1957

The Predictive Value of the Cattell Infant Intelligence Scale

Patricia Bledsoe

Loyola University Chicago

Recommended Citation

Bledsoe, Patricia, "The Predictive Value of the Cattell Infant Intelligence Scale" (1957). Dissertations. 446.
https://ecommons.luc.edu/luc_diss/446
THE PREDICTIVE VALUE OF THE CATTELL
INFANT INTELLIGENCE SCALE

by
Patricia Bledsoe

A Thesis Submitted to the Faculty of the Graduate School
of Loyola University in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION ..................................................</td>
<td>1</td>
</tr>
<tr>
<td>Current status of infant intelligence testing—Discussion of IQ inconstancy—Primary purpose of present research—Related investigation.</td>
<td></td>
</tr>
<tr>
<td>II. REVIEW OF THE LITERATURE - I ..................................</td>
<td>13</td>
</tr>
<tr>
<td>History of infant tests—Test content and factors influencing infant test construction—Reliability and validity of infant scales—Integrated summary of the literature.</td>
<td></td>
</tr>
<tr>
<td>III. REVIEW OF THE LITERATURE - II ...............................</td>
<td>42</td>
</tr>
<tr>
<td>The variability of mental growth—Environmental correlates of intelligence—Summary.</td>
<td></td>
</tr>
<tr>
<td>IV. DESIGN OF THE RESEARCH ........................................</td>
<td>62</td>
</tr>
<tr>
<td>Discussion of the Cattell Infant Intelligence Scale—Subjects of the present study—Status of true and adoptive parents—Statistical methodology.</td>
<td></td>
</tr>
<tr>
<td>V. ANALYSIS OF RESULTS ............................................</td>
<td>84</td>
</tr>
<tr>
<td>Analysis of Cattell and Stanford-Binet Scale performance—Correlation between Cattell and Stanford-Binet IQ's—Analysis of IQ point changes—Correlations between intelligence, and background and environmental variables.</td>
<td></td>
</tr>
<tr>
<td>VI. SUMMARY AND CONCLUSIONS .....................................</td>
<td>97</td>
</tr>
<tr>
<td>Restatement of the problem—Propose of research and methods used—Conclusions.</td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Title</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>I.</td>
<td>RETEST CORRELATIONS FROM PUBLISHED VALIDITY STUDIES OF SIX-MONTH-OLD INFANT TEST PERFORMANCES</td>
</tr>
<tr>
<td>II.</td>
<td>MEAN IQ'S OF ADOPTED CHILDREN AS REPORTED IN PUBLISHED STUDIES</td>
</tr>
<tr>
<td>III.</td>
<td>OCCUPATIONAL STATUS OF THE TRUE MOTHERS</td>
</tr>
<tr>
<td>IV.</td>
<td>MINNESOTA OCCUPATIONAL SCALE DISTRIBUTION OF 110 ADOPTIVE FATHERS AND 83 TRUE FATHERS</td>
</tr>
<tr>
<td>V.</td>
<td>EDUCATIONAL STATUS OF THE TRUE PARENTS AND THE ADOPTIVE PARENTS</td>
</tr>
<tr>
<td>VI.</td>
<td>CHANGES IN IQ POINTS</td>
</tr>
<tr>
<td>VII.</td>
<td>CORRELATIONS BETWEEN TRUE AND ADOPTIVE PARENT EDUCATION, SOCIO-ECONOMIC STATUS, AND THE CHILD'S IQ</td>
</tr>
<tr>
<td>VIII.</td>
<td>IQ'S OF 110 ADOPTED CHILDREN IN RELATION TO ADOPTIVE FATHER'S OCCUPATION</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

The writer wishes to express her sincere appreciation to Frank J. Kobler, Ph.D., Professor of Clinical Psychology, Loyola University, for his suggestions and direction of this study; to the Reverend Bernard M. Brogan, Director of the Catholic Home Bureau, Child-placing agency of the Catholic Charities of the Archdiocese of Chicago, for granting permission for the renewal of contact with families formerly known to the agency, as well as for the use of agency case records; to Ralph D. Bergen, M.D., Director of the Guidance Center of Catholic Charities, for his encouragement and his assistance in arranging for the gathering of the data; and to the children who served as the subjects, and their adoptive parents for their cooperation and their generous participation in this research.

Patricia Bledsoe
CHAPTER I

INTRODUCTION

The value of an infant scale as a clinical instrument depends almost entirely upon its ability to predict intelligence at later ages. The experimental psychologist working in the laboratory can be interested in examining infant behavior for its own sake, in order to investigate its nature or to attempt to isolate the beginnings of the more complicated higher order activities. Clearly, the tester of older age groups is concerned primarily with relating his findings to academic progress, work adjustment or some other feature of the current life situation. But the clinician who administers intelligence tests to infants has little or no interest in evaluating infant responses apart from the extent to which they forecast intellectual functioning in the school age child and the adult.

We know from indirect references in psychological and social casework literature that certain of the infant tests, in particular the Cattell Infant Intelligence Scale (20) and the Gesell Developmental Schedules (36), are in fairly common use in the clinic. Articles by Carter and Bowles (19), Escalona (26), Fischer (28), and Gallagher (32) refer to a regular clinical use of the test, usually for the purpose of determining the suitability of young infants for adoption. Recent prominent textbooks on psychological testing (1, 22, 68) have devoted special sections to the problems of infant testing, reflecting the increase of interest in the field.
Another likely area in which infant scales are being used is the diagnosis of mental deficiency, since the need for parental guidance and plans for training the retarded child make detection desirable at as early an age as possible.

In spite of this increased attention to infant tests, only the rather urgent need of adoption agencies for objectivity in evaluating the intellectual potentialities of the infant can account for their clinical usage, because research on infant scales has been limited and the findings on predictive efficiency have been unfavorable. Several infant tests have been published since Gesell (33) described his first group of developmental items in 1925, but very little follow-up work has been done on any of the scales and the studies which have been reported have not been of uniformly acceptable quality. Almost all have been conducted on very small groups and in some reports the results have not been clearly presented. But the more definitive studies have agreed that infant scales in their present stage of development are poor predictors of later intelligence. The highest correlation found for an age level under twelve months was .47, obtained by Nelson and Richard (59) who studied the six month items of the Gesell Schedules. Typically, the correlations have been much closer to zero, indicating that predictions based on early scores have little more than chance possibility of accuracy.

In investigating the predictive efficiency of infant scales, the common procedure has been to compare the infant scores with scores
obtained at a later age on a well established test, usually one of the forms of the Stanford-Binet Scales. But an inquiry into the relationship between early and later intelligence as measured by intelligence tests involves more than the mere application of statistics to two sets of IQ's and accepting the results as final evidence either in favor of or against the scale in question. There are certain factors of possible influence which the researcher must consider and perhaps retain as limitations upon his findings. Some of these derive from the testing devices used, while others pertain to the nature of intelligence itself. Since these introductory comments are prefatory to such a predictive study, it will be well to examine these factors further at this point.

It is now recognized that there can be considerable variability and inconsistency in the measurements of the intellectual development of any particular individual. There was a time when psychologists were convinced that scores on intelligence tests remained relatively the same, provided only that the tests were adequate and environmental conditions were not drastically changed. However, thinking about the so-called "constancy" of the IQ has been revised in recent years as research findings have continued to show that the IQ is not the stable index of brightness it was once considered to be, especially during the infancy and preschool periods. Such key studies as those by Bradway (15), Dearborn and Rothey (24), Honzik (49), and Bayley (10, 12) have indicated that individual variability occurs over long periods of time, although variability does decrease with
age. Perhaps the most striking evidence on this entire question can be found in the individual intelligence growth curves which have been published in the periodic reports on the Berkeley Growth Study (8, 10, 12). Wide fluctuations in intelligence were found for many of the children included in that study. Children who demonstrated early precocity sometimes proved later to be average or even slow learners, and sometimes the reverse was true, with some children showing initially slow development and a more rapid growth tempo later on. When the children in the Berkeley study were eight years of age, it was found that only a fifth of the group had maintained any stability in their relative status over the eight year period (12). Similar variations in measured intelligence can be found in the individual growth curves included in the Harvard Growth Study publications (24, 20).

As the evidence of variability in the young child accumulates, researchers are faced with the necessity of explaining it. It is probably unnecessary to point out that while IQ inconstancy and variability of intelligence are closely related concepts, they are not identical. Irregularities in an intelligence growth curve could conceivably be due almost entirely to the test instruments used. Under such conditions, intelligence as a basic property of the individual would remain the same, but successive tests would fail for one or more reasons to measure it in the same way, thus producing spurious changes in the IQ's. On the other hand, inconsistencies in IQ scores could reflect actual changes in the tempo or perhaps even the composition of the underlying intelligence.
The extent to which irregularities are a function of the test instruments is difficult to determine. Under present testing conditions different tests must be employed for the different age ranges in a longitudinal study. But tests are known to differ widely on several counts, such as kinds of abilities measured, relative difficulty at different age levels, susceptibility to environmental influences, and dependence upon social and cultural factors. In addition, standardization populations are different for different tests and scores are not directly comparable. Any or all of these factors might affect a particular measurement and create inconsistency or inconstancy of the IQ. Ideally, of course, these influences could only be controlled by a test which would measure the same abilities from birth on through maturity and for which scores would be expressed in standard units. Until such a test is found, if ever, these factors inherent in test instruments will continue to obscure actual irregularities which might be occurring.

An additional difficulty which is crucial for infant scale construction is the problem of identifying intelligent behavior in the infant and devising suitable test items to evaluate it. The activities usually considered to be manifestations of general intelligence—the abilities to retain and to recall, the abstract and to relate—can be presumed to exist in some potential form in the very young infant, but they are as yet unrealized due to immaturity of physical development and lack of experience. The scope of infant responses is limited. As far as can be
determined, most early activities are of a sensorimotor nature, and infant scales which test abilities under twelve months necessarily include many items which tap sensorimotor adjustments, despite special efforts to exclude them. Various studies have revealed the sensorimotor items of the infant scales to be poor predictors of later intelligence (6, 59). But there does remain the possibility that those aspects of infant behavior which do bear some direct relationship to later intelligence are still being overlooked; that further research may permit test constructors to isolate them and utilize them effectively as good predictors.

But even if our testing instruments were able to measure intelligence precisely, it is possible that inconsistencies in scores over a long period of time would occur as an innate characteristic of intellectual growth. Perhaps intelligence is composed of several kinds of abilities which are not all of the same strength in a given individual and which emerge at different times during the life span. Since infancy is the period of most rapid growth, one would therefore expect the greatest variability to occur in the infancy period. Further, if the separate abilities emerge at different times, there need be no qualitative agreement between infant behavior and later intelligence. Instead, the relationship may be something like that of a foundation to its superstructure, important to the existence and balance of the building but differing in composition.

