THE FUNCTIONING OF FAMILIAL AND ORGANIC MENTALLY DEFECTIVE CHILDREN ON THE STANFORD-BINET, L-M, WISC, AND GOODENOUGH DRAW-A-MAN TESTS OF INTELLIGENCE: A COMPARATIVE STUDY

Thesis for the Degree of M. A.

MICHIGAN STATE UNIVERSITY

Frederick William Rohrs

1961

ABSTRACT

THE FUNCTIONING OF FAMILIAL AND ORGANIC MENTALLY DEFECTIVE CHILDREN ON THE STANFORD-BINET,L-M, WISC, AND GOODENOUGH DRAW-A-MAN TESTS OF INTELLIGENCE: A COMPARATIVE STUDY

by Frederick William Rohrs

The purpose of this study was to determine the correlational relationships and mean differences between the Stanford-Binet, Form L-M (SB,L-M), the Wechsler Intelligence Scale for Children (WISC), and the Goodenough Draw-A-Man Test (Goodenough) using mentally retarded children of organic and familial etiologies. Differential SB,L-M functioning of these two types of subjects was also studied.

Two groups of institutionalized children were selected; twenty in the Familial Group and twenty-six in the Organic Group. Groups were equated by previous IQ, sex, chronological age, and length of institutionalization. Subjects were selected according to the following criteria: a) previous IQ between 50 and 70, b) current enrollment in institutional school programs, c) not crippled, blind, deaf, untidy, or subject to seizures, d) not tested during the previous year.

The three intelligence tests were administered in counter-balanced order and in accordance with the standardized instructions found in their respective manuals. Two recently reported minor scoring innovations were used where applicable in the calculation of IQs. Differences in organic and familial SB,L-M functioning were examined by performance on individual test items and by patterns of item failure ("scatter").

Findings for the Familial Group were:

a) SB,L-M vs. WISC correlations significant except in the case of the Performance IQ, and, contrary to expectation, a significantly lower mean WISC Full Scale IQ.

b) SB,L-M vs. Goodenough correlation and mean difference non-significant.

c) Goodenough vs. WISC correlations significant for the WISC Full and Performance Scales, and a significantly higher mean Goodenough IQ as compared with the WISC Full Scale IQ.

Findings for the Organic Group were:

a) SB,L-M vs. WISC correlations all significant and, contrary to expectation, a significantly lower WISC Full Scale IQ.

b) SB,L-M vs. Goodenough correlation and mean difference non-significant.

c) Goodenough vs. WISC correlations significant for the WISC Full and Performance Scales, but no significant mean IQ differences.

The following differential functioning between Organic and Familial Groups was found:

a) Organics exhibited significantly greater SB,L-M "scatter" than did familials, scatter being measured by number of intra-range item failures and by number of intra-range age levels.

b) Only one SB,L-M item distinguished between the groups, VIII,2 Memory for Stories: The Wet Fall. Failures of this item were by familials and a repressive tendency was suggested as the cause of failure.

c) SB,L-M mean IQ was significantly higher for organics than for familials. No significant organic vs. familial mean IQ differences were found on the WISC or Goodenough.

Approved: Gilbert R. Hawes
Thesis Advisor

Date: July 10, 1961

THE FUNCTIONING OF FAMILIAL AND ORGANIC MENTALLY DEFECTIVE CHILDREN ON THE STANFORD-BINET, L-M, WISC, AND GOODENOUGH DRAW-A-MAN TESTS OF INTELLIGENCE: A COMPARATIVE STUDY

By

Frederick William Rohrs

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Psychology

1961

ACKNOWLEDGEMENTS

The author is greatly indebted to Mr. Peter Ziebell, Chief Psychologist at the Lapeer State Home and Training School and to the staff members of that institution for their help in obtaining the subjects used in this study. Also very much appreciated are the many helpful ideas in the planning and execution of this study supplied by Dr. Mary R. Haworth, chairman of the thesis committee and by the committee members, Dr. Charles F. Wrigley and Dr. Gerald F. King.

Frederick W. Rohrs

TABLE OF CONTENTS

LIST OF TABLES

APPENDIX TABLES

INTRODUCTION

Whenever a new psychological test is published, the task becomes necessary to determine how this test is related to the older, more well-established instruments in the same general area. Such a task is imperative in the area of intelligence testing because of the strong emphasis placed upon the "IQ" as a measure of intellectual assessment by both professional and lay personnel alike. Whether this emphasis is warranted or not is a point of doubt today, particularly in regard to the individual case. Since, however, this emphasis does exist, psychologists must know what relationships prevail between their various instruments of intellectual assessment. The general task of this study is to help determine such relationships.

Specifically, the main purpose of this study is to determine the correlational relationships and mean differences between the new, Third Revision of the Stanford-Binet, the Form L-M (1960), henceforth called the SB,L-M, as compared with two older tests of intelligence, the Goodenough Draw-A-Man Test (1926) and the widely used Wechsler Intelligence Scale for Children (1949), henceforth called the Goodenough and WISC respectively. This comparison will be limited to the lower end of the intelligence distribution and will use as subjects two groups of mentally defective children, one of familial etiology and the other of

various organic etiologies. For the most part, the emphasis will be upon comparative test performance within the etiologically classified groups. However, when appropriate, differential test functioning between the Familial and Organic Groups will also be studied.

It is hoped that the results will help determine whether IQs as measured by the above tests have the degree of numerical equivalence often accorded them in actual work situations with mental defectives. Also, through the examination of differences in the more subtle quantitative aspects of SB,L-M performance, specifically "scatter" and the differentiating power of individual items, it is hoped that aids in differential diagnosis between mental defectives of familial and organic etiologies might be established.

Since the SB,L-M is the new test, emphasis will be placed upon it and its comparisons with the other two tests. However, comparisons between the WISC and the Goodenough are also reported as a matter of scientific interest since only one previously published study could be found which related the functioning of defectives on these two particular tests.

BACKGROUND OF THEORY AND RESEARCH

In this section is presented a) a brief discussion of the theoretical backgrounds of the SB, the WISC, and the Goodenough with emphasis on the types of intelligence which these tests are purported to measure, b) definitions and explanations of the familial and organic mental defective etiologies, c) a review of previous studies relating 1937 SB IQs with WISC and Goodenough IQs using both mental defectives and normal subjects, and d) a discussion of previous studies making use of SB scatter and individual items for differentiating between familial and organic mental defectives.

Background of the SB, WISC, and Goodenough

In 1960, the SB,L-M was made available as part of the repertoire of psychological tests. Perhaps the words of Terman and Merrill best describe and summarize the attributes and background of this latest revision.

> The Stanford Revision in 1960 retains the main characteristics of scales of the Binet type. It is an age scale making use of age standards of performance. It undertakes to measure intelligence regarded as general mental adaptability. The 1960 scale incorporates in a single form, designated as the L-M Form, the best subtests from the 1937 scales. The selection of subtests to be included in the 1960 scale was based on records of tests administered during the five-year period from 1950 to 1954. The main assessment group for evaluating the subtests consisted of 4498 subjects aged 2½ to 18 years. Changes in difficulty of subtests were determined by comparing the per cents passing the

> individual tests in the 1950's with the per cents passing in the 1930's constituting the original standardization group. Criteria for selection of test items were: (1) increase in per cent passing with age (or mental age); and (2) validity determined by biserial correlation of item with total score. Changes consisted in the elimination or relocation of tests which have been found to have changed significantly in difficulty since the original standardization; the elimination or substitution of tests which are no longer suitable by reason of cultural changes; further clarification of ambiguities of scoring principles and test administration; and the correction of structural inadequacies of the 1937 scale, first by introducing adjustments to make the average mental age that the scale gives more nearly equal to the average chronological age at each age level and second, by providing revised and extended IQ tables that incorporate built-in adjustments for atypical variability of IQs at certain age levels so that the standard score IQs provided are comparable at all age levels. (Terman & Merrill, 1960, pp. 39-40)

From the above statement it can be seen that the 1960 SB,L-M does not introduce any real innovations, but rather, it is an improved continuation of concepts and methods previously found useful in the field of intelligence testing.

While a detailed account of the many existing definitions and theories of intelligence is not within the scope of this study, a brief overview of the definitions and theories specifically pertaining to the SB, WISC, and Goodenough tests is in order.

In considering the long evolution of the Binet type tests the questions might be asked: What was "intelligence" to Binet, and what were his methods of measurement?

Peterson (1925, pp. 170-171) in his discussion of the early Binet-Simon scales quotes Binet as saying, "'There is in intelligence, it appears to us, a fundamental organ, any defect or alteration in which is of the greatest consequence to the practical life. This basic factor in intelligence is judgement, otherwise spoken of as good sense, practical judgement, initiative, the ability to adapt oneself. To judge well, to comprehend well, to reason well, these are the essentials of intelligence.'"

According to Terman (1916, p. 45), "Binet's conception of intelligence emphasizes three characteristics of the thought process: (1) Its tendency to take and maintain a definite direction; (2) the capacity to make adaptations for the purpose of attaining a desired end; and (3) the power of auto-criticism." In discussing Binet's methods Terman states, "...the Binet tests differ from most of the earlier attempts in that they are designed to test the higher and more complex mental processes, instead of the simpler and more elementary ones"(1916, p. 42). "Where others had attempted to measure memory, attention, sense discrimination, etc., as separate faculties or functions, Binet undertook to ascertain the <u>general level</u> of intelligence.... (He) undertook, so to speak, to triangulate the height of his tower without first getting the dimensions of the individual stones which made it up"(pp. 42-43).

In essence, the successive Stanford revisions have adhered to this original "tower" principle ever since Binet's time.

McNemar (1942), in his factor analyses of the 1937 SB Revision (Terman & Merrill, 1937), shows that the selection of items for the purpose of measuring general intelligence was mainly successful in that end. He states in his summary chapter, "It would appear that the items in the scale are not measuring such a hodgepodge of abilities as some have supposed. Presumptive evidence (fourteen factor analyses) was presented to the effect that the common factors at successive levels are nearly identical" (1942, p. 168). These nearly identical factors are taken to indicate the operation of general intelligence, whereas group factors are considered absent (p. 122). Thus the 1937 scales were shown to be mainly a measure of general intelligence which was Binet's original intent.

"In selecting tests for the L-M Form, the factor loadings of McNemar's analyses for the various age levels were taken into account. Though items with low first factor loadings were not entirely eliminated, all items which are highly saturated with the general factor were included" (Terman & Merrill, 1960, p. 35). The 1960 scales are presumed, then, to continue in the Binet tradition by attempting to measure general intelligence. Judging by the nature of the items included in this scale, general intelligence

is highly verbal in flavor, at least above the Year V level. This was also apparently true for the 1937 Revision because McNemar unsuccessfully attempted to construct a non-verbal scale from SB items. He reported that "Items chosen as depending less on the language factor were scored as a possible non-verbal scale of intelligence, but the small number of available items mitigated against respectable reliability and validity"(McNemar, 1942, p. 122). Higgins and Sivers comment on the highly verbal nature of the 1937 SB by referring to other of McNemar's findings. "McNemar reports that the first factor loading of vocabulary progressively increases (.59 to .91) from experimental age 6 through 18. This rules out an analysis of SB item discrimination on the basis of verbal and non-verbal content" (Higgins & Sivers, 1958, p. 465).

Wechsler defines intelligence operationally as "the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment"(1958, p. 7). He goes on to say that "Although intelligence is not a mere sum of intellectual abilities, the only way we can evaluate it quantitatively is by the measurement of the various aspects of these abilities"(1958, p. 7). He built the WISC as well as his other tests accordingly. The WISC contains groups of verbal and performance subtests which yield separate Verbal and Performance IQs and a combined Full Scale IQ.

