnesota Number Checking	Minnesota Name Checking	Errors in Number Checking	Errors in Name Checking	D. A. T. Abstract Reasoning
P.A. Acc.	.1700	.3189	.1171	.0456	-.1260	-.1730	.3510
P.A. Total	.2130	.2750	.1432	.0770	-.1502	-.1820	.3190
P. M. A. Verbal Meaning		.4590	.0070	.3097	-.2734	-.3318	.4808
P. M. A. Reasoning			.1229	.3474	-.2260	-.4146	.6223
Minnesota Number Checking				.6200	-.1503	-.2133	.1257
Minnesota Name Checking					-.0915	-.3382	.1926
Errors in Number Checking						.4370	-.0767
Errors in Name Checking							-.3091

APPENDIX 2

ABSTRACT OF

A Study of the Rationale of the Wechsler-Bellevue Picture Arrangement Subtest, Form I¹

The aim of this study was to determine the functions tapped by the Picture Arrangement subtest of the Wechsler-Bellevue and the extent to which these particular elements were present. A review of the literature suggested that the subtest measured attention, recognition and establishment of relationships, foresight and planning, judgement, reasoning, and social intelligence in addition to a verbal element.

However, because of the lack of suitable instruments, the investigation of foresight and planning, judgement, and social intelligence could not be undertaken. The remaining areas were studied through the use of the P.M.A. Verbal-Meaning and Reasoning, the D.A.T. Abstract Reasoning, and the Minnesota Clerical Test.

The sample selected consisted of eighty-four senior High School English-speaking male students, ranging in age from fifteen years and four months to eighteen years and six months. The test scores of these individuals were then

¹ M.A. Thesis presented by Pascal J. Delli Colli, in 1957, to the School of Psychology and Education of the University of Ottawa, 81 pages.

tabulated and studied to determine the intercorrelations. Multiple correlations were also computed between the test battery and the Picture Arrangement accuracy score and the Picture Arrangement total score. Analysis of the beta weights revealed the relative contribution of the component elements.

From an analysis of the data, it was concluded that Picture Arrangement apparently measures both reasoning and the ability to recognize and establish relationships. Moreover, it appears likely that this subtest taps a verbal factor in the particular group examined. The study of the attention and distractibility factors was inconclusive.