There is a sharp division of opinion on this point in the field of infant research. Gesell, whose observational studies of infant behavior conducted at Yale University were the pioneer efforts in this country and
the source of the majority of currently used infant test items, offers an explanation based on the integrity of the nervous system. According to Gesell, the infant develops as a unitary action system which is manifested in patterns of behavior governed by deep-seated ontogenetic laws of developmental sequence from birth to maturity (35, p. 4). Further, mental growth is orderly and progressive, differing from individual to individual only in tempo or developmental rate. Gesell has followed the mental growth careers of large numbers of children and has published several reports of his research (34, 36). He believes that mental growth is predictable if (1) the end products of mental growth at the various age levels are satisfactorily isolated and (2) clinical rather than psychometric appraisals are made. In keeping with his thinking on the latter point, Gesell has made very little use of quantitative scoring and statistical methods, which has made it difficult to evaluate his work and compare it with other research in the field.

Opposed to the thinking of Gesell is that of Bayley, whose research has been described as the most definitive study of infant testing. The subjects of the Berkeley Growth Study, twenty-five years of age at the time of the latest report, have been followed from early infancy by means of periodic examinations. For these subjects Bayley has found no significant relationships between early scores and later IQ's, and in her recent report she interprets her results as follows:

These findings give little hope of ever being able to measure a stable and predictable intellectual factor in the very young.
I am inclined to think that the major reason for this failure rests in the nature of intelligence itself. I see no reason why we should continue to think of intelligence as an integrated (or simple) entity or capacity which grows throughout childhood by steady accretions. (12, p. 807)

As an alternate explanation, Bayley proposes that intellectual growth is comprised of a series of developing functions or groups of functions, each growing out of but not necessarily correlated with previously matured behavior patterns, the whole being a process of dynamic or changing organizations (5). However, she believes that a "g" factor does appear soon after the second or third year since correlations of tests at those ages become positive with tests at later ages. She points out further that although, according to her findings, intellectual development is a highly individualized process, the patterning is not random. "After the infancy period, there is a strong underlying consistency or constancy," and after five or six years of age "children can be reliably classified into broad categories of normal, defective and bright" (12, p. 808).

It is evident that the issues outlined above can be resolved only by more thorough data on the nature of early intellectual functioning and the problems of measurement that we now possess. Longitudinal studies, such as the Berkeley and Harvard studies, are of particular value, since the growth careers of the same individuals can be charted over a period of years and irregularities in development can be discerned and evaluated. Obviously, if Bayley's conclusions are proved correct, it will never be possible to use infant behavior as a basis for predicting later status. The most that even the best infant scale could ever offer would be an accurate appraisal of
development for the particular age level tested; unfortunately, such information is of little value for clinical purposes.

**Purpose and Scope of the Present Study**

Despite the pessimistic outlook for infant intelligence scales, the present study of the predictive value of the Cattell Infant Intelligence Scale has been undertaken because the scale is in current clinical use and some data on its actual clinical efficiency should be available. No validity study of the Cattell Scale has ever been published. Cattell warned of doubtful validity under twelve months—although her correlation coefficients of .10, .34, and .18 for three, six and nine months were higher than those obtained by Bayley for the California First Year Mental Scale. However, Escalona (25) indicated that the correlations she had found in clinical practice between Cattell estimates of intelligence obtained in early infancy and during the preschool years were better than previous studies on infant tests had reported. (Escalona admitted that her actual number of cases was small and she did not publish any figures to substantiate her statements.)

One of the probable reasons for the clinical usage of the Cattell Scale is its resemblance to the well established Stanford-Binet Scale. In fact, it was so constructed as to form a downward extension of Form L of the Stanford-Binet, and its findings are expressed in the familiar MA and IQ units. However, another possible reason for its acceptance has been the claim made by Cattell (21) that even for the age levels under twelve months extreme variations from the norm "in the direction of both feeblemindedness
and superiority" can be identified with the use of the scale. This is contrary to the findings of Bayley who discovered relatively large shifts in position made by individual children along the continuum from low to high intelligence. Moreover, Cattell herself in her earlier publication (20, p. 56) described some instances of growth irregularities similar to those reported by Bayley. Since her statement is at variance with the only factual evidence we have, it should be examined critically with additional data.

Accordingly, this research will concern itself with the extent of agreement between the scores obtained on two intelligence tests - the Cattell Infant Intelligence Scale and the Revised Stanford-Binet Scale, from L - when administered to a group of one hundred and ten children, in order to determine (a) the predictive value of the Cattell Scale for this group of children and (b) the success of the Cattell Scale in identifying the extremes of inferior and superior intelligence at an early age. The children to be studied were examined with the Cattell Scale at six months of age and were later retested with the Stanford-Binet Scale when they were in the three to six year age range.

As a related investigation, the extent of agreement between the intelligence of these children and certain features of their natural backgrounds and their environment will be examined. The children in this group were adopted in early infancy, and fairly complete records of the backgrounds of both the natural parents and the adoptive parents are available. This affords an opportunity to explore and to compare the correlations between the intelligence of the children and the education and
socio-economic status of the adoptive parents, and the intelligence of the children and the education of the natural parents. Educational status is frequently substituted when no other measure of parental ability is available.

The results of these investigations can then be compared with the findings of other studies. For example, Honzik (50), studying eight-year-old children reared with their own parents, found that mental test scores correlated .36 with mid-parent education and .11 with parent socio-economic status. Bayley (11), for her group of Berkeley children reared with their own parents, found a correlation of .55 with socio-economic ratings and education of the parents when the children were ten years old and correlations of .64, .65, and .60 with mid-parent education when the children were sixteen, seventeen and eighteen years old respectively. Bayley had found no agreement between these factors while the children were infants, but discovered that correlations increased steadily as the children matured; she therefore found it reasonable to assume a hereditary core of parent-child similarities in mental characteristics, even though such similarities may not be evidenced during the first year or so of the child's life. Bayley agreed that the effects of environment were difficult to determine; however, she believed that her hypothesis was supported by a study conducted by Skodak and Skeels (62) on a group of adopted children. Skodak and Skeels found a correlation of .44 between the true mother's IQ and the child's score when the children studied were thirteen and a half years of age, whereas correlations with educational status of the adoptive parents
were .00 to .06. It should therefore prove interesting and profitable to examine these relationships for the one hundred and ten adopted children in this study and compare them with the earlier findings.
CHAPTER II

REVIEW OF THE LITERATURE - I

The period of infancy is usually considered to extend up to eighteen months, or by some authorities, up to twenty-four months. Beyond that age level is regarded as the preschool period. There has, of course, been overlapping of these age levels by intelligence scales. The Cattell Scale extends well up into the preschool period. Two of the well known preschool tests, the Merrill-Palmer Scale (64) and the Minnesota Pre-School Scale (45), present items for as low as eighteen months. But since the present research is concerned with an age level within the first year of life, only those studies and tests which have specific reference to that period will be considered in the following resume of the literature.

There are several points of difference between the development and behavior of the young infant and the preschool child which justify a separate consideration of their testing problems. The young infant must be tested either while lying in a crib or placed in a well supported sitting position. The normal preschool child can walk and move independently, and hence can assume a position at a table with the examiner. The infant does not engage in verbal communication. The preschool child uses speech as a tool and can participate in activities demanding verbal responses. The infant
reacts primarily to the testing equipment which must be especially designed to elicit the desired behavior. Only secondarily does he usually relate to the examiner, who thus acts in the testing situation as an observer. In contrast, the preschool child is capable of entering into a direct interpersonal relationship with the examiner and the quality of rapport assumes a much greater importance.

But most particularly for the purposes of test construction, the young infant's developmental rate is much more rapid and the nature of the test items themselves is quite different. As will be pointed out in more detail, the test items in the first six months are largely sensorimotor in character, but this component has been found to decrease gradually in importance after that age level. In contrast, the test activities designed for the preschool child are more clearly of a problem solving nature.

Historical Development of Infant Intelligence Tests

A survey of the history of infant intelligence test development reveals an early interest in the field. Over the years a large number of test items for evaluating different aspects of infant behavior have been devised and presented either as supplements at the lower end of tests for more advanced age levels or as groups constituting separate infant scaled.

As early as 1904, when he published his scale for measuring intelligence in the school child, Binet (14) presented four or five items which were suitable for evaluating reactions of infants under one year. However, he intended these items to be used in differentiating mental
defectives who were too retarded to perform on the regular scale; in consequence, Binet did not assign any exact age placement to these tests in terms of their applicability to infant behavior, but it is interesting to note that all of these items - namely, reaction to light and sound, prehension after tactile excitation and after visual perception, and imitation of movements and execution of simple orders in response to word or gesture - appear in present day infant scales in some form.

In his 1922 revision of the Binet scale, Kuhlmann (53) extended the test at the lower end from three years down to three months, presenting five items for scoring at each of the age levels of three, six, nine, twelve, eighteen and twenty-four months. In the 1939 revision of the Kuhlmann-Binet Scale (54), fourteen items are presented for the first year, arranged in the order of their difficulty. Scores are expressed in terms of growth curve values, made of equal units of measurement.

Over the years since 1919, when he began his investigations in the Yale Clinic of Child Development, Gesell has developed many of the standard items used in the current infant tests. The authors of the recent infant scales - Bayley, Cattell, Gilliland and Griffiths - have all indicated that they drew heavily on Gesell items. In 1925, Gesell published his first schedule of developmental norms (33), extending from birth up to five years. Over five hundred children were examined at four, six, twelve, eighteen, twenty-four, thirty-six, forty-eight and sixty months of age, and separate schedules, involving a total of one hundred and fifty normative items, were then arranged for each level of development. The chief objection
raised against this first scale was that no precise ratings of a child's developmental level could be obtained. Gesell described different grades of success for each item, so that a general idea of a child's developmental level could be established by examining the different degrees of success attained on the total scale but he did not indicate any exact at-age values for his items.

In 1947, Gesell published the most recent description of his revised and supplemented version - the Gesell Developmental Schedules (36). This revision presented over two hundred items representing behavior characteristics for the age levels between four weeks and forty-two months. As in the early scale, items are arranged to check behavior in four separate areas of development - language, motor, adaptive and personal-social. For example, at the six months (twenty-four weeks) level there are six motor items, six adaptive items, three language items and four personal-social items. The norms in each area were derived from observation of infants and young children, and were placed with objective reference to the age at which they elicited behavior patterns are normally expected to appear. It is important to remember that Gesell's schedules comprise a normative scale, rather than an intelligence scale in the strict sense. His method of tapping behavior at its expected level of appearance differs from the empirical method of item selection - that of assigning item selection on the basis of percentage of successes by a given age group. The Gesell scale does not lend itself to an IQ rating, although the infant's total score on the four areas of behavior can be divided by the chronological age to give a
developmental quotient, which indicates the proportion of normal development present at the time of the examination. There has been one study of the six months items of the Gesell schedules which will be reported later in this chapter.