Wechsler (1958, p. 119), in discussing the factorial composition of the Wechsler-Bellevue I, reports that factor analyses have produced three recurrent factors. "These are a broad verbal factor (verbal comprehension), a non-verbal organization factor (variously identified as performance, non-verbal, space and visual-motor organization) and g (sometimes referred to as the eductive or general reasoning factor)." He then relates these findings to the WISC by citing evidence from a study by Gault (1954) showing the "equivalence and continuity of the factors entering into both Scales"(WISC and WB-I).

Cohen (1959) in a recent factor analysis of the WISC has reported five factors in addition to g. There were two factors involving verbal comprehension, a perceptual organization factor, a factor involving freedom from distractibility, and a "quasi-specific" factor pertaining to the Coding and Picture Arrangement subtests.

With regard to the Full Scale, Verbal and Performance trichotomy, Cohen reports the following (p. 296):

> Both the Verbal and Full Scale IQs are excellent measures of G.... Either of these IQs represents well the G defined by the 12 subtests.... The modification of clinical practice suggested by the present results lies in the conception of the Verbal IQ as a measure of G, rather than of a 'group' or primary factor of verbal ability. Another way in which this can be put is that <u>insofar as the WISC is concerned</u> (italics his), no distinction in measurement function can be made between the Verbal IQ and the Full Scale IQ for the general population of children.

In the case of the Performance I_q, although the error variance is slightly larger, there remains about a quarter of the total variance specific to these tests, which is predominantly reflective of primary-specific Factor B, that of Perceptual Organization which does not enter into G. In clinical use, the comparison of Verbal with Performance I_q for a subtest is seen to be justified, in that the difference between the two reflects the battery specificity of the Performance Scale, the effect of G having been in effect 'partialled out' by subtraction.

The WISC and SB tests would both appear to measure a general intelligence factor. It also seems reasonable to expect that the SB, due to its highly verbal nature, would be more closely related to the WISC Verbal Scale than to the Performance Scale.

While the items of the SB and WISC were empirically selected, the total tests themselves are well-grounded in the theories of intelligence espoused by their respective originators. The Goodenough, on the other hand, seems to be more an empirically based instrument than a derivation from any particular theory of intelligence. However, from surveying the Goodenough manual a few ideas can be obtained concerning the intellectual processes involved in performance on this test.

1. Association by similarity. The child sees a resemblance between a series of lines on paper and the concrete object represented by them. This is the preliminary stage which must precede any active attempt at representation on the part of the child himself.

2. Analysis into its component parts of the object to be drawn.

3. Evaluation of these parts and selection of those which appear to be essential or characteristic. This process is largely an unconscious one as far as the child is concerned, but it is significant, since it is determined by the nature of his interests and by his fundamental habits of thought.

4. Analysis of spatial relationships; of relative position.

5. Judgments of quantitative relationships; of relative proportion.

6. Through further process of abstraction, reduction and simplification of the several parts into graphic outlines.

7. Coordination of eye and hand movements in the drawing act.

8. Adaptability; the capacity to adjust the drawing scheme to the new features which are added from time to time as the concept develops (Goodenough, 1926, pp. 78-79).

Elsewhere in the manual it is stated that "In young children a close relationship is apparent between concept development as shown in drawing, and general intelligence"(p. 12). While this may be true, it would appear from the processes listed above that the type of intelligence being "tapped" by this test is very much performance in character. This is not to deny, however, that g elements are present. Such processes as evaluation, abstraction, and adaptability are more g type functions than pure performance processes. In general though, the Goodenough seems to be a test that is highly performance in nature.

Birch (1949, p. 223) tends to disagree with this position. He says:

> Categorizing the (Goodenough) Drawing Test as a 'performance' test...may be doing violence to the concept of 'performance ability' as it is associated with the scales of Arthur, Cornell-Coxe, and Wechsler. Motor-speed is not at a premium in this test. Neither is attention-span nor manipulative skill. Any 'performance' quality in the test is there mainly because the subject does something and leaves a record of having done it.

The answer to this differing point of view is that while such things as attention span and manipulative skill are not in themselves being measured as separate entities, they necessarily contribute to the final score much as the individual performance processes involved on the WISC Performance subtests each contribute to the total Performance IQ. Therefore, it is this experimenter's position that the Goodenough is mainly a performance test even though the score depends upon the finished product rather than the separate processes which go into the task.

Bender (1952, pp. 144-145) seems to agree with this position. She says:

> It is probable that the capacity to draw the human form is not related to a simple visual gestalt but a more complicated gestalt which is based upon sensory impressions of all types coming from the surface as well as from the inside of the body. Besides the sensory impressions of the present, the sensory consciousness of the past are integrated into the present concept. However, it is a most important fact that motor impulses give the final shape to the body image.

To summarize what has been said about the theoretical implications of the SB,L-M, the WISC, and the Goodenough, the following conclusions can be drawn:

a) The SB,L-M in the tradition of Binet, is mainly a test of general intelligence. The items have been selected with McNemar's first factor loading in mind, a factor which seems to represent general intelligence. Judging by the nature of the items, particularly above the Year V level, the SB,L-M requires mainly abilities of a verbal nature.

b) As shown by factor analysis, the WISC, through its Full and Verbal Scales, seems also to measure general intelligence, but by its Performance Scale seems to measure performance ability as well.

c) While there are some general intelligence functions involved in behavior on the Goodenough, the test seems mainly to call performance abilities into play and, hence, it is mainly a performance test.

Familial and Organic Mental Defectives

For this study the definition of "familial" will be that used by the American Association on Mental Deficiency (1960, p. 2): "This category depends on multiple causative mechanisms of which the most distinctive is an inherited

sub-average intellectual status or adequacy.... The diagnosis is based on the presence of a relatively similar degree and type of sub-average intellectual status in one or both parents and in most of the siblings."

While it is somewhat at odds with the definition given above, a good summary statement concerning the etiology of the familial is given by Clarke and Clarke (1958, pp. 132-133). They used the term "subcultural defective," a term usually held as more or less synonymous with "familial." (Also known as "garden variety" mental deficiency - Sarason, 1959, p. 101; simple primary amentia - Tredgold and Soddy, 1956, p. 36; and endogenous mental deficiency - Strauss and Lehtinen, 1947).

> Studies relating to the aetiological factors in subcultural defect are so numerous, vary so much in method and validity, and have sometimes produced such apparently contradictory results, that it is very difficult to come to a satisfactory conclusion at the present time. Certain facts, however, seem to the writer to be fairly well established:
>
> (i) Subcultural mental defectives have a higher proportion of dull and deficient parents, siblings, and other relatives than do members of the normal population; the closer the blood relationship, the closer the similarity in intellectual and social status.
>
> (ii) A higher proportion of the children of subcultural defectives are dull or defective than are the children of normal people.
>
> (iii) But <u>the majority</u> of the parents, siblings, and children of subcultural defectives are not themselves defective nor even very dull. (italics theirs)

> (iv) The...large majority of inmates of mental deficiency institutions and pupils in special schools for the educationally subnormal are drawn from the lowest strata of society, have not infrequently been subjected to adverse experiences, and as a rule have suffered gross cultural deprivation.

In Benda's (1952) system of classification there is one group which seems to include the familial defective. "This group represents biologically normal persons who have low I.Q. ratings but are...a legitimate part of the population. Their low I.Q. ratings are genetically determined but, paradoxically, their genes are not pathologic from a biologic point of view"(Benda, 1952, p. 15). Hutt and Gibby in discussing the children of this group say that "These children have no demonstrable biologic involvement, but have a low degree of intelligence. They are a normal part of our population"(Hutt and Gibby, 1958, p. 97). This means that if intelligence is normally distributed throughout the population at large, these children are merely the lowest end of the distribution.

The IQ range of the familial defective is usually thought of as being from 50 to 70 - what was formerly called the Moron range of intelligence (Sarason, 1959, p. 101). However, the American Association on Mental Deficiency (1960, p. 2) says that "the mental level is usually in the borderline or mild category," i.e. from 55 to 85.

The definition of "organic" for this study will be that used by Beck and Lam (1955, p. 154): "The term 'organic' is defined as: any invlovement of the central nervous system resulting from structural changes caused by disease, injury, or malformation occurring before, during, or after birth." Organic disorders with gross neurological or physiological malfunctioning will not be included.

Excerpts from a discussion by Barnett, Ellis and Pryer (1960) will expand the above definition for clarification.

> A survey of the literature would suggest that two, possibly three, 'classes' of organics have been employed in research. The first class (excluded from the present study) consists of brain-injured Ss who are so grossly affected that even the lay person would probably have little difficulty in 'seeing' the S's disability. In addition, neurological reports and the developmental history and available to support the diagnosis (p. 894).
>
> A second class of organics utilized in research are those of the 'Straussian' (Exogenous) variety. These Ss are lacking in gross physical stigmata and are diagnosed on the basis of minor neurological signs and symptoms. The developmental history usually corroborates such a diagnosis (p. 895).
>
> Possibly a third class of organics is that involving Ss diagnosable only on the basis of developmental history. These Ss appear to be lacking even minor symptoms (p. 895).

The second and third classes of organics will be considered in the present study.

In describing the exogeneity syndrome, class two above, Strauss and Lehtinen (1947, p. 15) say that "First, it was found that these children came from families with

parents and siblings of normal mentality, second, that they showed a history of prenatal, natal, or postnatal damage to the brain, and third, that medical examination revealed slight neurological signs although no conspicuous motor impairment."

The familial and organic etiologies have been defined and briefly described. The next topic of importance is the functioning of mental defectives in previous studies utilizing the 1937 SB, the WISC, and the Goodenough.

Previous Research

The literature dealing with the functioning of mental defectives on the 1937 SB, WISC, and Goodenough is rather sparse. In the studies existing, if any note was taken of etiologies, it was to the exclusion of organics as defined in this study. In view of this fact, the composite trends derived from the studies reviewed below are probably best considered as applicable to mental defectives in general, or at least to those at the upper end of the defective range.

SB vs. WISC

A review of the literature produced the following studies which, while not exactly comparable to the present study, have utilized mental defectives and their functioning

on the WISC and 1937 SB. These studies are summarized in Table 1, p. 18. Since the 1960 SB,L-M uses the same items from either the 1937 Form L or Form M, the results of these previous studies can with appropriate caution be generalized to what might be expected from the mental defectives of the present study.

Sloan and Schneider (1951) report a study which dealt in part with a comparison of the WISC and SB,L. The subjects were forty institutionalized high grade mental defectives divided equally as to sex and classified either as familial or undifferentiated etiology. The mean CA for the group was 13.5 years with a range from 9.1 to 15.5. Unfortunately no attempt was made to equate either the sexes or the etiological classifications by IQ beforehand; a shortcoming which the authors acknowledge. Because of this the results are merely applicable to the forty subjects as a group. Any further breakdown seems unjustified for IQ comparison purposes. The results for the group showed that the SB,L was significantly correlated with the WISC Full, Verbal, and Performance Scales. The highest correlation, .751, was obtained with the Verbal Scale, and the lowest, .493, was obtained with the Full Scale. The correlation obtained with the Performance Scale was .641. Regarding means and mean differences, all the means of the WISC scales were higher than the SB,L, the means of the Verbal and Performance Scales being significantly higher.

TABLE 1 Comparison of SB & WISC IQs in Five Previous Studies Utilizing Mentally Defective Children

Study, N, & type of subject		SB	FSIQ	WISC VIQ	PIQ
Sloan and Schneider (1951) N=40 High grade familials & undifferentiated CA 9.1 to 15.5	M	56.3	58.3	59.7	64.6
	S.D.	4.8	9.5	6.2	12.7
	r (SB vs WISC)		.493	.751	.545
	t (mean diff.)*		1.53	5.239**	7.28**
Stacey and Levin (1951) N=44 Morons CA 7-2 to 15-4	M	62.5	61.2	63.6	65.4
	S.D.	6.78	5.48	6.16	8.37
	r	.....	.60	.56	not given
	t*	.....	1.54	1.20	
Nale (1951) N=104 Mental defectives CA 8-10 to 15-11	M	55.38	57.97		
	S.D.	9.85	10.15		
	r	.....	.909	not given	
	t*	.....	6.6 **		
Sandercock and Butler (1952) N=90 Defectives in school programs CA 10 to 16	M	58.5	59.0	62.8	62.6
	S.D.	9.0	11.4	9.7	12.4
	r	.....	.76	.80	.66
	t*	.....	.639	6.858**	4.158**
Vanderhorst, et al. (1953) N=38 Familials & undifferentiated CA 11 to 16	M	59.34	62.18	61.74	70.05
	S.D.	5.07	7.15	7.15	9.92
	r		not given		
	t	.....	1.73	1.52	5.31***

* t calculated by this experimenter from data reported.

** .001 level of significance

*** .05 " " "

Stacey and Levin (1951) report a study in which they compared a moron group and a borderline group with respect to their functioning on the WISC and SB,L. Since the interest of the present study is with the moron group, discussion will pertain to that group. The mean CA of the 44 institutionalized subjects in this group was 11 yrs. 4 mos. and the range was from 7-2 to 15-4. No information concerning sex or etiology is given. Results indicate significant correlations between the SB,L and the WISC Full and Verbal Scales. For some reason the authors neglected to give the SB,L vs. WISC PIQ correlation. Regarding means and mean differences, results indicated that the mean SB,L IQ was lower than either the mean WISC VIQ or PIQ, but not significantly so in the case of the VIQ. The degree of significance in the case of the PIQ could not be calculated since the r was not given. The WISC FSIQ was lower than the mean SB,L IQ but not significantly so.

In a study by Nale (1951) on 104 institutionalized mentally defective children ages 8-10 to 15-11 the following results were found: a) a significant correlation of .909 between the 1937 SB (no form given) and WISC FSIQ, and b) a significant difference between the SB and WISC FSIQ means. There was no information given concerning etiology, and there were no comparisons of the SB with the WISC Verbal and Performance Scales. The results of this study seem merely to be that the mentally defective

children functioned significantly higher on the WISC Full Scale than on the SB and that the two tests were significantly correlated.

Sandercock and Butler (1952) used 90 institutionalized mentally defective children who were enrolled in school programs. There were 58 males and 32 females, and they ranged in CA from ten to 16 years with a mean CA of 13.0 years. This study used the SB,M and the IQs obtained ranged from 45 to 86 with a mean of 58.5, SD 9.0. This mean IQ was lower than the mean IQs obtained on all of the WISC scales; significantly lower than the mean Verbal and Performance IQs. The correlations between the SB,M and WISC scales were .76 with the Full Scale, .80 with the Verbal, and .66 with the Performance, all significant.

As part of a study by Vanderhorst, Sloan and Bensberg, SB IQs (form not given) were compared with WISC IQs for mean differences. The subjects were 38 mentally defective institutionalized children, 22 males and 16 females between the ages of 11 and 16 years with a mean age of 15-1. They were diagnosed as either familial or undifferentiated. The mean SB IQ was lower than all of the mean WISC IQs, significantly lower than the mean Performance IQ. No correlations were reported between the SB and WISC scales.

To summarize the results of the five previous studies, the following trends are to be noted where the information is available:

a) In all the studies where results were reported, the SB was significantly correlated with the WISC FSIQ, VIQ, and PIQ. SB vs. WISC correlations ranged from .493 to .909 for the FSIQ, .56 to .80 for the VIQ, and .545 to .66 for the PIQ.

b) In four of the five studies reporting, the mean WISC FSIQs were higher than the mean SB IQs; in one case significantly higher.

c) In the four studies reporting, all mean WISC VIQs were higher than the mean SB IQs; in two cases significantly higher.

d) In the four studies reporting, all mean WISC PIQs were higher than the SB IQs; in three cases significantly higher.

As a matter of interest, the results of the following SB vs. WISC studies utilizing normal children are presented in Table 2, pp. 22-23: Frandsen and Higginson (1951); Weider, Noller and Schramm (1951); Pastovic and Guthrie (1951); Krugman, Justman, Wrightstone and Krugman (1951); Cohen and Collier (1952); Holland (1953); Arnold and Wagner (1955); Gehman and Matyas (1956); Harlow, Price, Tatham and Davidson (1957).

Table 2 Comparison of SB & WISC IQs in previous studies utilizing normal children

Study, N & subject		SB	WISC FSIQ	WISC VIQ	WISC PIQ
Frandsen & Higginson (1951) N=54 4th grade	M	105.8	102.4	100.9	103.5
	S.D.	11.15	11.15	12.25	11.20
	r (SB vs WISC)		.80	.71	.63
Wieder, et. al. (1951) N=106 School children, age 5 to 12	M	93.1	90.0	91.9	89.8
	S.D.	19.56	18.90	17.07	18.14
	r	.....	.89	.89	.77
Pastovic & Guthrie (1951) N=50 Kindergarten	M	113.02	101.58	104.24	103.16
	r	.....	.63	.57	.71
Pastovic & Guthrie (1951) N=50 2nd grade	M	115.08	108.56	112.68	111.50
	r	.....	.82	.71	.88
Krugman, et. al. (1951) N=332 School children, age 5 to 15	M	108.45	101.23	103.37	98.28
	S.D.	15.85	12.75	13.17	15.05
	r	.....	.817	.739	.644
Cohen & Collier (1952) N=51 Age 6 to 8	M	104.8	99.8	98.5	101.1
	S.D.	15.07	14.23	14.48	14.51
	r	.....	.85	.82	.80

(Continued on next page)

Table 2 Continued

Study, N & type of subject		SB	WISC FSIQ	WISC VIQ	WISC PIQ
Holland (1953) N=52 Ages 6 to 8	r (SB vs WISC)		.87	.88	.73
Arnold & Wagner (1955) N=50 3rd & 4th grades	M	104.52	101.88	104.70	103.34
	S.D.	15.66	12.75	15.40	13.59
	r	.....	.85	.75	.88
Gehman & Matyas (1956) N=60 5th grade	M	96.17	98.13	96.90	99.87
	S.D.	14.29	11.20	10.84	13.19
	r	.....	.73	.78	.46
Gehman & Matyas (1956) N=60 9th grade	M	98.25	99.82	96.72	103.23
	S.D.	14.38	12.79	11.14	14.44
	r	.....	.77	.76	.64
Harlow, et. al. (1957) N=90	Age	r			
	6½	SB	.64	.64	.51
	10	vs	.83	.88	.52
	14	WISC	.83	.79	.71

From these studies the following trends are to be noted:

a) In all of the studies the SB was significantly correlated with each WISC scale. SB vs. WISC correlations ranged from .63 to .89 for the FSIQ, .57 to .89 for the VIQ, and .46 to .88 for the PIQ.

b) In seven of the nine studies where the information was reported, the mean SB IQs were higher than the mean WISC FSIQs, VIQs, and PIQs.

From this information a difference between the functioning of normals and defectives on the SB and the WISC can be seen. Defectives tend to score lower on the 1937 SB than on any of the WISC scales whereas normals tend to score higher. This conclusion seems to give partial support to Nale's suggestion (1951, p. 421) "that Binet results are higher at the upper limits of the normal ability range, that WISC and Binet results are nearly identical at the lower limits of the normal ability range and that WISC results are slightly higher at the upper level of the defective range." Littell (1960), in his comprehensive review of the WISC literature, reaches essentially the same conclusion.

An explanation of the phenomenon of defectives' scoring higher on the WISC than on the SB can be obtained from Anderson's (1953) criticism of the WISC. Here he

emphasizes the fact that for the lower age levels on the WISC credit is given on subtests even if all the items are failed. He states that "A zero scaled score cannot be earned until the age of 6-8 through 6-11 and here on only one test. Not until we reach the norms for 12-0 through 12-3 do we find that credit is no longer given for zero performance. Nothing could more clearly indicate the inapplicability of this material for the younger age groups" (Anderson, 1953, pp. 363-364). From this standpoint the WISC would also be inappropriate for defectives who score at age levels below 12-0 since they too would be given gratuitous scaled scores for zero performance. The differences in mean SB and WISC IQs found for defectives seem to stem from the simple fact that at the lower age levels poor performance is rewarded on the WISC whereas it is penalized on the SB with a resultant lower SB IQ. If this is true, the treating of SB and WISC IQs as equivalents is not a valid procedure for mental defectives. Part of the task of the present study is to determine whether the same SB vs. WISC mean IQ difference trends obtained by mental defectives with the 1937 SB can be found with the 1960 L-M Revision.

SB vs. Goodenough

A review of the literature produced the following previous studies which show the comparative functioning of mental defectives on the 1937 SB and the Goodenough. These results are summarized in Table 3, p. 26.

TABLE 3 Comparison of SB and Goodenough IQs in Previous Studies Utilizing Mental Defectives

Study, N & type of subject		SB	Goodenough	
Birch (1949)				
N=68, Low IQ	M	53.75	57.647	
students, public	S.D.	11.73	15.22	
school	r	.....	.62	
CA 10-6 to 16-3				
Johnson, et.al. (1950) N=209				
Institutional	r		.4808	
school program				
CA 6-9 to 17				
Murphy (1956)	M	62	66	
N=30 Institution-	S.D.	15.3	18.5	
alized male adults	r	....	.78	
CA 20 to 54	t*	....	1.89	.05

*t calculated by experimenter from data reported.

Birch (1949) reports a study comparing Goodenough IQs with SB IQs obtained from both the Form L and Form M. The subjects were 68 children examined in a public school system. There were 43 boys and 25 girls all with SB IQs below 70. Their ages were 10-6 to 16-3 with a mean age of 150.7 months (12-5). A Pearson r of .62 was found between the two sets of IQ scores. Unfortunately any test of significance of the mean differences is of doubtful meaning since the range of the SB IQs was restricted to 70 or below when the sample was selected while the Goodenough IQ range was unrestricted. This study shows, however, that SB L or M IQs are significantly correlated with Goodenough IQs.

As part of a study using the Goodenough as a projective device, Johnson, Arthur and Lahey (1950) give some information on the comparison of SB and Goodenough IQs. Two-hundred-nine institutionalized children in school programs were used. They ranged in CA from 6-9 to 17, average CA of 12-11. MAs ranged from 2-6 to 13-9 with an average MA of 6-7. They were of several etiologies including feeble-minded, epilepsy, post-encephalitis, and brain damage. A product-moment correlation of .4808 was reported between the two tests, and means which "each approximated 50" were obtained. Thus, the Goodenough seems to be significantly correlated with the 1937 SB when used with mentally defective children.

Murphy (1956) as part of a larger study reports the following comparison between the SB,L and Goodenough. A significant r was found as well as a significant mean difference in favor of the Goodenough. The subjects were 30 institutionalized non-organic mentally defective male adults ranging in CA from 20 to 54 years with a mean CA of 30. Murphy, on the basis of her information, makes the statement that "The relatively high correlation (.78) between drawing scores and the Stanford-Binet results for the mentally deficient group suggests that the Draw-A-Man Test may serve as an adequate measure of the general intellectual ability of adjusted mentally deficient adults" (p. 398).