Another normative scale was published in 1928 by Hetzer and Wolf (47). On the basis of twenty-four hour observations of infants in the laboratories of the Psychological Institute of Vienna, these authors devised monthly norms through eleven months. In 1930 and in 1935, Buhler (16, 17) published other versions of these tests, in which the scale was extended up to two years and revised along the general lines of the Binet scale. A series of ten items was presented for each month level up to twelve months, items being selected to evaluate four general lines of development. The tests were drawn up after ten preliminary trials were made for each month and then given to thirty children at each month level. The final score is expressed in terms of a developmental age, obtained by adding credits to a basal score. Certain of the Buhler tests were used by Cattell and Gilliland in the construction of their scales. However, the Buhler scale has been criticized for clinical usage because it was standardized on institutional babies and because it contains many situations which are frustrating or frightening to the child. Further, using Buhler's own classification of behaviors, Cattell found less than half of the items to relate mainly to mental development (20, p. 22).

In 1928, Linfert and Hierholzer (56), graduate students at the Catholic University of America, published their point scale for the first
twelve months, based mainly on the Gesell tests. It was claimed to be the first standardized scale with age norms published for that period of life. The scale was divided into two series, and included tests for one, two, four, six, nine and twelve months. Tables indicating percentages of successes in the various tests were presented for calculating age norms, the final results to be expressed in I-H Quotient. According to the test authors, the total point scores showed a linear increase with age. Only one study, indicating poor predictive value for this scale, has been published; it will be described later in this chapter. The extent of the clinical usage of the Linfert-Hierholzer Scale is not known.

As an outgrowth of her longitudinal study, which involved over 1142 tests on sixty-one children during the first three years, Bayley published the California First Year Mental Scale (4), covering the age range on one month to eighteen months. Using a large number of Gesell items, Bayley included tests of adaptability or learning, and tests of sensory acuity and fine motor coordination. She placed her items on a continuous scale in order of difficulty by the Thurstone Method of Absolute Scaling, and she indicated exact at-age values for each item. Results are expressed in terms of a cumulative point score based on the number of the child's successes. Bayley's scale is considered to be fairly well standardized and to include a sufficient number and variety of items. However, the scale has been found to have poor predictive value.

On the basis of examinations conducted on "several hundred" children at the Iowa Child Welfare Research Station, Fillmore (27)
published the Iowa Tests for Young Children in 1936. These tests, including forty-nine items, covered the age range of four months to two years, but they were never adopted extensively for infant testing in clinical practice. One possible explanation is the very small number of items presented for the first twelve months - only ten items were given for the period from five to nine months, two of them for the six-month level. Originally, the test items were arranged according to the percentage of successeses, and mental age credits were found for each item by dividing the age range covered by the number of items in the particular age range. This system was then discarded in favor of a point scale with the items arranged in the order of difficulty, according to Thurstone's Method of Absolute Scaling. According to the author, the Iowa tests measure some ability which increases with age. Fillmore found that her tests failed to correlate highly with later Stanford-Binet IQ's, however. No studies, aside from the author's original presentation, have been reported on this scale.

The Cattell Infant Intelligence Scale, published in 1940, will be described in detail in Chapter IV. The most recently developed infant scale in this country has been the Northwestern Infant Intelligence Scale, originally described in 1943 by Gilliland and Shotwell (41), and later presented in its revised form by Gilliland (37). The authors began their work on the scale at the request of a child care institution interested in determining the suitability for adoption of the very young infant. A large number of items from existing scales, mainly those of Gesell, Cattell and
Buhler, as well as some new items, were administered to approximately five hundred children, mostly institution babies. On the basis of tests for 276 babies whose records were complete, a final revision was made which consists of forty items arranged in two overlapping series (38, 39), covering the age groups from four to fourteen weeks and thirteen to thirty-six weeks. An IQ can be computed for any age, the raw score of the test being the number of items passed. In placing his items, Gilliland used the method of increase in percentage of passes with chronological age. If seventy-five percent of the infants at a given age level could pass an item, it was considered to be correctly placed. Gilliland claimed evidence of high validity for his scale on the basis of later Stanford-Binet results, but he did not publish data to amplify this statement.

In 1954, a British psychologist, Ruth Griffiths, published the Griffiths Mental Development Scale (46), covering the age range from two to twenty-four months. Based on the author's research on over 1000 infants between the ages of two weeks and two years, and standardized on 604 "representative" London infants, this scale is the most elaborate and detailed infant test to date. The author indicated in the manual her belief that all aspects of mental development are represented during the first year of life. Five separate scales are presented - Locomotor, Personal-Social, Hearing and Speech, Eye and Hand, and Performance; each scale contains fifty-two separate items arranged on month levels in order of difficulty (on the basis of the percentage of babies passing them month by month), a total of 260 items in all. The scales can be scored separately and
separate quotients obtained. The author points out the usefulness of this procedure for the differential diagnosis of handicapped children. As a total score on the five scales, a General Quotient (GQ) is obtained by totaling the number of items passed in all and computing a mental age. The manual includes detailed instructions for administering and scoring the items.

The Nature of Infant Test Items

Item selection poses a particular difficulty for infant test construction because of the lack of suitable external criteria. In selecting infant test items and placing them on the scale, the chief criterion has been progressive increase in the percentage of successes at successive age levels. Additional criteria used by most test authors include (a) appropriateness, in so far as can be determined, as indicators of intelligence, (b) freedom from the influences of home training, and other social and cultural factors, (c) internal item consistency and correlation with the total scale, and (d) convenience in administration and scoring.

Because of the use of the criterion of progression with age, infant scales have always included many motor items. Motor behavior is readily observable and, as Gesell's studies have revealed, there is an orderly progression from gross to fine activity in relation to age maturation. However, motor behavior has not been found to correlate well with intelligence at later ages. Bayley (6) found some similarity of function between mental and motor scores during the first fifteen months – correlations were in the
vicinity of .5 - but the relationship dropped markedly after that age level. Nelson and Richards (59), in a study of the six months items of the Gesell Developmental Schedules, found that motor items, including posturo-locomotor and manipulatory activities, did not correlate highly with mental development at two and three years of age, although they had correlated fairly well with the total test performance at six months. Correlations for the motor-manipulative items dropped from .65 for the total six months test score to .30 with Stanford-Binet scores at three years. Similarly, correlations for posturo-motor items dropped from .55 to .21.

Bayley sought to overcome the influences of motor items by arranging tests of this nature into a separate motor scale (7). Cattell also sought to eliminate items which appeared to be related chiefly to motor ability. Gesell separated infant behavior into several categories, among them motor development, so that separate evaluations could be made. Similarly, Buhler separated her test items and designated certain subtests as involving "bodily control". However, the difficulty in making a clear-cut distinction between motor and other types of behavior during the first year of life, and hence of excluding entirely the influence of motor development, has been pointed out by Gesell in the following comment:

Motor and adaptive behavior are intimately combined in early life, because under the pressure of growth, a normal infant feels impelled to put each newly attained motor ability to repetitive use, and to exercise it with experimental variations. For example, an eight week old infant cannot reach for a rattle but will briefly retain a rattle placed in his hand - a slight bit of adaptive behavior which is not altogether pure reflex. At twelve weeks he will hold the rattle actively and even glance in its direction. At sixteen weeks he regards it immediately and
intently. He also deploys his eyes in a roving manner to "contact" his surroundings. In the next two months he reaches out to contact, to grasp and to hold. Thus by subtle growth stages which begin very early the infant's visual and manual behavior take on voluntary and adaptive characteristics (35, p. 58).

An inspection of the existing infant scales reveals a similarity in the kinds of items included, especially for the age levels under twelve months. In the first place, the items are limited by the small range of behavior that can be elicited from the infant. Further, most of the recent test authors have borrowed heavily from Gesell's normative items, although they have frequently placed them at other than the originally designated age levels in accord with their own findings. The following items from Bayley's scale, together with the exact at-age placement for each item, are presented here as typical of the activities expected in testing the six-month-old infant:

<table>
<thead>
<tr>
<th>Item</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaches persistently</td>
<td>6.05</td>
</tr>
<tr>
<td>Turns after spoon</td>
<td>6.1</td>
</tr>
<tr>
<td>Mirror image, approach</td>
<td>6.1</td>
</tr>
<tr>
<td>Picks cube deftly</td>
<td>6.1</td>
</tr>
<tr>
<td>Several syllables</td>
<td>6.3</td>
</tr>
<tr>
<td>Bangs in play</td>
<td>6.4</td>
</tr>
<tr>
<td>Sustained attention to ring</td>
<td>6.4</td>
</tr>
<tr>
<td>Unilateral reach</td>
<td>6.15</td>
</tr>
<tr>
<td>Vocalizes satisfaction</td>
<td>6.5</td>
</tr>
<tr>
<td>Lifts cup by the handle</td>
<td>6.6</td>
</tr>
<tr>
<td>Exploits string play</td>
<td>6.7</td>
</tr>
<tr>
<td>Rotates wrist</td>
<td>6.7</td>
</tr>
<tr>
<td>Scoops pellet</td>
<td>6.8</td>
</tr>
</tbody>
</table>

Most of the test authors have avoided any logical explanation of the nature of the behavior underlying their test items. An exception to this is Gesell who utilized the four categories of behavior - language, personal-social, adaptive and motor behaviors. Buhler also attempted
some classification, by labeling the individual items in accord with the behavior they were intended to evaluate. Her system included four general lines of development—bodily control, mental ability, manipulation of objects and social development.

Bayley attempted several classifications of her test items but found them to be unsatisfactory. "In many cases an adequate response to a test situation requires abilities of more than one kind, so that items may be equally well assigned to two or more classes." (5, p. 63)

As has already been pointed out, Gesell found this to be true but he attempted to select the type of behavior chiefly involved in order to provide a means of evaluating progress in certain general areas of growth.

An added difficulty noted by Bayley in classifying tests into sub-groups was that no two areas of infant development show parallel development. She made a broad classification of test items into two categories—sensorimotor and adaptive behavior (the first group involving sensory acuity and fine motor adjustments, the second demanding learning and problem solving)—and studied their influences on test performances during the first year. She discovered that development during the first six or eight months was largely sensorimotor in character, whereas the more truly adaptive behavior is measured by tests only after that period. (5, p. 63).

Bayley's findings are probably applicable to all of the infant intelligence scales. Watson (68, p. 334), for example, has indicated in a general way the chief abilities demanded by the Cattell sub-tests during the first twelve months. According to his designation, the tasks are
largely perceptive in nature at the earliest levels – for example, attending to sounds or visually following a moving ring. Beginning at about five months, there is a gradual change to more manipulatory (adaptive) tasks. The first verbal type of test appears at nine months, involving adjustment to words – that is, performing an activity in response to a spoken request. The first speaking vocabulary item appears at eleven months. From that point on, more verbal tests are utilized, although manipulatory tests still predominate.

There have been two studies of infant tests in which factorial procedures have been utilized in examining the nature of infant test items. Working with the items of the six month level of the Gesell Schedules, Nelson and Richards (60) found that three factors seemed to account for most of the variance in the behavior of the six-month-old infant, as tested by the seventeen items in the middle range of difficulty – 25 percent to 75 percent passing. These factors were designated as (a) testability or "halo effect" (b) alertness and (c) "motor ability." In the second study, Hofstaetter (48), working with Bayley's data from the Berkely Growth Study, found one factor predominating during the first two years of life, which he named "sensory motor alertness." Two additional factors were analyzed from the 18 year Berkeley data – "persistence," predominating from two to four years, and "manipulation of symbols," accounting for most of the variance after four years. Bayley considered these successively appearing factors to be representative of the "complete break between the kinds of function measured in infants and in school age children." (12, p. 808)
Reliability of Infant Scales

The problems of test reliability have some added significance during the infancy period because of the developmental factors affecting test scores which can operate even over brief periods of time. Of particular importance among these factors are growth tempo and consistency. Unquestionably, the older the age group the more reliable the measuring instrument is expected to be because rapidity of growth and irregularities of growth patterns decrease with age. But infancy is a time of rapid behavioral development. Gesell describes the period of infancy as one of changing and fugitive behavior, exceeding all other age intervals in the wealth of phenomena displayed. Referring to the difficulties involved in selecting suitable norms for infant behavior, he states that "even with simplification we must recognize in the first year of life at least three developmental intervals and devote attention to the stages of maturity presented at four months, six months and nine months." (33, p. 4)

Thompson (66), a co-worker of Gesell in the Yale Clinic of Child Development, concluded as a result of daily observations of infant behavior under well controlled conditions that behavior growth proceeds fully as rapidly as physical growth. Growth increments may occur in different functions on successive days or they may occur in more than one function on the same day. Further, the behavior growth increment may manifest itself in one of the four following ways: (a) the greater frequency of one item of behavior; (b) the improved performance of an activity; (c) the appearance of a new activity, and (d) the integration of previous activities.
A summary presented by Gesell of the progress of prehensory development during the first year of life—in which growth proceeds from the tight reflex grasp in the neonate to the precise finger-thumb opposition at twelve months—provides a clear illustration of the span of development in one type of behavior.

Clearly, such rapid growth increments would have some bearing on infant test reliability. Both Bayley and Cattell found their lowest test-retest reliability figures for their scales in the one to three month age range. In his review of Bayley's study, Cronbach concluded that her neonate tests were unreliable because at a level where a new activity is just emerging "the pattern is diffuse, varied and inconsistent from time to time; measurement of such functions is therefore unstable" (22, p. 169). Cattell likewise proposed that some of the variations noted in test-retest scores resulted from changes in the tempo of development rather than from inadequacies in the tests themselves (20, p. 60).

Cronbach had pointed out further that a scale showing unreliability at certain levels may have satisfactory overall reliability. The Cattell and California Scales, as well as the more recent Northwestern Scale, have fared surprisingly well with respect to split-half reliability coefficients found for most age levels under eighteen months. For her scale, Bayley reported reliability figures ranging from .75 to .95 for the age intervals from four to twelve months. The median value was .86. Cattell found reliability coefficients of .88, .86, .89, and .90 for the six, nine, twelve and eighteen month age levels respectively. In general, as
expected, reliability figures tended to increase at higher age levels. Contrary to the findings of Bayley and Cattell, Gilliland obtained fairly high reliability figures for the Northwestern Scale, even for early age levels. He reported split-half reliabilities ranging from .79 to .94 for each week between four and twelve weeks of age, with an overall reliability of .84 for the total group. These figures were obtained with the Northwestern Test A, administered to 276 infants in the maternity home. When Test B was administered to a group of 200 infants, an overall reliability of .80 was found. Anastasi (1, p. 288) emphasizes the large number of items included as a partial explanation of the higher reliability coefficients obtained for the Northwestern Tests. However, Anastasi also suggests that items selected to sample a wide variety of functions may produce such heterogeneous test content that comparable halves for the computation of split-half reliability would not be obtained. If closely comparable forms were available for the infant scales, the reliability coefficient might be even higher than obtained. But on another aspect of this same issue, Anderson (3) points out that while high reliability may be obtained by the inclusion of a great many items, the items in themselves may not necessarily be significant indicators of later intelligence ratings. This would explain why infant scales may have good overall reliability but poor predictive value.

Gilliland (40) indicated that additional factors operating to produce low reliability figures at the infant level are (a) temporary physical factors, (b) late maturation observed in some infants, and (c)
poor rapport during the examination.

Validity of Infant Scales

The validation of Infant scales has been a problem because the variety of validating criteria - such as independent achievement, ratings of brightness, and scholastic records - which can be used for work with tests for older age groups are not available with infants. Therefore, the only validating criteria ordinarily used by infant test constructors have been (a) increase in percentage of passes with chronological age, used in the original selection of items, and (b) the predictive value of the total test. Gilliland was the only test author to vary the usual procedure for infant tests by making use of an independent validity criterion; he compared the performances of defective infants (in recognizable clinical groups, such as mongolism) with performances of infants not regarded as defective.

For the Cattell Scale, a total of sixty percent of infants in a particular age group was considered suitable for placing an item, provided that a lesser number of infants at the adjacent lower month level and a higher number at the adjacent higher month level were able to pass it. Gilliland selected seventy-five percent. Age differentiation was of course the basis for such point scales as the California and Iowa tests. The test authors on the whole have been successful in incorporating items which show satisfactory age difference from month to month, even when motor items which more readily satisfy this criterion have been excluded. However, Anderson (2, p. 376) pointed out the following drawback in the use of such a criterion:
"since (infant) development is a timed series of reactions or sequences, there are for many functions periods below which only a small portion of the function can be measured and above which a progressively larger portion can be measured. Hence the possibilities of prediction are limited and progression with age is not an infallible indicator of the value of a measurement."

A small number of studies, including the descriptions of standardization work done by the test authors, have reported the results of correlating infant test scores with scores at later ages on more established tests, usually the Stanford-Binet. Standardization populations and study samples have been very small; typically, only one report is available for a particular infant scale and for most of the tests - such as the California Scale, the Iowa Tests, the Cattell Scale and the Northwestern Scale - no validity work beyond that of the authors has been published. There has been some small variation in the correlations reported, but the findings in general have agreed in establishing the conclusion that infant tests in their present form have poor predictive value. These findings can be considered in summary form below.

Cunningham (23) examined the results of the Kuhlmann-Binet given to a group of 27 children at twelve months in the 1916 Stanford-Binet at a median age of eight years. Retest age range was seven years to eight years, seven months. She found the correlations to be .55.

Hubbard (51) studied the Buhler Tests and their predictive value for a group of infants followed in a well baby clinic of a hospital. The
children were tested twice with the Buhler scale; all were of varying ages under twenty months at the time of the first test. First test ratings for twenty-five children correlated .37 with later ratings on the Merrill-Palmer Preschool Scale and second test ratings for fifteen children agreed with Merrill-Palmer results to the extent of .70. (This study has little value for comparative purposes because Hubbard did not present her data clearly in terms of ages at the time of testing and retesting, nor did she indicate the length of intervals between test and retest. Since she was apparently comparing performances on the upper levels of an infant scale with performances on a preschool scale, it appears likely that her results should be regarded as reliability figures, rather than validity coefficients.)

Furfey and Muhlenbein (31) retested 71 of the 131 children used in the standardization of the Linfert-Hierholzer Scale. Mean age at the time of retesting with the 1916 Stanford-Binet was four years, eight months; twenty-seven of the children had been tested with the L-H Scale at six months, twenty-six at nine months and twenty-eight at twelve months. They found negative correlations of -.11, -.34, and -.20 for six, nine and twelve month scores respectively.

Fillmore (27) reported a correlation of .32 between performance on the Iowa Tests at six months and the earliest IQ obtained (at two to three years). She also correlated certain of her six month test items with this criterion, obtaining coefficient ranging from .02 to .43. The size of her retest group was not specified.

Anderson (3) worked with a group of infants followed in a
longitudinal health and development study at Western Reserve University. Ninety-one children were tested at regular intervals between three and twenty-four months with a battery of Gesell and Buhler tests, and retested at five years with the 1916 Stanford-Binet. Average IQ at five years was 116.12. She found little or no correlation between scores at three, six, nine and twelve months and later IQ standings (correlations centered around zero). By selecting and correlating certain items with later scores, she found that correlations increased to .31, .41, .20, and .22 at three, six, nine and twelve months, with the lower correlations at the two later ages ascribed by the author to the fact that fewer significant items could be found for those ages. Anderson considered the most marked characteristic of the significant items to be "alertness to external environmental stimulation . . . especially those involving bodily orientation." She also suggested that early language development (at eighteen to twenty-four months) appears to be more closely related to later intelligence than any other grouping of tests and that it may be impossible to predict intelligence before the age when language development can be measured. Anderson further offered her findings as proof that an infant scale constructed only with regard to the inclusion of items showing satisfactory increase of successes in relation to age does not necessarily have value in predicting intelligence.

Nelson and Richards (59) examined the Gesell performances at six months and 1916 Stanford-Binet Mental ages at thirty-six months for a group of thirty-one children and found agreement to the extent of .47. When the
Biserial technique of correlation was used to determine Binet relationships to individual Gesell items at six months, the correlation coefficients ranged from .00 to .58, centering about .20 to .25. The highest coefficient (.58) was found for an item which correlated very low with the total test at six months. In general, the "motor" items tended to have lower coefficients than the "awareness" or distance-perception items. Results of multiple correlation work done by the authors suggested that predictive efficiency of certain single items was better than that of the total test, and was raised by combining items. When five items, passed by 25 to 75 percent of the group and correlating relatively highly (about .54) with the Binet mental age of thirty-six months, were combined, a correlation of .80 was obtained. "If we regard the abilities tested by the Gesell series at six months as basic to those measured by the Binet at three years, we may conclude that as much as sixty percent of the variance in mental age at the later age level is sampled at six months." (59, p. 322)

On Bayley's California First Year Mental Scale, the correlations found for her standardization group of 61 children, subjects of the Berkeley Growth Study, were -.09, .10, .22 and .45 for four age groupings under twelve months and a retest with the California Preschool Scale at twenty-seven, thirty and thirty-six months. Bayley used an average of the sigma scores on three consecutive tests as a basis for comparisons over wider age intervals in order to rule out chance variations on a single test. Bayley's work has been the only research to employ standard scores in correlating the results of infant tests and tests at later ages in order to eliminate
changes in test scores which were due to difference in variability at the various age levels. Thus a child's relative position in the group is used as the basis for comparison, rather than his IQ or some equivalent score.