In interpreting Murphy's results, one must keep in mind, however, that although she does not specify, she probably used a CA of 13 in calculating the Goodenough IQs as the Goodenough manual suggests. In a recent study, Mitchell (1959) has demonstrated an MA increase above the age of 13 as measured by the Goodenough. Because of this she states that her results "indicate that a maximum CA divisor of 13.0 is inappropriate in calculating Draw-A-Man IQs; the significant MA increment after the age 13.0 should be counterbalanced by a corresponding CA increment." She goes on to say that "an appropriate pattern for CA increase through the years from 13 to 16 was found to be that used by Terman and Merrill in constructing the Stanford-Binet (1937) IQ tables"(p. 557). Thus if Murphy did use a CA of

13.0 as her divisor, the significant mean difference may be a function of that rather than any true mean difference.

Bender, in discussing body image problems of the brain damaged child, had this to say about his performance on the Goodenough. "The inability to integrate the perception of himself into an adequate body image is seen in the Goodenough drawing of a man.... This poor organization of the body image is sometimes reflected in a drawing which may be two years or more below the mental age of the child" (Bender, 1956, p. 99). Elsewhere Bender presented essentially the same idea and documented her discussion with the results of five case studies of chronic encephalitis in children. In four of the five cases, the Goodenough MA was approximately one-half or less of the 1937 SB MA (Bender, 1952).

Of additional interest are relationships between the Goodenough and the 1916 SB for retarded subjects. Yepsen (1929) reports an r of .60 between SB MAs and Goodenough MAs for 37 institutionalized feebleminded boys, ages 9-0 to 18-11. McElwee (1932) reports an r of .717 between SB and Goodenough MAs for 45 fourteen year old "subnormal" school children in ungraded classes. She also reports median MAs of 8-0 and 7-3 for the SB and Goodenough respectively. Isrealite (1936) reports an r of .71 between SB

and Goodenough MAs. She also reports average Goodenough MAs as being "slightly lower" than the SB MAs. Actual values are not given.

For normal subjects the Goodenough and 1937 SB IQs have been demonstrated to be significantly correlated. McHugh (1945) reports an r of .41 using 90 kindergarten children. Havighurst, Gunther and Pratt (1946) report an r of .50 using 66 midwestern white children. Harris found rs of .65 and .75 using groups of 23 and 35 kindergarten children. Harris' rs are reported by Ansbacher (1952). No means or mean differences were reported in these studies.

In summary, from the preceding discussion, the following trends can be noted in the relationship between the SB and the Goodenough for retardates:

a) Significant correlations were obtained between the Goodenough and both the 1916 and 1937 SBs. These ranged from .60 to .717 for the 1916 SB and from .48 to .78 for the 1937 SB.

b) A tendency was shown for MAs to be lower on the Goodenough than on the 1916 SB.

c) Non-organic defectives may receive higher IQs on the Goodenough than on the 1937 SB, but results are inconclusive due to methodological ambiguity.

d) Organic defectives can be expected to perform more poorly on the Goodenough than on the SB due to body image problems.

These relationships might be expected to hold for the 1960 SB,L-M as well.

Goodenough vs. WISC

Only one previous study comparing the Goodenough and the WISC could be found which used mental defectives as subjects. This is surprising when one considers the fact that mental defectives usually find drawing a fairly facile medium of expression.

The study in mind is that by Warren and Collier (1960). Part of their interest was a comparison between Goodenough and Wechsler IQs using 49 female patients who ranged in CA from 9 to 30 years. No etiologies were given. The WISC was used for the younger subjects and the WAIS for the older ones. (Ns were not specified.) The results are reported in such a manner that specific separate Goodenough vs. WISC and Goodenough vs. WAIS comparisons are not possible. A correlation of .43 was found between the Goodenough IQs and the IQs of the WISC and WAIS pooled together. With a maximum CA of 13 being used, the mean Goodenough IQ for the entire group was 65.30, SD 15.29. The mean WISC FSIQ was 60.60, SD 12.44, and the mean WAIS FSIQ was 63.14, SD 12.07. About all that can be said from these results for purposes of the present study is that the Goodenough and WISC Full Scale seem to be significantly correlated, and female defectives seem to score somewhat higher on the Goodenough than on the WISC.

Differential SB Functioning of Organics and Familials

With regard to the more subtle quantitative differences in the test performance of organic and familial retardates the following information is to be considered.

In describing the functioning of brain-injured children, Bender (1956, p. 87) states, "The 'scatter' in the Stanford-Binet is well known. Failures usually occur in the so-called performance portions or the maturational tests, or the perceptual motor patterns."

Berko (1955, p. 20) has put the SB scatter to the test, so to speak, by comparing 46 brain-injured aphasic retardates, mean CA 9-3, IQ 48, with a control group of 46 endogenous retardates, mean CA 9-2, IQ 46. Scatter scores were "obtained by counting the number of items missed between the last consecutive item passed and the upper limit of each child's total performance range." The mean scatter score for the organic group was 12.54 and for the non-organic group was 6.78, the difference being statistically significant. This leads Berko to conclude that "... the brain-injured mental retardate can be psychometrically differentiated from the endogenous retardate on the basis of intra-range scatter on the Stanford-Binet IQ test." The question for the present study is whether the scatter on the 1960 SB,L-M will also differentiate organic from non-organic retardates.

Lord and Wood (1942), Strauss and Lehtinen (1947), Bensberg (1950), Lewis (1951), Bender (1956), Kaliski (1959), and Eisenson (1960), to name a few, refer to the perceptual difficulties commonly associated with organicity. Gallagher (1957), on the other hand, failed to find any group differences in perceptual functioning between familial and brain-injured retarded children "although a minority of the group displayed definite perceptual problems" (p. 64). Lewis sums up the situation regarding aberrations in the functioning of brain-injured children by saying, "It is probable that when damage does occur in the brain, it alters all or many basic functions in varying degrees. In our research, however, we have found four basic deviations in the mental make-up of the brain-injured child.... These are disturbances in perception, in concept formation, in language, and in emotional behavior" (1951, p. 15). These conclusions are based mainly on the earlier research of Strauss, e.g. Strauss (1943) and Strauss and Werner (1942), which involved such things as sorting tasks and impressions from behavioral observations.

One might expect such disturbances as described above to be of diagnostic value when reflected in the performance of mental defectives on individual SB items.

Hoakley and Frazeur (1945) report a study in which the item VII,3 Form L, Copying a Diamond was taken

to be the only item which differentiated between a group of exogenous and a group of endogenous male defectives. Eighteen pairs of children, one member of each pair being from each group, were matched on CA, MA, and IQ. CAs ranged from 9-1 to 15-1, MAs from 6-2 to 10-10, and IQs from 50 to 75. The level of significance was set at .01. However, if it were lowered to .05, two additional items became significant: VII,4 (Form M), Repeating 3 Digits Reversed, and IX,3 (Form L), Memory for Designs (1st design). No significance difference between the groups was found in the range of items covered, i.e. in the scatter.

Gallagher (1957) also attempted to ascertain which SB items, if any, differentiated between a brain-injured group and a familial group of defectives. Each group consisted of 24 subjects; CAs ranged from 6-9 to 14-5 and IQs from 35 to 76. Gallagher states, "It can be seen...through the five levels and 30 items reported, only two items differed significantly between the two groups. These are item 2 at year level IV-6, Repeating Four Digits, and item 6 at year VI, Maze Tracing. Both of these results are reasonable in terms of the initial general hypotheses of distractibility and perceptual difficulties of the brain-injured child" (p. 50).

In the present study the differentiating power of the items of the SB,L-M will be investigated for familials and organics.

HYPOTHESES TESTED

From the results of previous studies, the following relationships are expected to exist between the SB,L-M and the WISC for the Familial and Organic Groups considered separately:

I The SB,L-M is significantly correlated with the WISC Verbal, Performance, and Full Scales.

II The mean SB,L-M IQ is significantly lower than the mean IQ of the WISC Verbal, Performance, and Full Scales except in the case of the organics where the mean SB,L-M vs. PIQ difference is not expected.

From the results of previous studies, the following relationships are expected to exist between the SB,L-M and the Goodenough for the Familial and Organic Groups considered separately:

III The SB,L-M is significantly correlated with the Goodenough.

IV For the Organic Group, the mean Goodenough IQ is significantly below the mean SB,L-M IQ due to posited perceptual and body image difficulties of this group.

The following relationships are expected to exist between the Goodenough and WISC for the Familial and Organic Groups considered separately:

V The Goodenough is significantly correlated with the WISC Verbal, Performance, and Full Scales.

VI For the Organic Group, the mean Goodenough IQ is significantly below the mean WISC Verbal and Full Scale IQs due to posited perceptual and body image difficulties in this group.

From the results of previous studies, the following factors of differential test functioning between the Familial and Organic Groups are expected:

VII The Organic Group exhibits a larger SB,L-M "scatter" than the Familial Group, scatter being measured by the number of items failed between the basal level and the level where all items are failed.

VIII The SB,L-M items which differentiate between the Organic and Familial Groups are VII,3 Copying a Diamond; IX,3 Memory for Designs I (1st design); and VI,6 Maze Tracing. (Repeating 3 Digits Reversed and Repeating 4 Digits which previously differentiated between the groups are not included in the regular SB,L-M administration.)

IX Both the mean PIQ and mean Goodenough IQ are significantly lower for the Organic Group than for the Familial Group due to posited perceptual and body image difficulties of the Organic Group.

METHOD

Samples and Population

The samples used in this study were one group of 26 organic mentally defective children whose diagnoses came under the definition of "organic" previously presented, and one group of 20 mental defectives who had been previously diagnosed as familial. All subjects were in residence at the Lapeer State Home and Training School, Lapeer, Michigan, and their diagnoses had been made by a staff representing the departments of medicine, nursing, psychology, and social service of that institution.

Besides etiological classification, the subjects were selected according to the following criteria:

a) Their previous IQs fell in the range of 50 to 70. These IQs were mainly SB,L IQs. Since, however, SB,L IQs were not available for all subjects, previous WISC IQs were used for approximately 20% of the subjects in each group - five familials and six organics.

b) They were currently enrolled in the institutional school programs.

c) They were not crippled, blind, deaf, untidy, or presently subject to seizures.

d) They had not been given any sort of psychological test for at least one year prior to this experiment. The mean

number of years since previous testing was 2.4 for the Familial Group with a range of 1.2 to 4.3 years, and 2.5 for the Organic Group with a range of 1.1 to 4.8 years.

The two groups were equated with regard to previous IQ scores, sex, CA, and length of institutionalization. Table 4, p. 39, summarizes the characteristics of the samples. T-tests showed no significant mean differences between the two groups for these characteristics, and F_{max} tests indicated homogeniety of variance.

An analysis of the 26 organic cases showed that 20 are cases of congenital cerebral maldevelopment with or without accompanying cranial anomalies, three cases of post-natal cerebral infections, and there is one case each of cerebral trauma, pre-natal infection, and previous convulsive disorder.

Generalization of results to any particular populations would have to take the above described sample characteristics into account for each of the separate Familial and Organic Groups.

Procedures

Testing was done by this experimenter during the summer months of 1960. Each subject was administered the

Table 4 Sample Characteristics

Characteristics	Familials N = 20	Organics N = 26
Sex:		
Males	n = 10	n = 13
Females	n = 10	n = 13
Chronological age (years):		
Mean	12.57	12.43
S.D.	2.18	1.88
Range	9.1 - 15.6	9.2 - 15.6
Previous IQ:		
Mean	61.05	61.15
S.D.	5.61	5.44
Range	50 - 69	52 - 70
Length of institutionalization (years):		
Mean	3.55	3.07
S.D.	1.52	1.39
Range	1.6 - 7.0	1.3 - 6.4

three tests of this study on two consecutive days. In all cases the Goodenough was attempted first to "break the ice." For the two initial refusals incurred, the administration of the Goodenough was delayed until the subjects in question signified that they were ready to "try" the test. Since the administration was individual, the instructions for the Goodenough were modified accordingly. They were: On this paper I want you to make a picture of a man. Make the very best picture you can. Take your time and work carefully. Try very hard and see what a good picture you can make.