In a recent report, Bayley published the correlations between the earlier tests and the Wechsler-Bellevue, administered when the children in the group were eighteen years of age. The correlation for six and twelve months and the eighteen year scores were -.12 and .25. Bayley has found that correlation coefficients for successive tests have increased steadily with age for this group; the correlation between seventeen and eighteen year scores for thirty-six subjects was .90 (12). These children have proved to be a superior group as far as general intelligence is concerned; their mean Wechsler-Bellevue IQ at sixteen years was 117, the mean Stanford-Binet, Form M, IQ at seventeen years was 129.

No complete item by item analysis has been done with the Berkeley infant measurements. However, in a preliminary item study, the infant test records of the six brightest and the six slowest seventeen year olds were examined, and thirty-one discriminative items were selected. The cumulative point scores composed of those thirty-one items still did not reliably differentiate the bright from the dull children during the first year. When scores for the total sample (45 cases) on this 31-item scale were computed for the three ages of six, nine, and twelve months and compared with the mean of the sigma scores at ages sixteen, seventeen and eighteen years, the correlations were .09 at six months, .32 at nine months, and .30 at twelve months. (12)
Cattell (20) reported the following validity coefficients for her standardization group, obtained with Stanford-Binet scores at thirty-six months: .10, .34, .18, .56 and .67 for the ages of three, six, nine, twelve, and eighteen months respectively. The number of children tested at each age ranged from 42 to 57.

As stated earlier, no validity data on the Cattell Scale has been published, although two reports have claimed evidence of satisfactory predictive value. Escalona (25) stated that the correlations between Cattell estimates of intelligence obtained in early infancy and during the preschool years which she had found in clinical practice were better than previous studies on infant tests had reported. Escalona admitted that her actual number of cases was small and she has not published any actual figures in support of her statement.

A recent report by MacRae (58) concluded that the predictive value of infant scales had been underestimated by investigators who have attempted to interpret infant test ratings in terms of specific IQ's. He found the Cattell Scale and the Gesell Schedules to have "definite value in the prediction of later mental ability" when categorical ratings of infant performances were correlated with similar categorical ratings for the retest data. The categorical ratings used by MacRae were 1. Superior (IQ 120 and above), 2. Above average (110-119), 3. Average (90-109), 4. Below average to borderline (70 to 89), and 5. Mental defective (69 or below). His study group was composed of 102 children who had received infant tests under thirty-five months of age. Ninety had received the Gesell Schedules and twelve
children were given the Cattell Scale. They were retested with either the Wechsler Intelligence Scale for Children or the Stanford-Binet at a median age of nine years, two months. MacRae did not treat the two infant tests separately in handling his data. When categorical infant ratings for the entire group of 102 were compared, a correlation of .65 was obtained. When separate infant age groupings were used, correlations of .56, .55 and .82 were obtained for 0–11 months, 12–23 months and 24–35 months. These three groups were composed of 40, 41 and 21 children, respectively. MacRae found further that the infant ratings compared very closely with retest IQ's in 65 out of 102 cases. In 97 out of 102 cases the deviations between ratings was not over one category. In no case was the deviation over two categories.

MacRae's study is difficult to evaluate because of the broad age groupings used in handling the data. It appears that the correlation coefficient for the total group was probably raised by the inclusion of "infant" tests administered at or near the age of thirty-five months, since the coefficient for the separate age range of twenty-four–thirty-five months was .82, as compared with lower coefficients for the younger age groups. The age range of twenty-four to thirty-five months is usually regarded as falling within the preschool period, and researchers have almost always found that validity coefficients derived from retest scores are higher for the preschool ages than for the infancy period.

Table I contains a summary of the studies reporting correlations between six-months performances on infant scales and tests at a later age.
### Table I

Retest Correlations from Published Validity Studies of
Six-month-old Infant Test Performances

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Test</th>
<th>N</th>
<th>Criterion</th>
<th>Retest Age</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Furfey and Muhlenbein (31)</td>
<td>L-H Scale</td>
<td>42</td>
<td>S-B (1916)</td>
<td>1 yr. 2 mos.</td>
<td>-.20</td>
</tr>
<tr>
<td>Fillmore (27)</td>
<td>Iowa Tests</td>
<td>91</td>
<td>S-B (1916)</td>
<td>2 or 3 yrs.</td>
<td>.32</td>
</tr>
<tr>
<td>Anderson (3)</td>
<td>Buhler &amp; Gesell Items</td>
<td>31</td>
<td>S-B (1916)</td>
<td>5 yrs.</td>
<td>.08</td>
</tr>
<tr>
<td>Nelson and Richards (59)</td>
<td>Gesell</td>
<td>31</td>
<td>S-B (1916)</td>
<td>36 mos. (MA)</td>
<td>.47</td>
</tr>
<tr>
<td>Bayley (5)</td>
<td>California Scale</td>
<td>61</td>
<td>Cal. Preschool Scale</td>
<td>27, 30 &amp; 36 mos.</td>
<td>.10</td>
</tr>
<tr>
<td>Cattell (20)</td>
<td>Cattell Scale</td>
<td>42</td>
<td>S-B (1937)</td>
<td>36 mos.</td>
<td>.34</td>
</tr>
</tbody>
</table>

The following explanations have been offered for the low predictive value of infant intelligence tests:

(a) The influence of motivational and temporary physical factors on individual infant test performances.

(b) The failure of existing infant scales to sample infant behavior adequately or to isolate behavior which is intellectual in character.

(c) The comparatively greater susceptibility of the infant to environmental
influences, leading to the modification of measured abilities during the infancy period. (Thorndike, 67)

(d) The influence of home environment which do not greatly affect the infant during the first year, but which exert a differential effect on intelligence in the older child. (Furfey, 30)

(e) The possibility of a common matrix of abilities at infant and later age levels which can be sampled at each point, but which never forms all of the abilities covered by the test, thereby limiting the possibilities of prediction. (Nelson and Richards, 59)

(f) "The simple addition of infant scores without regard for the predictive value of the individual test items may result in the favorable effects of the significant items being overbalanced by the unfavorable effects of the nonsignificant items. These last may not only act as dead wood but may decrease the predictive value of a scale. (Anderson, 3)

(g) The replacement of the abilities manifested in the infant's behavior repertoire by different sets of abilities at more mature age levels. (Bayley, 5, 12)

Summary:

Since 1904, when Binet published a few items suitable for evaluating infant behavior, there has been a large number of test items presented either as supplements at the lower end of tests for more advanced age groups or as separate infant scales. Kuhlmann extended the Binet Scale from three years down to three months in his revision and offered items for several
levels during the first twelve months. The most widely known infant tests are those of Gesell, published originally in 1925 and in revised form in 1947. The Gesell Developmental Schedules comprise a normative scale, rather than an intelligence test in the strict sense, but a developmental quotient evaluating the child's developmental level in four different areas of growth can be derived.

Another normative scale was published in Vienna in 1928 by Hetzer and Wolf, and revised along the lines of the Binet Scale in 1930 by Buhler. Some of the Buhler items have been incorporated in more recent infant tests but her scale has never been widely used in this country.

Other infant scales have included the Linfert-Hierholzer Scale (1928), the California First Year Mental Scale (1933), the Iowa Tests for Young Children (1936), the Cattell Infant Intelligence Scale (1940), the Northwestern Infant Intelligence Scale (1943) and the Griffiths Mental Development Scale (1954).

One of the important characteristics of the infancy period which bears upon the general problem of evaluating infant mental development is the transient nature of infant behavior. Researchers at the Yale Clinic of Child Development have concluded as a result of their observations that behavior growth proceeds fully as rapidly as physical growth and that even with simplification, at least three developmental intervals - four, six and nine months - must be recognized in evaluating rate of maturity during the first year. Normative tests of behavior can be utilized since behavior growth proceeds in accord with laws of orderly developmental sequence,
according to Gesell.

The fluctuations occurring in infant growth are considered to have important implications for intelligence test construction. Most of the infant scales have good overall reliability, but lower reliability at certain age levels and individual variations in IQ curves have been explained in terms of the diffuse nature of new behavior and changes in the tempo of development. Anderson has suggested that the changing nature of infant growth lowers the predictive value of infant scales since there necessarily occurs an irregularity in the amount of a given function that can be measured at different age levels.

Because of the validity criterion of progression with age, infant scales have always included many motor items. However, motor development has not been found to correlate well with intelligence at later ages. Most test authors have tried to reduce the influence of motor behavior on test performance, but it is difficult to exclude entirely because of the close relation between motor and other types of behavior during the first year of life.

There is a similarity in the kinds of items included in infant intelligence scales because of the limited range of infant behavior and because test authors have all borrowed heavily from the normative items devised by Gesell. Most authors have avoided presenting any logical explanation of their items. Gesell separated his tests into four general areas of behavior - motor, language, adaptive and personal-social - but he emphasized the impossibility of drawing hard and fast lines of distinction.
among these areas. A study by Bayley presented additional evidence that a response to a given test situation required activities of more than one kind, making classification difficult. She found that sensory and motor items predominated during the first six months of life, while adaptive items gradually gained prominence after that period. This general classification is probably applicable to most of the infant intelligence scales.

The only validating criteria ordinarily used by infant test constructors have been (a) increase in percentages of passes with chronological age, used in the original selection of items, and (b) the predictive value of the total test. The test authors on the whole have been successful in incorporating items which show satisfactory age differentiation from month to month, even when motor items which more readily satisfy this criterion have been excluded. However, infant scales have been found to have poor predictive value. The highest validity coefficient reported in the literature for an infant scale was .47, obtained by Nelson and Richards for the unreviewed Gesell tests with a very small group of children.

Various explanations have been offered for the low predictive value of infant tests, ranging from the influence of temporary physical and motivational factors to the changing composition of intelligence at different age levels. Validity coefficients do tend to increase with age, even within the infancy period, and various authors have concluded from their limited findings that the use of the multiple correlation technique with available infant test items offers some possibilities for improving item groups below twelve months.
The present chapter will be devoted to a discussion of some of the pertinent literature on the variability of mental growth, and environmental correlates of intelligence.

The Variability of Mental Growth

As stated in the introduction, a study which involves a comparison of intelligence test scores on the same individuals at different ages must take into account the possibility of variations in scores occurring either (a) as a function of inadequacies of the test used, (b) as a reflection of changes occurring in the tempo and/or the nature of the underlying intelligence, or (c) as a combination of these two factors. Bayley has handled this issue in somewhat greater detail in her introductory comments to a study of variability and inconsistencies found in the Berkeley mental growth curves (10). She lists three possible conditions which militate against a child's maintaining a "constant IQ" throughout his growth:

(1) Differences in standardization from one test to another, with differences in relative difficulty, cause spurious changes in IQ's.
(2) Age changes in variability of the tested mental functions, since if relative intellectual status is expressed either by scaled point scores or by the ratio MA/CA, the scores of exceptional children are necessarily brought closer to the average during the periods when variability is reduced.