All drawings were done on standard 8½" x 11" blank white paper with an HB medium soft pencil. IQs for subjects of CA above 13 were calculated according to Mitchell's suggestion by using the tables from the 1937 SB (See p. 28 above for discussion).

The SB,L-M and the WISC were administered according to the standard procedures detailed in their respective manuals. Ten of the twelve WISC subtests were given, those excluded being the optional Digit Span and Mazes subtests. In order to eliminate possible practice effects, the SB,L-M and the WISC were administered in counter-balanced order to consecutive subjects; i.e. Subject 1 was given the SB,L-M on the first day and the WISC on the second day, whereas Subject 2 was given the WISC on the first day and the SB,L-M on the second day, and so on for the rest of the subjects.

On the WISC, nine subjects obtained scaled scores less than the minimum given in the WISC manual. In order to determine their FSIQs, Ogdon's (1960) downward extension of the WISC was consulted. In order to determine their Verbal and Performance IQs, a downward extension for these scores was developed. This was done simply by extrapolation. In Wechler's Scaled Score - IQ conversion table (1949, p. 25), 25 VIQ units equals 20 scaled score units for the Verbal Scale, or one scaled score unit equals 1.25 VIQ units. For the Performance Scale, 25 PIQ units equals 18 scaled

score units, or one scaled score unit equals 1.39 PI_q units. Hence, to achieve the downward extension, for every scaled score unit subtracted, 1.25 VI_q units and 1.39 PI_q units were subtracted from the spread of VI_qs and PI_qs respectively. (See Table I, Appendix for Verbal and Performance downward extension values).

The subjects, whenever possible, were tested within their own cottages. In this way, the testing became more a game for the children than a stress situation. Since most of the subjects were drawn from only two cottages, the attention received in the testing situation became a "badge of distinction" within the cottage, and the subjects actually became eager to be tested. In effect, rapport was many times established before the examiner officially met the subject.

Analysis of Data

The statistical procedures used in this study were fairly simple and straight-forward. The correlational hypotheses, Hypotheses I, III, and V, were tested by use of the Pearson *r*. The mean difference hypotheses, Hypotheses II, IV, VI, and IX, were tested by use of Fisher's *t*-test. The "scatter" hypothesis, Hypothesis VII, was also tested by the *t*-test. Scatter scores were computed for the Familial

and Organic Groups by obtaining the mean munber of items failed between the basal level and the level where all items were failed.

For Hypothesis VIII, the item differentiation hypothesis, the chi-square statistic was used. Frequencies of passes and failures were tallied for each SB,L-M item attempted by at least half of the subjects from each group. The arbitrary number of at least ten subjects from the Familial Group and at least thirteen subjects from the Organic Group was chosen in order to carry greater statistical weight than any lesser frequency would carry. Chi-squares were computed from 2x2 tables with Pass and Fail categories, and Familial and Organic categories for the respective rows and columns of the tables. Where the observed frequency for any given cell turned out to be less than two, Fisher's Exact Probability Test was used due to the inappropriateness of chi-square in such cases.

RESULTS

The specific hypotheses tested in this study were those where significance was expected on the basis of previous research. However, since the expressed purpose of this study was a full comparison of the SB,L-M, WISC, and Goodenough for the Familial and Organic Groups, results where significance was not expected are also reported tangentially to the results of the specific hypotheses.

For reading ease and continuity, each hypothesis is repeated and results both specific and tangential are given immediately thereafter. For additional information, e.g. individual IQs, MAs, etc., see the tables in the Appendix.

Hypothesis I: (For the Familial and Organic Groups considered separately), The SB,L-M is significantly correlated with the WISC Verbal, Performance, and Full Scales.

Hypothesis I was accepted for the Organic Group (See Table 5, p. 44). All correlations between the SB,L-M and the WISC were found to be significant at the .01 level of significance for this group. The highest were the SB,L-M vs. WISC VIQ and FSIQ correlations, while the lowest was the SB,L-M vs. PIQ correlation.

For the Familial Group, Hypothesis I had to be rejected in the case of the SB,L-M vs. PIQ correlation which was non-significant. The other two rs were significant but only at the .05 level.

Table 5 SB,L-M, WISC, and Goodenough Correlations for Familial and Organic Groups (Pearson r)

	Familials		Organics	
WISC	SB,L-M	Goodenough	SB,L-M	Goodenough
VIQ	.467*	.432	.810**	.268
PIQ	.384	.670**	.642**	.458*
FSIQ	.483*	.625**	.804**	.389*
SB,L-M		.311		.344

* .05 level of significance
** .01 " " "

Hypothesis II: The mean SB,L-M IQ is significantly lower than the mean IQ of the WISC Verbal, Performance, and Full Scales except in the case of the Organic Group where the SB,L-M vs. PIQ mean difference is not expected.

Hypothesis II was entirely rejected (See Tables 6 and 7, pp. 45 and 46). The only mean difference in the predicted direction was that between the SB,L-M and the PIQ for the Familial Group. This difference was not significant, however.

The most surprising result of the study was the SB,L-M vs. WISC FSIQ mean difference. For both the Organic and Familial Groups, the mean FSIQ turned out to be significantly lower than the mean SB,L-M IQ, i.e. in the direction opposite to what was originally predicted. In the case of the familials, a 2.95 mean difference yielded a *t* of 1.82 which was significant at the .05 level of significance. In the case of the organics, a 5.08 difference yielded a *t* of 3.82 which was significant at the .001 level.

Table 6 Means, Standard Deviations, and Standard Errors of SB,L-M, WISC, and Goodenough IQs

	Familial			Organic		
	Mean	SD	SE	Mean	SD	SE
SB,L-M	54.80	5.53	1.24	58.54	6.62	1.30
WISC						
VIQ	54.25	6.37	1.42	58.11	10.48	2.06
PIQ	58.15	11.06	2.47	57.08	11.24	2.20
FSIQ	51.85	8.19	1.83	53.46	10.83	2.12
Goodenough	57.65	12.52	2.80	55.54	11.13	2.18

Because it was unexpected, the SB,L-M vs. FSIQ difference trend was investigated to determine whether it was established by a few highly deviant individuals or by a significant number of individuals scoring in the same direction. The statistic applied was the Sign Test as described by Siegel (1956). Results showed that in each

Table 7 SB,L-M, WISC, and Goodenough Means and Mean Differences for Familials and Organics

Familials			SB,L-M	Mean Diff.	t	p
WISC	FSIQ	51.85	54.80	2.95	1.82	.05
	VIQ	54.25	"	0.55	0.40	ns
	PIQ	58.15	"	3.35	1.46	ns
Goodenough		57.65	"	2.85	1.06	ns*
Familials			**Goodenough**			
WISC	FSIQ	51.85	57.65	5.80	2.66	.02*
	VIQ	54.25	"	3.40	1.35	ns*
	PIQ	58.15	"	0.50	0.23	ns*
Organics			**SB,L-M**	**Mean Diff.**	**t**	**p**
WISC	FSIQ	53.46	58.54	5.08	3.82	.001
	VIQ	58.11	"	0.43	0.34	ns
	PIQ	57.08	"	1.46	0.86	ns
Goodenough		55.54	"	3.00	1.38	ns
Organics			**Goodenough**			
WISC	FSIQ	53.46	55.54	2.08	0.87	ns
	VIQ	58.11	"	2.57	1.00	ns
	PIQ	57.08	"	1.54	0.61	ns

* Two-tailed since there was no rationale for expecting results to be in any particular direction + or -. The remainder are one-tailed since results were predicted in specific directions in the hypotheses tested.

group, Organic and Familial, a significantly greater number of individuals obtained lower FSIQs than SB,L-M IQs. For the Organic Group, twenty subjects scored lower FSIQs while only six scored higher ($p = .005$). For the Familial Group, fourteen subjects scored lower FSIQs while only five scored higher ($p = .032$). There was one tie. Sign Tests applied to the other SB,L-M vs. WISC IQ differences were non-significant (See Table J, Appendix).

The next point of investigation was to determine whether the significantly lower WISC Full Scale scoring was due to the fact that Ogdon's (1960) downward extrapolation was used for four of the familial subjects and for five of the organic subjects. A recalculation using non-extrapoted FSIQs, i.e. the lowest FSIQs given in the manual conversion tables, produced a mean of 55.04 for the Organics which was still significantly below their SB,L-M mean of 58.54 ($t = 3.25$, $p = .005$). The Sign Test for lower scoring individuals was also still significant ($p = .015$). Although the Sign Test for the Familial Group remained unchanged ($p = .032$), the mean difference became non-significant. The new mean FSIQ of 52.60 was 2.20 points below the mean SB,L-M IQ of 54.80. The resulting t was 1.45.

Recalculation of mean VIQs and PIQs using unextrapolated scores produced negligible differences. The largest difference was only .73 and the rest were considerably less.

Hypothesis III: (For the Familial and Organic Groups considered separately), The SB,L-M is significantly correlated with the Goodenough.

Hypothesis III was entirely rejected. The SB,L-M vs. Goodenough correlations for both groups, though positive, were not significantly so (See Table 5, p. 44).

Hypothesis IV: For the Organic Group, the mean Goodenough IQ is significantly below the mean SB,L-M IQ due to posited perceptual and body image difficulties of this group.

Hypothesis IV was rejected. Although for the organics, the mean Goodenough IQ was three points below the mean SB,L-M IQ, the difference was not found to be significant. Of additional interest, the difference for the familials, though 2.85 IQ points, in the opposite direction, was also non-significant (See Table 7, p. 46). Sign Tests for individual scoring differences were also non-significant for both groups.

Hypothesis V: (For the Familial and Organic Groups considered separately), The Goodenough is significantly correlated with the WISC Verbal, Performance, and Full Scales.

Hypothesis V was accepted for both groups regarding the Goodenough vs. WISC FSIQ and PIQ correlations. The Goodenough vs. VIQ correlations for each group were not significant, however, and the hypothesis had to be rejected in this case. The Goodenough vs. FSIQ and PIQ correlations were highest for the Familial Group, being at the .01 level of significance (See Table 5, p. 44).

Hypothesis VI: For the Organic Group, the mean Goodenough IQ is significantly below the mean WISC Verbal and Full Scale IQs due to posited perceptual and body image difficulties of this group.

Hypothesis VI was rejected. Though the mean Goodenough IQ was below the mean WISC VIQ, the difference was not significant. The mean Goodenough IQ was actually somewhat higher than the mean FSIQ but not significantly so. Of additional interest, there were no significant mean Goodenough vs. WISC VIQ or PIQ differences for the Familial Group. However, the mean Goodenough IQ turned out to be significantly higher than the mean FSIQ for this group (See Table 7, p. 46). Sign Test analyses of Goodenough vs. WISC differential scoring by individuals were all non-significant.

Hypothesis VII: The Organic Group exhibits a larger SB,L-M "scatter" than does the Familial Group, scatter being

measured by the number of items failed between the basal level and the level where all items are failed.

Hypothesis VII was accepted. The organics failed an average of 4.05 more intra-range items than did the familials. A t of 2.63 was obtained which was significant at the .01 level of significance (See Table 8. See also Table G, Appendix for individual scatter scores).

Table 8 SB,L-M Scatter (Mean number of items failed between basal and ceiling*)

	Familial	Organic	t	p
Mean	7.30	11.35	2.63	.01
SD	4.12	5.90		

*Ceiling is defined as the level where all items are failed by a given individual.

Since a significant scatter was obtained, an attempt was next made to determine at what point, if any, in the number of intra-range item failures, the organics and familials could be differentiated. A X^2 of 4.70 ($p = .05$) indicated the point to be between eight and nine items failed. Sixteen of the 26 organics failed nine items or more while only five of the twenty familials did so.