(3) Changes in mental organization, so that different functions are being measured on different segments of the mental growth span (10, p. 180).

The findings of both the Harvard and the Berkeley Growth Studies suggested that differences among the various tests used were in part responsible for changes in scores when the IQ was employed as the index of growth. In the Harvard study, both group and individual tests were used; it was found that the group mental tests yielded higher IQ's than the Stanford-Binet test, and that, with respect to all the tests used in the study, each of the different mental tests was characterized by its own single and peculiar differences with respect to the problems of practice effect and its relation to individual test problems (24, p. 342). In the Berkeley study, considerable differences in mean IQ's for the different tests used reflected this source of spurious changes in IQ. Means for the Berkeley group ranged from 99.1, obtained at six months with the California First Year Scale, through 117.8 with the Stanford-Binet at five years, to 129.1 with the Wechsler-Bellevue at 17 years. "It is obvious from their shifts, which range from 116 to 132 on the standard tests given after five years of age that the norms are not of equivalent difficulty at all
ages. The Stanford-Binet IQ's average considerably higher than either the Terman-McNemar or the Wechsler." (10, p. 171) Bayley's findings with respect to the Stanford-Binet and the Wechsler-Bellevue tests are in agreement with those of other investigators. Most infant scale researchers have indicated that lack of comparability between the infant and school age tests is at least in part responsible for the poor agreement between scores.

With respect to the second factor listed by Bayley as a possible cause of IQ changes, differences in variability within an intelligence scale, as revealed by fluctuating S.D.'s for the separate age levels, are well known even in the well established tests. While the SD of the Revised Stanford-Binet IQ fluctuates around a median value of 16, it is not constant at all ages (1, p. 85). In an examination of the SDs published by Terman and Merrill (65), the differences among the SDs are apparent; the SD at CA 2-6 is 20.7, whereas the SD at CA 6-0 is 13.2. In other words, an IQ of 120 at age two and one-half becomes 113 at age 6, if the child maintains his relative status in the group. McNemar (57) has published a correction table for use with the Stanford-Binet Scale, in order to reduce errors which result from variability.

Bayley examined the trend of the SDs for her group from birth through 17 years and she found that they did not increase at the constant rate which is necessary if IQ's are to remain constant. She found that SDs were too small during most of the first year and too large after seven years, especially after nine, ten and eleven years. "The SDs show strikingly why the IQ is a poor indicator of later intelligence. When IQ's are used
the children's scores were most variable at one month (when the SD was 20) and around nine to eleven years (SDs were as high as 24), and least variable around one year (when it drops below seven IQ points). The variability tends to diminish again when maturity is reached (SD at eighteen years was 12.28)." (10, p. 173) Similar fluctuations in SDs were noted in the Harvard Growth Study data.

With further reference to this second factor, it can be pointed out that while age changes in variability of test scores can be related to the test instruments themselves - because of differences in difficulty, inadequate items and peculiarities of sampling - there is also the possibility that variability is in part a reflection of age changes occurring in the underlying mental functioning. Bayley considered changing mental organization to be a valid explanation of the restricted variability which she discovered at twelve months of age and during adolescence. She compared the SDs for her infant scale with those reported for the Iowa Tests and the Gesell Scale, and found that for all tests and samples and different methods of scoring, there was evidence of decreased variability in scores at or near one year of age, with SDs increasing above and below that age. "The consistency of these trends suggests that children are less variable in behavior-maturity patterns at one year than earlier or later." (10, p. 176) Bayley explains these trends for the infancy and adolescence periods as follows:

... It seems quite probable that both of the clear-cut periods of restricted variability in the Berkeley Growth Study intelligence scores - toward the end of infancy and adolescence -
are due to the approach to maturity of the particular processes being measured. Mental processes during the first year are largely sensorimotor and though they form the basis for further intellectual development, precocity or retardation in them is not necessarily related to rates of development in the more complex processes which we call intelligence in the school-age child and adult. By one year of age most of the slow developers have caught up with those who were precocious in these simple coordinations. The SDs then become restricted to individual differences in mature functions. In the same way the approach to mature intellectual status after eleven or twelve years could reduce the variability of performance as the children whose mental growth is more accelerated reach their own ceilings” (10, p. 178-79).

In that portion of the Harvard Growth Study concerned with the mental development of girls between the ages of eight and sixteen, evidence of another age change in variability due to inner processes was found. At approximately the age of puberty the girls appeared to show much less variation in mental age than they did before and after that time. In the two years following puberty, the variations increased until it was almost twice as much as it was at puberty (24, p. 166).

In an effort to reduce spurious changes in scores which resulted from differences in test construction and from age changes in variability, both the Harvard and Berkeley research groups made use of standard scores in handling their test findings, thereby determining the extent to which the children maintained constant positions in the total group. Bayley, for example, found for the school age children that consistency of their intellectual status relative to each other was very little influenced by the use of different tests, true even though the IQ’s as computed by the several tests’ norms were often quite variable. Because of her use of standard scores, Bayley’s data on the predictive value of her infant scale
is probably more meaningful than those of other infant scale investigators who used IQ's as a growth index.

As stated above, the third condition causing changes in IQ is the possibility that different mental functions, differing in strength and in time of emergence, are being measured on different segments along the growth span. Reference has already been made in Chapter I to the fact that this is the main conclusion from the infancy data of the Berkeley Growth Study, based upon the findings of shifts in relative status within the group and age changes in variability. Changes in the character of mental growth was also one of the conclusions of the Harvard Growth Study. It is worthwhile at this point to quote in full one of the main conclusions of the Harvard Study:

Our findings and experience with the individual growth data do not seem to offer much promise of reliable prediction in the individual case. This is attributable to two possible sources: the crudeness of the instruments by which we measure intelligence and the cyclic character of growth. To be sure there has been improvement over some of the instruments used in the earlier part of this investigation. But anyone who has observed the variability of a group test scores from repeated measures of the same individual is impressed with the hazard of prediction. The second source of difficulty is that growth, both mental and physical, seems to be characterized by cycles. Whether these cycles are due to a biological mechanism or to environmental factors, defined so as to include internal stimuli, or both, remains yet to be determined. Our guess is that they are due to both. It is conceivable, however, that environment, as defined above, is the dominant factor." (24, p. 232)

If inconsistencies in scores can be taken as evidence of changing mental organization, some similar shifting apparently occurs throughout the preschool period as well as during infancy, since tests at the preschool ages have proved to be considerably less effective as predictors than
school-age measurements. Bayley, for example, found that five to six years was the minimum age at which classification into broad categories could be made. Below that age her correlations were too low for significance although they were increasing steadily with age. Two other studies of preschool children concurred with these findings. Bradway (15) who studied a group of 136 preschool children, concluded that an individual IQ obtained prior to the age of six years must be interpreted with caution. "The chances are one in four if the child is four to five years old and one in three if the child is two or three years old, that the IQ will change 15 points or more in the following ten years." (15, p. 215) Honzik (19) noted marked individual differences in test constancy in a group of 252 children of preschool age participating in a longitudinal study of mental growth. "Twenty percent of the children making extremely high or extremely low scores at twenty-one months maintained that position to the six year test. On the other hand, there were many instances of extremely marked changes in mental test scores. One child gained over three sigma between twenty-one months and six years."

Most investigators agree that the IQ becomes more reliable for the child of school age and the adult. This was the finding of the Berkeley and Harvard studies, as well as of numerous other studies conducted on the Stanford-Binet Scale. For example, in the Harvard study, the individual growth curves for 266 girls of school age were examined and it was found that they tended to remain throughout the period of their mental growth to age sixteen in the same classification as they were at age eight.
although a regression toward the mean was noted, with extremely high and extremely low scores moving consistently toward the group average (24, p. 186). (This is in agreement with Bayley's finding of decreased variability toward later adolescence, which she explained on the basis of approach to maturity in mental functioning.) This relative constancy of the IQ in the school age child suggests that any changes in intelligence which might be occurring are very gradual and cover a long period of time.

Environmental Correlates and Intelligence

Many studies have reported on the influences of environment on intelligence test performance of the preschool and the school age child. Institutionalization in particular has been found to have retarding effects on the intellectual as well as the emotional functioning of children in the younger age groups. Two studies have reported on the effects of institutional care upon infant test scores. Fischer (28) conducted a study of 62 infants who had been cared for from birth to a period of beyond six months in a maternity hospital. The mean Cattell IQ of this group of infants when tested in the institution at six months was 76.11. All of the infants tested below the normal range. Thirty-six of these children who were then placed in adoptive homes were later found to have a mean Cattell IQ of 97.54. On the basis of an analysis of the six month test records and the behavior reported at the time of the examination, Fischer concluded that a definite "hospitalism" syndrome occurs in a large number of institutionalized infants of six months which inhibits cooperation in developmental
examinations but which is not yet in the nature of an irreversible pattern. Fischer found that muscle development in these children was not seriously retarded but that their grasping behavior - an adaptive activity - was seriously affected.

Gilliland (40) reported similar findings. When a group of babies six to twelve weeks of age and of similar socio-economic status were compared on the basis of being reared at home or in an institution, the home reared infants were found to be significantly superior in scores by a mean of five IQ points. The infants were tested with the Northwestern Tests.

In line with these results, two studies reported the results of permissive rearing and parental attitudes on test scores. Klatskin (52) offered the fact of flexible methodology in the rearing of a group of infants she studied as a partial explanation of their high mean IQ (112), obtained on the Cattell Scale. She examined the test performances of 316 twelve-month-old infants followed in the Yale Rooming-In project. However, Klatskin also suggested that the Cattell Scale itself was inadequate and in need of restandardization, since her group of babies obtained higher percentages of success on many items than did Cattell's infants. Williams and Scott (69) found significant differences in gross motor development, as tested by the Gesell Developmental Schedules, between two groups of Negro infants, which they ascribed to permissiveness in methods of child care and the home atmosphere.

Gilliland (40) also reported three studies comparing the performances of Negro and white infants on the Northwestern Tests. The
mean IQ of the Negro infants was slightly higher than that of the white infants, but the difference was not significant.

Socio-economic status of the parents has no influence on scores during the infancy period, as reported by Furfey (30), Gilliland (40) and Bayley and Jones (13). Furfey rated the socio-economic status of the parents of 277 infants on the Chapman-Sims Scale. He found that correlations between scores on the Linfert-Hierholzer Scale and the Chapman-Sims ratings were too small to be significant, for the ages under twelve months. Gilliland reported the results of three studies comparing the performances of infants rated in two groups - high and low - in terms of parental socio-economic status. The socio-economic status of the parents was found to have no demonstrable influence on intelligence test scores below the age of thirty-six weeks, which was the entire age range studied.