The question next arose as to whether the number of items failed occurred within a few age levels or whether

the range of SB,L-M age levels itself was greater for the organics than for the familials. Accordingly, the number of age levels between the basal level and the level at which all items were failed was found to be significantly greater for the Organic Group than for the Familial Group. The mean number of age levels covered by the Organic Group was 3.58 as compared with 2.50 for the Familial Group. A $\underline{t}$ of 2.61 was found to be significant at the .05 level of significance. However, analysis by X^2 showed no specific number of intra-range age levels at which the organics and familials could be differentiated.

Hypothesis VIII: The SB,L-M items which differentiate between the Organic and Familial Groups are VII,3 Copying a Diamond; IX,3 Memory for Designs I (1st design); and VI,6 Maze Tracing.

Hypothesis VIII was rejected. The items which were expected to differentiate between the groups failed to do so. However, another item, VIII,2 Memory for Stories: The Wet Fall, differentiated between the two groups beyond the .001 level of significance by Fisher's Exact Probability Test. The organics tended to pass this item, fifteen passes to eight failures, whereas the familials tended to fail it, one pass to fourteen failures (See Table H, Appendix for other item pass-fail frequencies). It must be noted that the hypothesis concerning item IX,3 Memory for Designs I

could not be tested because an insufficient number of familials attempted this item.

Hypothesis IX: Both the mean WISC PIQ and the mean Goodenough IQ are significantly lower for the Organic Group than for the Familial Group due to posited perceptual and body image difficulties of the Organic Group.

Hypothesis IX was rejected (See Table 9). Although the differences between the groups were in the predicted direction, they were not large enough to be significant. It is interesting to note that in all other IQ comparisons, the organics scored higher than the familials, significantly so in the case of the SB,L-M.

Table 9 SB,L-M, WISC, and Goodenough Mean Differences Between the Familial and Organic Groups

	Familial	Organic	Diff.	t	p
SB,L-M	54.80	58.54	3.74	2.04	.05*
WISC					
VIQ	54.25	58.11	3.86	1.46	ns*
PIQ	58.15	57.08	1.07	0.32	ns
FSIQ	51.85	53.46	1.61	0.56	ns*
Goodenough	57.65	55.54	2.13	0.61	ns

*Two-tailed since no specific prediction of direction was made.

In addition to the results of the specific hypotheses, two areas remain to be considered, first the comparison between the subjects' present SB,L-M IQs and their previous IQs, and second, the comparison between the SB,L-M and SB,L. Actually these two areas are closely related since the previous IQs were mainly SB,L IQs.

As can be seen from Table 10, p. 54, for both groups the mean SB,L-M was significantly below both the mean previous IQ and the mean SB,L IQ. For the SB,L vs. SB,L-M difference, Sign Tests were then used and the results for both organics and familials were significant. Of the 21 organics for whom SB,L IQs were available, twelve scored lower on the SB,L-M, four scored higher and there were five ties ($p = .038$). Of the sixteen familials, twelve scored lower on the SB,L-M, three scored higher and there was one tie ($p = .018$). Thus the trend is set by many individuals rather than a few highly deviant scorers.

Table 10 Comparison of Means of the SB,L-M; All Previous IQs; and the Previous SB,L IQs

	All Previous IQs		SB,L-M	Diff.	t	p
Organics	n=26	61.15	58.54	2.61	2.10	.05
Familials	n=20	61.05	54.80	6.25	4.66	.001
	Previous SB,L IQs		Adjusted SB,L-M *	Diff.	t	p
Organics	n=21	61.05	57.71	3.34	2.54	.01
Familials	n=16	60.00	53.62	6.38	3.85	.005

*SB,L-M IQs of subjects whose previous IQs were WISC were omitted in the computation of these means. The effect is a comparison of the SB,L and the SB,L-M using the same subjects.

DISCUSSION OF RESULTS

The discussion of results will relate to each hypothesis in succession. Again, for continuity and ease of reading, each hypothesis is repeated, and again, pertinent tangential material is introduced wherever appropriate.

Hypothesis I: (For the Familial and Organic Groups considered separately), The SB,L-M is significantly correlated with the WISC Verbal, Performance, and Full Scales.

The SB,L-M vs. WISC correlations seem to bear out, in a relative manner at least, what was previously noted about the factorial composition of these tests. The fact that the highest correlations occurred between the SB,L-M and the WISC Full and Verbal Scales is consistent with the studies of McNemar (1942) and Cohen (1959) which showed these tests to be highly saturated with g. The lower SB,L-M vs. PIQ correlations would be consistent with Cohen's finding that the WISC Performance Scale entails a Perceptual Organization factor as well as g.

With regard to magnitude, the correlations for the Organic Group are not out of line with previous research utilizing the 1937 SB (See Table 1, p. 18). However, the lower correlations for the Familial Group are not at all impressive, the SB,L-M vs. PIQ being non-significant.

These results may be a function of sample size, or of some unknown peculiarity of this particular sample. The non-significant SB,L-M vs. PIQ correlation may be due to the fact that the SB,L-M has more verbal items included than did the SB,L in order to boost its ability to measure g. These verbal items are included at the expense of performance items and, hence, the relationship with any measure of performance ability would seem necessarily to be lower. Which of these speculations is closest to the truth cannot be ascertained at this time and would be an area for further research.

Hypothesis II: The mean SB,L-M IQ is significantly lower than the mean IQ of the WISC Verbal, Performance, and Full Scales except in the case of the Organic Group where the SB,L-M vs. PIQ mean difference is not expected.

The rejection of this hypothesis is inconsistent with the previous SB,L research. With the exception of the SB,L-M vs. PIQ mean difference trend, all other WISC means were lower; significantly lower in the case of the mean FSIQ for both groups. The value of these findings is somewhat diminished by the fact that the standard errors of the means were large in relation to the mean differences, and by the finding that the SB,L-M vs. FSIQ mean difference became non-significant when non-extrapolated IQs were used. However, even when non-extrapolated IQs were used, the

SB,L-M vs. FSIQ mean difference remained significant for the organics, and analysis of individual scoring indicated a significantly lower scoring trend on the WISC Full Scale for both groups. The trend would, therefore, seem to be a real one rather than an artifact of extrapolation.

The SB,L-M mean IQs of the present study are not much out of line with the 55.38 to 62.5 range of SB,L IQ means in the previous studies cited (See Table 1, p. 18). However, most of the WISC means are considerably below those of the previous studies. The explanation for this phenomenon remains obscure. It may be due to some peculiarity of the present sample even though the CAs and SB IQs seem consistent with those of the previous samples.

The point to be made then, for Hypothesis II is that mental defectives as described in this study tend to score about three to five IQ points lower on the WISC Full Scale than on the SB,L-M. This deviation from previous findings may be due to some peculiarity of the particular sample used, or it may represent a difference in structure between the SB,L and the SB,L-M. In any case, the trend for lower WISC scoring would be of considerable import in institutions where IQs from different tests are considered equivalent and where IQ cut-off points are maintained in the selection of candidates for particular types of training or even for consideration for discharge.

Hypothesis III: (For the Familial and Organic Groups considered separately), The SB,L-M is significantly correlated with the Goodenough.

The non-significant SB,L-M vs. Goodenough correlations for both groups seem to be at odds with previous studies (See Table 3, p. 26). However, these correlations tend to support the contention that the Goodenough is a performance test rather than a measure of general intelligence. Further support is gained from the results of Hypothesis V (See discussion below). Speculation as to the deviation from previous SB,L studies involves possible differences of the present sample as well as the fact that the SB,L-M has fewer performance items than the SB,L.

Hypothesis IV: For the Organic Group, the mean Goodenough IQ is significantly below the mean SB,L-M IQ due to posited perceptual and body image difficulties of this group.

Although a trend existed in the predicted direction, the fact that it was non-significant leads to a question of the exact nature of this particular organic sample. It will be remembered that by far the majority, 20 of the 26 subjects consisted of developmental anomalies rather than any sort of direct brain-injury. The situation may be that this type of organic, the third class of

Barnett, et al. (1960), manifests the same lack of perceptual difficulty which Gallagher (1957) noted in his study. This conclusion is further born out by the results of Hypotheses VI, VIII and IX discussed below.

Hypothesis V: (For the Familial and Organic Groups considered separately), The Goodenough is significantly correlated with the WISC Verbal, Performance, and Full Scales.

The pattern of the Goodenough vs. WISC correlations seems to be in line with what has been previously said about the performance nature of the Goodenough. For both groups, the highest correlations were between the Goodenough and the PIQ and the lowest (non-significant) were with the VIQ. The fact that the Goodenough vs. FSIQ correlations were also significant is at first sight rather enigmatic since the FSIQ is supposedly a measure of g (Cohen, 1959) and since the Goodenough was not significantly correlated with SB,L-M which is also a measure of g. However, the explanation of this phenomenon may lie in the fact that half of the subtests which make up the WISC Full Scale are performance subtests whereas the performance items on the SB,L-M are minimal. From this, one would expect the Goodenough to be more highly related to the WISC Full Scale than to the SB,L-M.

Hypothesis VI: For the Organic Group, the mean Goodenough IQ is significantly below the mean WISC Verbal and Full Scale IQs due to posited perceptual and body image difficulties of this group.

Again the expected effects of perceptual difficulties did not materialize for the Organic Group since no significant mean differences appeared between the Goodenough and the WISC Verbal and Full Scale IQs. Analysis for individual scoring by use of the Sign Test likewise produced non-significant results. This seems to further reinforce the idea that organics with minor developmental anomalies do not as a rule develop the perceptual distortions associated with other types of organicity.

Of tangential interest is the fact that the Goodenough was significantly higher than the WISC FSIQ for Familials. Although Sign Test results were non-significant, drawing the human figure seems to be a facile mode of expression for at least some familials.

Hypothesis VII: The Organic Group exhibits a larger SB,L-M "scatter" than does the Familial Group, scatter being measured by the number of items failed between the basal level and the level where all items are failed.

The results of this study agree with the Berko study (1955) in that the organics experienced more intra-range item failure than the familials. The organics also covered a greater number of age levels. What this seems to indicate in terms of the intellectual functioning of organics is that they do not experience an "across the board" intellectual deficit. Rather, their deficit depends upon the extent and location of injury or malformation.

The finding that organics tended to fail nine or more intra-range items while the familials tended to fail eight or less, can be taken only as a very tentative diagnostic criterion. The overlap was considerable, 25% in the case of familials failing nine or more items and 38% in the case of organics failing eight or less. Never-the-less, the finding can be used as a diagnostic sign in conjunction with other signs in the differential diagnosis of organicity.

Hypothesis VIII: The SB,L-M items which differentiate between the Organic and Familial Groups are VII,3 Copying a Diamond; IX,3 Memory for Designs I (1st design); and VI,6 Maze Tracing.

Since the organics exhibited greater intra-range item failure, the question arises as to whether there were any particular items which they failed more significantly than did the familials. The answer is that there were no

such items, not even the perceptual-motor Copying a Diamond item. This again points out the lack of perceptual disability in this particular sample of organics.

The only item which differentiated between the two groups was VIII,2 Memory for Stories: The Wet Fall, and the failures in this case were on the part of the familials. A tentative explanation of the familials' failures can be seen by the nature of the item and its relation to their family background. The item consists of a story about a boy, his sister and a pony which is given to them by their father. After the story some questions are asked, one of which is "Who gave the pony to them?" Only three of the 23 organics who attempted failed to answer this question correctly whereas nine of the fifteen familials said they did not know the answer or named some other male figure besides the father. This failure to remember the answer, "their father" may be an act of repression because in general fathers of familial defectives are almost by definition incompetent and undesireable as heads of households. Thus, unpleasant emotion is aroused by the thought of father giving anything, either because he has never actually given anything of value, or because what he has given has been punishment. This unpleasant emotion leads to the act of repression and failure to answer the question correctly. To be sure, this explanation is speculative, but it does not seem altogether unrealistic.

Hypothesis IX: Both the mean PIQ and the mean Goodenough IQ are significantly lower for the Organic Group than for the Familial Group due to posited perceptual and body image difficulties of the Organic Group.

The fact that Hypothesis IX was totally rejected serves to reinforce what has been previously said about the apparent lack of perceptual and body image difficulties of this particular group of organics. Although a trend in the predicted direction was present, it was not significantly established.

In addition to the discussion to the specific hypotheses, a consideration of some other findings is in order. It will be remembered that both organics and familials scored SB,L-M IQs significantly lower than their previous IQs. Since these previous IQs were obtained mainly on the SB,L, the question can be raised as to whether a real difference exists between the SB,L and the SB,L-M. At first glance such a difference would seem to exist because, when the SB,L-M was compared directly with the SB,L by omitting subjects with previous WISC IQs, the differences remained significant. However, when one considers the fact that the present WISC and Goodenough mean IQs are also considerably below the previous IQs, the difference seems to be one of deterioration in intellectual functioning rather than a difference in SB,L vs. SB,L-M test structure. This

deterioration might be due to the constricting effects of institutional living, or possibly due to an earlier termination of intellectual growth (MA) in these subjects than in the public at large. In the latter case, the fact that the organics obtained significantly higher SB,L-M IQs than did the familials would indicate that MA growth may terminate even earlier for familials than for organics.

With the results of this study in mind, several points for future research can be envisioned. Cross validation would be necessary using larger samples of composition similar to those of the present study. If subject by subject matching could be obtained on the pertinent variables, perhaps more familial vs. organic test differences could be discovered than were found by merely equating the groups.

Another point would be greater emphasis on the study of mental defectives who manifest minor developmental anomalies. One might wonder if they represent an exaggerated form of familial mental deficiency since both of these types of deficiency have suspected hereditary origins.

The last area for suggested future research is SB,L-M, WISC, and Goodenough comparisons using subjects of both normal and superior intelligence. Only in this way would the full range of relationships between these tests become known.

SUMMARY

The purpose of this study was to determine the correlational relationships and mean differences between the Stanford-Binet, Form L-M (SB,L-M), the Wechsler Intelligence Scale for Children (WISC), and the Goodenough Draw-A-Man Test (Goodenough) using mentally retarded children of organic and familial etiologies. Differential SB,L-M functioning of these two types of subjects was also studied.

Two groups of institutionalized children were selected; twenty in the Familial Group and twenty-six in the Organic Group. Groups were equated by previous IQ, sex, chronological age, and length of institutionalization. Subjects were selected according to the following criteria: a) previous IQ between 50 and 70, b) current enrollment in institutional school programs, c) not crippled, blind, deaf, untidy, or subject to seizures, d) not tested during the previous year.

The three intelligence tests were administered in counter-balanced order and in accordance with the standardized instructions found in their respective manuals. Two recently reported minor scoring innovations were used where applicable in the calculation of IQs. Differences in organic and familial SB,L-M functioning were examined by performance on individual test items and by patterns of item failure ("scatter").

Findings for the Familial Group were:

a) SB,L-M vs. WISC correlations significant except in the case of the Performance IQ, and, contrary to expectation, a significantly lower mean WISC Full Scale IQ.

b) SB,L-M vs. Goodenough correlation and mean difference nonsignificant.

c) Goodenough vs. WISC correlations significant for the WISC Full and Performance Scales, and a significantly higher mean Goodenough IQ as compared with the WISC Full Scale IQ.

Findings for the Organic Group were:

a) SB,L-M vs. WISC correlations all significant and, contrary to expectation, a significantly lower WISC Full Scale IQ.

b) SB,L-M vs. Goodenough correlation and mean difference nonsignificant.

c) Goodenough vs. WISC correlations significant for the WISC Full and Performance Scales, but no significant mean IQ differences.

The following differential functioning between Organic and Familial Groups was found:

a) Organics exhibited significantly greater SB,L-M "scatter" than did familials, scatter being measured by number of intra-range item failures and by number of intra-range age levels.

b) Only one SB,L-M item distinguished between the groups, VIII,2 Memory for Stories: The Wet Fall. Failures of this item were by familials,and a repressive tendency was suggested as the cause of failure.

c) SB,L-M mean IQ was significantly higher for organics than for familials. No significant organic vs. familial mean IQ differences were found on the WISC or Goodenough.

REFERENCES

American Association on Mental Deficiency. Etiological Classifications of the Committee on Nomenclature. Mimeo. Lapeer State Home and Training School, Lapeer, Mich., 1960. Based on Herber, R. A manual on terminology and classification in mental retardation. Amer. J. ment. Defic., 1959, 64, monogr. suppl. No. 2.

Anderson, J. M. In Buros, O. K. (Ed), The fourth mental measurements yearbook. Highland Park, N.J.: Gryphon Press, 1953, pp. 363-364.

Ansbacher, H. L. The Goodenough Draw-A-Man Test and Primary Mental Abilities. J. consult. Psychol., 1952, 16, 176-180.

Arnold, F. C. & Wagner, Winifred K. A comparison of Wechsler Children's Scale and Stanford-Binet scores for eight- and nine-year olds. J. exp. Educ., 1955, 24, 91-94.

Barnett, C. D., Ellis, N. R. & Pryer, Margaret W. Learning in familial and brain-injured defectives. Amer. J. ment. Defic., 1960, 64, 894-901.

Beck, H. S. & Lam, R. L. Use of the WISC in predicting organicity. J. clin. Psychol., 1955, 11, 154-158.

Benda, C. E. Developmental disorders of mentation and cerebral palsies. New York: Grune & Stratton, 1952.

Bender, Lauretta. Child Psychiatric Techniques. Springfield, Ill.: Thomas, 1952.

Bender, Lauretta. Psychopathology of children with organic brain disorders. Springfield, Ill.: Thomas, 1956.

Bensberg, G. J. A test for differentiating endogenous and exogenous mental defectives. Amer. J. ment. Defic., 1950, 54, 502-506.

Berko, M. J. A note on "Psychometric Scatter" as a factor in the differentiation of exogenous and endogenous mental deficiency. Cerebral Palsy Rev., 1955, 16, 20.

Birch, J. W. The Goodenough drawing test and older mentally retarded children. Amer. J. ment. Defic., 1949, 54, 218-224.

Clarke, Ann M. & Clarke, A. D. B. (Eds). Mental deficiency, the changing outlook. Glencoe, Ill.: Free Press, 1958.

Cohen, B. D. & Collier, Mary J. A note on the WISC and other tests of children six to eight years old. J. consult. Psychol., 1952, 16, 226-227.

Cohen, J. The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. J. consult. Psychol., 1959, 23, 285-299.

Eisenson, J. When and what is aphasia? In Wood, Nancy E. Language development and language disorders: A compendium of lectures. Monogr. soc. Res. Child Developm., 1960, 25, No. 3, 89-95.

Frandsen, A. N. & Higginson, J. B. The Stanford-Binet and the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 236-238.

Gallagher, J. J. A comparison of brain-injured and non-brain-injured mentally retarded children on several psychological variables. Monogr. soc. Res. Child Develpm., 1957, 22, No. 2.

Gault, Una. Factorial patterns on the Wechsler Intelligence Scales. Aust. J. Psychol., 1954, 6, 85-90.

Gehman, Ila H. & Matyas, R. P. Stability of the WISC and Binet tests. J. consult. Psychol., 1956, 20, 150-152.

Goodenough, Florence L. Measurement of intelligence by drawings. Yonkers, N.Y.: World Book, 1926.

Harlow, J. E. Jr., Price, A. C., Tatham, L. J. & Davidson, J. F. Preliminary study of comparison between Wechsler Intelligence Scale for Children and Form L of Revised Stanford-Binet Scale at three age levels. J. clin. Psychol., 1957, 13, 72-73.

Havighurst, R. J., Gunther, Minna K. & Pratt, Inez E. Environment and the Draw-A-Man Test: The performance of Indian children. J. abnorm. soc. Psychol., 1946, 41, 50-63.

Higgins, C. & Sivers, Catherine H. Comparison of Stanford-Binet and Colored Raven Progressive Matricies IQs for children with low economic status. J. consult. Psychol., 1958, 22, 465-468.

Hoakley, Z. Pauline & Frazeur, Helen A. Significance of psychological test results of exogenous and endogenous children. Amer. J. ment. Defic., 1945, 50, 263-271.

Holland, G. A. A comparison of the WISC and Stanford-Binet IQs of normal children. J. consult. Psychol., 1953, 17, 147-152.

Hutt, M. L. & Gibby, R. G. The mentally retarded child. Boston: Allyn & Bacon, 1958.

Israelite, Judith. A comparison of the difficulty of items for intellectually normal children and mental defectives on the Goodenough drawing test. Amer. J. Orthopsychiat., 1936, 6, 494-503.

Johnson, A. P., Arthur, A. E., & Lahey, T. H. The Goodenough Test as an aid to interpretation of children's school behavior. Amer. J. ment. Defic., 1950, 54, 516-520.

Kaliski, Lotte. The brain-injured child - Learning by living in a structured setting. Amer. J. ment. Defic., 1959, 63, 688-695.

Krugman, Judith I., Justman, J., Wrightstone, J. W., & Krugman, M. Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children. J. consult. Psychol., 1951, 15, 475-483.

Lewis, R. S. The other child. New York: Grune & Stratton, 1951.

Littell, W. M. The Wechsler Intelligence Scale for Children: Review of a decade of research. Psychol. Bull., 1960, 57, 132-156.

Lord, Elizabeth & Wood, Louise. Diagnostic values in a visuo-motor test. Amer. J. Orthopsychiat., 1942, 12, 414-428.

McElwee, Edna W. The reliability of the Goodenough intelligence test used with sub-normal children fourteen years of age. J. appl. Psychol., 1932, 16, 217-218.

McHugh, G. Relationship between the Goodenough drawing a man test and the 1937 revision of the Stanford-Binet Test. J. educ. Psychol., 1945, 36, 119-124.

McNemar, Q. The revision of the Stanford-Binet scale. New York: Houghton Mifflin, 1942.

Mitchell, Anna C. A new maximum CA for the Draw-A-Man Test. J. consult. Psychol., 1959, 23, 655-557.

Murphy, Mary M. A Goodenough scale evaluation of human figure drawings of three non-psychotic groups of adults. J. clin. Psychol., 1956, 12, 397-399.

Nale, S. The children's Wechsler and the Binet on 104 mental defectives at the Polk State School. Amer. J. ment. Defic., 1951, 56, 419-423.

Ogdon, D. P. WISC I.s for the mentally retarded. J. consult. Psychol., 1960, 24, 187-188.

Pastovic, J. J. & Guthrie, G. M. Some evidence on the validity of the WISC. J. consult. Psychol., 1951, 15, 385-386.

Peterson, J. Early conceptions and tests of intelligence. Yonkers, N.Y.: World Book, 1925.

Sandercock, Marian G. & Butler, A. J. An analysis of the performance of mental defectives on the Wechsler Intelligence Scale for Children. Amer. J. ment. Defic., 1952, 57, 100-105.

Sarason, S. B. Psychological problems in mental deficiency. (3rd ed.) New York: Harper, 1959.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Sloan, W. & Schneider, B. A study of the Wechsler Intelligence Scale for Children with mental defectives. Amer. J. ment. Defic., 1951, 55, 573-575.