Bayley and Jones studied the relationship of the mental test scores obtained with children in the Berkeley Growth Study to several variables. The factors included were: education of the mother, education of the father, mid-parent education (the average of both parents' years of schooling), father's occupation, family income, social rating, and a total socio-economic rating. They found that all factors revealed a zero or slightly negative correlation with intelligence test scores up to eighteen months, but beyond that age level certain factors - especially mid-parent education - began to show positive correlations with test scores. For example, at 48 months the correlations between test scores and mother's education, father's education, mid-parent education and father's occupation
were .50, .37, .50, and .31; at 72 months the correlations between test scores and these factors had become .58, .50, .59, and .38. Bayley and Jones concluded that perhaps inherited parent-child resemblances and environmental influences become evident only after a certain point in maturational processes has been reached. "Probably environment and inheritance each have some validity. The growth of children involves both an increasing assimilation of environmental pressures and an increasing manifestation of complex hereditary potentialities." (13, p. 336) The authors believed that their data could not at that time be utilized to distinguish between the nature-nurture variables, or to define their relative importance.

A number of studies have attempted to investigate the "nature-nurture" question by studying the relationship between certain features of the environment and the mental test scores earned by children of varying age groups. Some of the studies have restricted their investigations to children in a single age group; other studies have been of a longitudinal character, in which attempts were made to trace the cumulative effect of environmental influences. Some of the study groups have been confined to children reared with their own parents, whereas others have used adopted children in investigating the relative influences of characteristics of true and foster parents. Since this present study will include such an investigation with adopted children, some of the outstanding studies utilizing the foregoing methods of approach will be summarized in the following paragraphs.

Three studies have investigated the influences of environment
on test scores of children beyond the infancy level and reared with their own parents. Goodenough (42) studied the relationship between mental test scores and parent education for a group of 213 children who were given the Kuhlmann Revision of the Binet Scale between the ages of 18 and 54 months. She discovered correlations as high as .35 with mother's education and .349 with father's education for those preschool ages. Honzik (50) has reported the results of a longitudinal study of 252 children who were given from eight to ten mental tests between the ages of twenty-one months and eight years. She found that the factors considered—namely, the mother's intelligence, parent education, and socio-economic index—showed only a negligible relation to test scores of the children at twenty-one months, but by three and one-half years statistically significant relationships were evident (correlations were .26, .21, .25, and .24 between test scores and mother's education, father's education, mid-parent education and socio-economic index). The most marked increase in relationship occurred between three and three and one-half years but the relationships continued to increase gradually up to eight years, at which time the correlations between test scores and mother's education, father's education, mid-parent education, and socio-economic index were .33, .35, .36 and .41 respectively, and the correlation with the mother's intelligence rating was .54.

In her most recent report of parent-child similarities noted for the Berkeley group, Bayley (11) determined the relationships between test scores up to eighteen years and mid-parent education. She discovered the
correlations between mid-parent education and scores at sixteen, seventeen, and eighteen years had increased to .64, .65 and .60 respectively. Considering parent education as a rough index of parent mental ability, Bayley proposed from her data that a hereditary core of parent-child similarities in both mental and physical characteristics exists, even though it is not observed during the first year and, further, that differences in the mental organization of the infant and the adult could contribute to the changing and increasingly positive parent-child relations in mental ability. Bayley believed that support was lent to her hypothesis by the fact that a study of adopted children (62) had revealed increasing true mother-child relationships which could not be attributed to environmental influences.

Several studies have investigated the relationships between environmental factors and intelligence test scores of adopted children. With only one exception, these studies have found little or no evidence of positive relationships between features of the adoptive home environment and the child's intelligence. Burks (18) studied the influences of environment upon the test scores of a group of 214 children placed in their adoptive homes before twelve months (average placement age was three months) and tested between the ages of five and fourteen years. She compared her findings on the adopted children with data obtained with a control group of 105 children reared by their own parents. She found correlation coefficients between mid-parent MA, father's education, mother's education, and cultural index and the test scores of the adopted children to be .20,
.01, .07, and .25 respectively, whereas correlations for the same factors with the control group were .52, .27, .27, and .44. As a general conclusion, Burks stated that home environment accounts for about 17 percent of the variance in IQ, whereas parental intelligence alone contributes about 33 percent. "The total contribution of heredity is probably not far from 75 or 80 percent."

The highest correlations between environmental influences and test scores of foster children to be found in the literature were reported by Freeman et al (29), who studied a group of 401 foster children living in the Chicago area. Freeman investigated the following features of the foster families: father's intelligence, mother's intelligence, mid-parent intelligence, father's vocabulary, mother's vocabulary, mid-parent vocabulary, home rating, mid-parent education and father's occupation. He found correlations to be .37, .28, .39, .27, .37, .36, .48, .42 and .37. Using the results of studies of children reared with their own parents as a basis for judgment, these figures indicated resemblances as high as those to be expected with offspring from whom both heredity and environment were operative.

Prompted by the discrepancies noted between the Burks and Freeman studies, Leahy (55) studied a group of 194 children placed in their adoptive homes at six months of age or younger. She also studied a control group of natural children for whom environmental conditions were comparable to the study group. She found her data to agree more closely with those of Burks. The correlations between the adopted child's IQ and environmental status,
cultural index, father's education, mother's education, mid-parent education, father's occupation, father's Otis score and mother's Otis score were .19, .21, .16, .21, .20, .12, .15 and .20, whereas correlations for the control group were .53, .51, .48, .50, .54, .45, .51 and .51. The mean age of the children was nine years and four months. Leahy pointed out that the expected familiar resemblance demonstrated in the control group reflected the combined effects of heredity and environment, whereas the correlations for the adopted group represented the effects of environment alone. Further, she concluded that the average coefficient of .18 found between the adopted child's IQ and features of the environment was largely or wholly the result of selective placement. "If further analysis supports this hypothesis, we must conclude that the influence of environment on measured intelligence is relatively insignificant" (55, p. 287).

In 1949, Skodak and Skoels (62) published a final report on a group of one hundred children placed for adoption under six months of age and given a series of four intelligence examinations between the ages of 2-2 and 13-6. They studied the influences of the foster mother's education, foster father's education, true mother's intelligence and true mother's education on test scores. They discovered that figures on the foster parents ratings remained close to zero on all four tests, whereas correlations between the child's IQ and the mother's intelligence rose from .00 at 2-2 to .28 at 4-3, .35 at 7-0, and .44 at 13-6, and correlations between the child's IQ and the true mother's education were .04, .31, .37 and .32 for the four age levels. Somewhat contrary to the hypothesis set
forth by Hayley, as well as by Burks and Laehy, these authors made the following comment with respect to their findings: "This one set of figures must not be permitted to overshadow the more significant finding that the children are consistently and unmistakably superior to their natural parents, and in fact follow and improve upon the pattern of mental development found among own children in families like the foster families." (62, p. 116) Skodak and Skeels believed that this was brought about by certain "dynamic aspects" in the adoptive environment, such as an environment rich in intellectual stimulation, a well-balanced emotional relationship and intellectual agility on the part of the foster parents - features of the environment which are difficult to measure in any way.

Snygg (63) also noted the superiority of the adopted child's IQ when compared with the mother's IQ, when he studied a group of 312 children placed from a Canadian urban institution before four years of age. The mean IQ of the children was 97.17, as compared with a mean IQ of 78.30 found for their true mothers. The children were tested at varying ages from one year to over five years of age. He found a correlation of only .13 between the IQ's of the mothers and the children, and concluded that the mother's IQ could not be used as a basis for prediction in selecting an adoptive home for the child. However, Snygg indicated that the inclusion of test scores on very young children may have masked a true higher correlation between mother's and child's IQ's.

The mean IQ's of several groups of adopted children, as reported in the studies summarized above, can be read from Table II. According to
Table II
Mean IQ's of Adopted Children, As Reported in Published Studies

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Number</th>
<th>Age at Placement</th>
<th>Age at Test</th>
<th>Criterion</th>
<th>Mean IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snygg (63)</td>
<td>312</td>
<td>under 4 years</td>
<td>between 1 and 5 yrs.</td>
<td>Kuhlmann &amp; 1916 Binet</td>
<td>95.7</td>
</tr>
<tr>
<td>Burks (18)</td>
<td>214</td>
<td>before 12 months</td>
<td>between 5 and 14 yrs.</td>
<td>1916 Binet</td>
<td>107.4</td>
</tr>
<tr>
<td>Leahy (55)</td>
<td>194</td>
<td>6 mos. or younger</td>
<td>9 yrs. 4 mos. (mean age)</td>
<td>1916 Binet</td>
<td>110.5</td>
</tr>
<tr>
<td>Skodak &amp; Skeels (62)</td>
<td>100</td>
<td>under 6 mos.</td>
<td>13½ years (mean age)</td>
<td>1937 Binet</td>
<td>116.8</td>
</tr>
</tbody>
</table>

...these investigations, the intelligence levels of children placed in early infancy are superior to the levels found for similar age groups in the general population. This is undoubtedly the result of the careful placement policies of the child-placing agencies.

The majority of the investigations of environmental influences on intelligence test scores of children beyond the infancy period found that higher correlations were obtained between estimates of the parents' intelligence and/or parent education and the child's IQ than between socio-economic...
status and the child's IQ. The exceptions to this were Burks and Leahy who found that the IQ's of the groups of adopted children they studied showed closer agreement with the cultural status of their adoptive parents than with the adoptive parents' educational levels, but these correlations did not approach the levels of significance noted for correlations between child's IQ and own parents' educational and intelligence ratings.

But within the range of socio-economic status, two studies of own children and one study of adopted children have noted agreement between group averages of mental test scores and the father's occupational status. Goodenough (43) classified the occupations of the fathers of a group of two, three and four year old children according to the Minnesota Scale of Paternal Occupations and then computed the mean IQ's of the children for each group. She found the mean IQ's of children whose fathers were in the professional and managerial groups to be significantly higher than the mean IQ of children whose fathers were in the semi-skilled labor group; mean scores for the three groups were 116.1, 111.7 and 96.0. Monszik (50) also found significant differences in mean IQ's when the children were grouped according to their father's occupations, with the mean IQ's decreasing down the scale from professional to the unskilled labor group. When Monszik's children were eight years old, the mean IQ's for the professional and managerial groups were 112 and 124 and the mean IQ's for the semi-skilled and unskilled labor groups were 106 and 104. Skeels (61) found these significant differences to be operating in a group of 73 adopted children placed before six months of age. In agreement with the findings
of other investigators, Skeels found no relationship between father's occupation and mean test scores when the children were under two years of age, but at thirty-six months of age the mean IQ's of the children whose fathers were in the professional, managerial and skilled labor groups was 112.1, as compared with a mean IQ of 98.3 obtained by children whose fathers were in the semi-skilled and unskilled labor groups.