Stacey, C. L. & Levin, Janice. Correlational analysis of scores of subnormal subjects on the Stanford-Binet and the Wechsler Intelligence Scale for Children. Amer. J. ment. Defic., 1951, 55, 590-597.

Strauss, A. A. & Werner, H. Disorders of conceptual thinking in the brain-injured child. J. nerv. ment. Dis., 1942, 96, 153-172.

Strauss, A. A. Ways of thinking in brain-crippled deficient children. Amer. J. Psychiat., 1943, 100, 639-647.

Strauss, A. A. & Lehtinen, Laura E. Psychopathology and education of the brain-injured child. (vol. I) New York: Grune & Stratton, 1947.

Terman, L. M. The measurement of intelligence. Boston: Houghton Mifflin, 1916.

Terman, L. M. & Merrill, Maude A. Measuring intelligence. Boston: Houghton Mifflin, 1937,

Terman, L. M. & Merrill, Maude A. Stanford-Binet intelligence scale. Boston: Houghton Mifflin, 1960.

Tredgold, R. F. & Soddy, K. A text-book of mental deficiency. Baltimore: Williams & Wilkins, 1956.

Vanderhorst, Leonette, Sloan, W. & Bensberg, G. J. Performance of mental defectives on the Wechsler-Bellevue and the WISC. Amer. J. ment. Defic., 1953, 57, 481-483.

Warren, Sue A. & Collier, H. L. Suitability of the Columbia Mental Maturity Scale for mentally retarded institutionalized females. Amer. J. ment. Defic., 1960, 64, 916-920.

Wechsler, D. Wechsler intelligence scale for children. New York: Psychological Corp., 1949.

Wechsler, D. The measurement and appraisal of adult intelligence. (4th ed.) Baltimore: Williams & Wilkins, 1958.

Weider, A., Noller, P. A., & Schramm, T. A. The Wechsler Intelligence Scale for Children and the Revised Stanford-Binet. J. consult. Psychol., 1951, 15, 330-333.

Yepsen, L. N. The reliability of the Goodenough drawing test with feeble-minded subjects. J. educ. Psychol., 1929, 20, 448-451.

APPENDIX

In the following tables where information is given for individual subjects, the first ten subjects for the familial group are female, the last ten are male. For the organic group the first thirteen subjects are female and subjects fourteen through twenty six are male.

TABLE A Previous IQs

Familial				Organic			
Female		Male		Female		Male	
S	IQ	S	IQ	S	IQ	S	IQ
1	66	11	56	1	62 W	14	62
2	69	12	67	2	59	15	65
3	51	13	61	3	67	16	70 W
4	65	14	65 W	4	58	17	56
5	64 W*	15	59	5	52	18	65
6	61	16	64	6	65	19	53 W
7	62	17	67 W	7	58	20	67
8	58	18	56	8	63	21	66
9	52	19	65 W	9	69 W	22	67
10	63	20	50	10	60	23	58
				11	68	24	55
				12	56	25	54
				13	61	26	54 W

*W = WISC FSIQs. Approximately 20% of the previous IQs for each group were WISC FSIQs. The other 80% were SB,L IQs.

Familial Mean 61.05, SD 5.61, Range 50 - 69

Organic Mean 61.15, SD 5.44, Range 52 - 70

TABLE B Stanford-Binet, L-M CAs, MAs, and IQs for Familial Subjects

S	CA	MA	IQ	S	CA	MA	IQ
1	10-1	5-6	54	11	14-5	6-0	47
2	11-0	6-4	60	12	10-4	5-6	54
3	12-7	5-10	51	13	15-7	9-10	66
4	13-4	6-0	50	14	12-7	6-8	57
5	15-3	8-10	62	15	10-7	5-0	49
6	9-1	5-1	54	16	13-6	6-8	54
7	15-6	7-0	49	17	11-0	6-4	60
8	13-5	5-10	49	18	9-8	6-2	63
9	15-1	6-8	49	19	10-2	6-0	59
10	12-10	6-10	58	20	15-5	7-2	51

CA Mean 12.57, SD 2.18, Range 9-1 - 15-7

MA Mean 6.5 , SD 1.15, Range 5-0 - 9-10

IQ Mean 54.80, SD 5.53, Range 47 - 66

TABLE C· Stanford-Binet, L-M CAs, MAs, and IQs for Organic Subjects

S	CA	MA	IQ	S	CA	MA	IQ
1	14-2	8-4	63	14	15-2	8-8	62
2	14-9	8-6	62	15	14-10	8-10	65
3	9-7	6-10	70	16	12-7	8-8	71
4	12-0	7-4	64	17	11-11	5-10	53
5	11-7	4-6	44	18	11-1	7-0	65
6	10-9	5-8	55	19	15-4	7-4	52
7	12-1	5-10	53	20	12-8	8-0	66
8	15-1	7-6	54	21	15-7	8-8	59
9	9-10	5-11	59	22	10-10	5-6	53
10	12-10	5-8	50	23	10-8	6-0	58
11	11-6	6-2	57	24	11-0	5-9	55
12	13-7	6-6	53	25	12-0	7-0	62
13	9-3	4-11	52	26	12-6	7-8	65

CA Mean 12.43, SD 1.88, Range 9-3 - 15-7

MA Mean 6.9 , SD 1.28, Range 4-6 - 8-10

IQ Mean 58.54, SD 6.62, Range 44 - 71

TABLE D WISC Verbal, Performance, and Full Scale IQs for Familial Subjects

S	VIQ	PIQ	FSIQ	S	VIQ	PIQ	FSIQ
1	62	51	53	11	41	55	43
2	52	55	49	12	50	61	51
3	55	46	46	13	63	62	59
4	51	51	46	14	58	82	67
5	56	61	54	15	52	51	47
6	46	43	39	16	62	79	67
7	61	55	54	17	63	71	64
8	50	46	43	18	51	55	48
9	52	46	44	19	60	71	62
10	55	54	50	20	45	68	51

FSIQ Mean 51.85, SD 8.19, Range 39 - 67

VIQ Mean 54.25, SD 6.37, Range 41 - 63

PIQ Mean 58.15, SD 11.06, Range 43 - 82

TABLE E WISC Verbal, Performance, and Full Scale IQs for Organic Subjects

S	VIQ	PIQ	FSIQ	S	VIQ	PIQ	FSIQ
1	55	72	59	14	57	60	54
2	62	58	56	15	67	71	66
3	71	55	60	16	84	74	77
4	65	67	62	17	44	43	38
5	39	32	29	18	62	50	52
6	62	69	62	19	45	62	48
7	56	50	48	20	76	75	73
8	47	62	50	21	57	62	56
9	71	62	64	22	55	62	54
10	47	46	41	23	58	58	54
11	53	51	48	24	53	39	41
12	47	43	40	25	53	54	49
13	62	46	50	26	63	61	59

FSIQ Mean 53.46, SD 10.83, Range 29 - 77

VIQ Mean 58.11, SD 10.48, Range 39 - 84

PIQ Mean 57.08, SD 11.24, Range 32 - 75

TABLE F Goodenough MAs and IQs for Familial and Organic Subjects

Familials			Organics		
S	MA	IQ	S	MA	IQ
1	6-9	67	1	8-9	64
2	5-9	52	2	7-9	55
3	5-6	44	3	4-0	44
4	6-0	45	4	6-9	56
5	6-6	45	5	5-0	45
6	4-6	49	6	6-3	58
7	7-0	48	7	7-6	62
8	6-9	51	8	8-0	55
9	8-3	57	9	5-0	51
10	6-3	49	10	6-0	47
11	5-0	36	11	5-3	46
12	8-3	80	12	6-6	48
13	9-3	63	13	4-3	46
14	8-0	64	14	9-0	62
15	6-0	57	15	7-6	53
16	11-6	86	16	10-3	81
17	7-3	66	17	6-0	50
18	6-3	65	18	6-0	54
19	6-9	66	19	9-6	65
20	9-3	63	20	6-9	53
			21	5-6	37
			22	5-9	53
			23	8-9	82
			24	4-6	41
			25	8-3	69
			26	8-3	67

Familial

MA Mean 7.0 , SD 1.65, Range 4-6 - 11-6

IQ Mean 57.65, SD 12.52, Range 36 - 86

Organic

MA Mean 6.8 , SD 1.70, Range 4-0 - 10-3

IQ Mean 55.54, SD 11.13, Range 37 - 82

TABLE G Stanford-Binet, L-M Scatter Scores for Familial and Organic Subjects (Number of items failed and number of levels between basal and ceiling*)

Familials			Organics		
S	Number Items Failed	Number Levels	S	Number Items Failed	Number Levels
1	3	1	1	28	7
2	4	2	2	3	2
3	1	1	3	7	2
4	6	2	4	22	6
5	13	4	5	6	3
6	6	2	6	8	2
7	6	3	7	7	2
8	1	1	8	21	6
9	8	2	9	14	5
10	13	3	10	2	1
11	6	2	11	12	4
12	10	4	12	9	3
13	7	3	13	9	5
14	8	2	14	14	4
15	8	4	15	13	3
16	8	3	16	14	5
17	4	1	17	14	6
18	5	2	18	6	2
19	12	3	19	16	4
20	17	5	20	12	3
			21	14	3
			22	10	4
			23	6	2
			24	8	3
			25	12	3
			26	8	3

*Ceiling is defined as the level at which all items were failed.

Familials	Number items failed	Mean 7.30, SD 4.12
	Number levels	Mean 2.50, SD 1.15
Organics	Number items failed	Mean 11.35, SD 5.90
	Number levels	Mean 3.58, SD 1.55

TABLE H SB,L-M Item Discrimination Between Familial and Organic Groups

Item		Familial		Organic		
		Pass	Fail	Pass	Fail	
V	1	12	2	13	2	
	2	14	0	15	0	
	3	14	0	15	0	
	4	13	1	11	4	
	5	14	0	12	3	
	6	13	1	14	1	
VI	1	11	7	11	10	
	2	13	5	17	4	
	3	16	2	17	4	
	4	15	3	15	6	
	5	10	8	15	6	
	6	13	5	16	5	
VII	1	10	10	14	9	
	2	6	14	10	13	
	3	8	12	10	13	
	4	11	9	8	15	
	5	8	12	15	8	
	6	4	16	9	14	
VIII	1	4	11	12	11	
	2	1	14	15	8	.001
	3	1	14	4	19	
	4	4	11	2	21	
	5	3	12	7	16	
	6	4	11	8	15	

TABLE I Downward Extension of WISC Verbal and Performance Scales

Verbal Scale*		
Scaled Score	IQ	IQ rounded
1	38.75	39
2	40.00	40
3	41.25	41
4	42.50	42
5	43.75	44
Performance Scale*		
1	31.89	32
2	33.28	33
3	34.67	35
4	36.06	36
5	37.45	37
6	38.84	39
7	40.23	40
8	41.62	42
9	43.01	43

*Verbal
One scaled score unit = 1.25 VIQ units.

*Performance
One scaled score unit = 1.39 PIQ units.

TABLE J Sign Test Results for Familial and Organic Subjects

	SB,L-M vs VIQ	SB,L-M vs PIQ	SB,L-M vs FSIQ	SB,L-M vs G	G vs VIQ	G vs PIQ	G vs FSIQ
Familials							
+'s	11	10	5	11	6	11	7
-'s	9	10	14	9	12	9	13
Ties	-	-	1	-	2	-	-
N =	20	20	19	20	18	20	20
p =	.412	.588	.032	.412	.119	.412	.132
Organics							
+'s	10	11	6	10	14	12	11
-'s	14	14	20	14	11	13	14
Ties	2	1	-	2	1	1	1
N =	24	25	26	24	25	25	25
p =	.271	.345	.005	.271	.345	.500	.345