Summary of Findings on Hereditary and Environmental Correlates

Two studies (52, 69) have reported increases in test scores to correlate positively with permissive rearing practices during the infancy period and two studies (28, 40) have reported on the retarding effects of institutional care on infant scores. Otherwise, no positive relationships between test scores during infancy and such factors as education of the parents, socio-economic ratings and occupational status of the father have been found (30, 13, 40). However, relationships between these factors and mental test scores have been found to increase steadily beyond the infancy period, when children reared by their own parents have been studied (11, 50). For example, Bayley (11), who has followed individual growth careers over a long period of time, found that correlations between parent education and test scores at sixteen, seventeen and eighteen years of age had reached a fairly high level of significance. A comparison of the findings of different investigators indicates that there are fairly wide individual differences in the ages at which children reared with their own parents become like that which is representative of the child's home and family.
On the other hand, either zero or very low positive correlations have been found in three out of four studies which examined the relationships between the intelligence of adopted children and factors of their adoptive environments. (18, 55, 62) In the fourth study, positive relationships comparable to those found for children reared with their own parents were reported (29). In the only study which examined the relative influences of both true and adoptive backgrounds (62), a greater agreement between the intelligence of the children and true parent characteristics was found to exist than between the intelligence of the children and aspects of their adoptive environments.

Several of the investigators have concluded from these findings that heredity thus emerges as a stronger force than environment in the intellectual development of adopted children (11, 18, 55), with Leahy (53) also suggesting that the "selective placement" practices of adoption agencies could account for any resemblances which may be found between adoptive parents and their adopted children. However, Skodak and Skeels (62) suggested that resemblances between the intelligence of adopted children and their foster parents were greater than could be measured because of the "dynamic aspects" of the environment which are involved. All studies reporting on adopted children placed in early infancy found the children to be of better than average intelligence (18, 55, 62).
CHAPTER IV
DESIGN OF THE RESEARCH

The Cattell Infant Intelligence Scale

The Cattell Infant Intelligence Scale was published in 1940 as an outgrowth of a longitudinal study of child health and development conducted at the School of Public Health of Harvard University. Constructed as an age scale and a downward extension of the Revised Stanford-Binet, Form L, the scale covers the age range of two to thirty months. Since Stanford-Binet items are interspersed with other items between the ages of twenty-two and thirty months, the author proposes that a continuous intelligence scale from early infancy to maturity has been attained. Five regular items and either one or two alternate items are presented for age levels one month apart during the first year, two months apart during the second year, and for the additional age levels of twenty-seven and thirty months.

In standardizing the scale, 1346 examinations on 274 children were used. Tests were administered at the ages of three, six, nine, twelve, eighteen, twenty-four, thirty and thirty-six months (Stanford-Binet). It was not possible to test all of the children at those ages but they averaged five examinations each. Percent passing was the only method of item analysis used by Cattell in placing her tests on the scale. For the age levels between the standardization ages - two, four, five, seven, eight,
ten, and eleven months during the first year — items were placed by estimation, based on the percent passing at the adjacent standardized age levels.

Items were adapted largely from Gesell and Buhler. A lesser number were taken from other sources. Items were eliminated from the scale if they failed to show sufficient increase in the percentage of passes from one age group to another, or if they increased irregularly in the number of passes from age to age, showed plateaus or failed to approach closely the one hundred percent mark at any age. Additional reasons for eliminating items were the following: (a) items which were difficult to administer or score, or which required an undue amount of subjective judgment on the part of the examiner; (b) items which did not hold the attention of the child; (c) items which required cumbersome apparatus; (d) items which were thought to be unduly influenced by home training; (e) items planned to test control of the large muscle groups; (f) items which appeared to test abilities similar to those covered by other items at the same age level, and (g) items for which a sufficient number of more or equally satisfactory items were available.

On the basis of Stanford-Binet results with 35 children whose test records were complete, Cattell rearranged her items to bring the median IQ for each age level as close as possible to the median IQ of 106 obtained on the Stanford-Binet at thirty-six months. She found that at no age did the median IQ differ by more than two points from the Stanford-Binet median. The mean Stanford-Binet IQ at thirty-six months for this group of children was 105.
Cattell found her scale to be of doubtful validity before twelve months, but of increasing validity thereafter. She measured validity in terms of the scale's ability to predict later Stanford-Binet scores. For the age levels of six, nine, and twelve months and the Revised Stanford-Binet at thirty-six months, the correlations were .10, .34, and .18, respectively. The median IQ changes were found to be greater before than after twelve months. The correlations between scores at the age levels of twelve, eighteen, twenty-four and thirty-months were much higher, .56, .67, .71 and .83 respectively.

The corrected odd-even reliability coefficients found by Cattell were as follows: .56 at three months, .88 at six months, .86 at nine months, .89 at twelve months, .90 at eighteen months, .85 at twenty-four months, .71 at thirty months and .87 with the Stanford-Binet at thirty-six months.

According to Cattell, the fairly rigorous requirements for enrollment in her study group probably resulted in a standardization sample somewhat above the general population in composition, a conclusion which she considered to be partially substantiated by the mean IQ of 105 obtained at thirty-six months. In general, Cattell described her group as being of the "lower middle classes." Enrollment requirements included good physical health and normal delivery, a background of primarily North European stock, more or less permanent employment of the father, and willingness of the mother to cooperate with the study group over a period of years. A few of the parents were professional people, but the majority
were employed in such positions as policeman, clerks, storekeepers, and the like.

The administration of the Cattell Scale is similar to that of the Stanford-Binet, with the exception that serial testing is permitted in order to secure the infant's best efforts and attention. Several items can be scored on the basis of observation of one activity, such as the degree of fine motor coordination displayed by the child in securing a small sugar pellet. The testing manual includes a complete description and an accompanying photograph for every item, thereby lessening the possibility of inadequate administration and scoring. Record forms for the complete scale are available.

Scoring is the same as for the Stanford-Binet. A basal age is established, and to this month level are added additional credits for all succeeding credits when computing the mental age. Since there are five items placed one month apart during the first year, each item receives one fifth or .2 of a month's credit. Thus, an infant who achieves a basal age at the six month level and has three additional successes beyond that level has a mental age of 6.6 months. Similarly, the chronological age is estimated in terms of tenths of months, every three days comprising an additional one tenth of a month. The IQ is computed in the same manner as for the Stanford-Binet.

Only five studies, apart from the original presentation of Cattell, have been published in which a detailed report of some use of the Cattell

1. See Appendix A for sample record form.
Scale is given. Two of these studies (28, 52) investigated environmental influences on infant test performances. These studies were summarized in Chapter III. Two other studies (26, 32) investigated the relationship between test performance and satisfactoriness of the examination, and one article (19), actually a detailed report on psychological examining, presented some limited Cattell findings for comparison with the standardization group.

Carter and Bowles (19) reported that percentages of successes on the Cattell and Gesell tests tended to be consistently higher than those reported by the authors when the tests were administered to two and three-month-old infants at the Wichita Guidance Clinic. Sixty-six two-month-old infants attained an average Cattell test age of 2.8 months and 66 three-month-old infants attained an average test age of 3.7 months. Carter and Bowles concluded that, to a considerable extent, these differences appeared to result from different examining procedures. They also offered two criticisms of the Cattell Scale in relation to their data: (a) The items placed at the two, three and four month levels are heavily weighted with visual tasks, often resulting in high scores for infants who have unusual visual alertness and responsiveness, but only average or even below average abilities in other areas, and (b) the failure of Cattell to make allowances for refusals of tasks decreases the value of the quantitative scores.

Although Klatskin (52) failed to find any relationship between test performance and the satisfactoriness of the examination, two other studies have reported positive findings in this area. Emphasizing the
importance of the infant's test response to the test situation, Escalona (26) reported an attempt to demonstrate an assumed positive relationship between "Optional functioning" of the infant during the administration of an intelligence examination and the predictive value of the examination. Seventy-two children were tested in early infancy with the Cattell Scale and the Gesell Schedules, and a judgment was made in each case as to whether such functioning had been elicited from the child. These children were later retested from one to six times. When the two groups of test-retest series were compared for predictive accuracy, it was found that predictive value was greater for that group of tests initially considered to have elicited optimal functioning. Of the non-optimal group, only nineteen percent were found to remain in the same intelligence range upon retesting, fifty-three percent moved into the higher adjacent range, and twenty-seven percent were in ranges one step removed or more. In determining the quality of test functioning, the following aspects of the test situation were recorded: (a) Quality of the child's motility; (b) his fatigibility and capacity for muscular relaxation; (c) respiratory and circulatory phenomena; (d) quality of responsiveness to objects and persons, and (e) degree of differentiation shown in test behavior. As a general conclusion, Escalona urged the "Gestalt" view of psychological testing, in which the infant's test behavior is considered in conjunction with his actual performance for more effective prediction of future developmental events.

A study by Gallagher (32) on the question of infant responsivity in the test situation reported findings in essential agreement with those
of Escalona. Forty-three infants ranging in age from 4.1 months to 21.1 months were placed in two groups for retests - a Mandatory Retest group, including all of those infants who were suspected of not doing their best on the original test, and a Routine Retest group, for which no special reason for retesting was noted. The Mandatory Retest group made a mean gain of 8.53 IQ points on the retest, a difference significant at the one percent level of confidence. The mean IQ on the first test was 88.05, on the second test, 96.58. The Routine Retest group made no significant gains in scores. The mean IQ for this group on the original test was 100.62, the mean IQ for the retest, 101.25. Changes in range placement of IQ were reported for both groups but much less for the Routine Retest group.

Selection of the Subjects

The subjects of the present study were 110 children - 58 boys and 52 girls - who had been placed for adoption during their first month of life from St. Vincent's and Misericordia Hospitals, the two maternity and infant homes operated under the auspices of the Catholic Charities of the Archdiocese of Chicago. In accord with agency procedures, these children were supervised in their adoptive homes by social workers for a probationary period of six months following placement, and approximately one month before the completion of the legal adoption they were brought by their adoptive parents to the Guidance Department of the Catholic Charities for a psychological examination. In the ordinary course of events, the agency has no further contact with an adoptive family after legal proceedings are over, unless the contact is initiated by the family. However, for this group of
children, the agency contacted the adoptive families and asked them to return the children for a second examination.

The infant testing program has been a part of the agency adoption practices since 1948, and the records of several hundred administrations of the Cattell Scale, all given by the same psychologist, were available to the writer for study. The general aim in selecting records to be used in the study was to obtain a sampling of valid tests administered at the same age level. Further, since relationships between the intelligence of the children, and certain environmental and background characteristics were to be investigated, the sampling had to be representative of all of the children placed in early infancy by the agency. Accordingly, selection was first restricted to records of children who had been tested with the Cattell Scale between January 1, 1950 and June 30, 1952, in order to allow for two previous years of supervised infant testing experience by the administering psychologist. From the tests administered during the 1950-1952 period, all of the records of children who had been tested within one week of their six month birthday were examined in the light of the following additional criteria:

(a) Reasonable indication, based on a consideration of test behavior and the opinion of the examiner included in the report accompanying each test record, that the responsiveness of the infant permitted complete and, in so far as could be determined, valid testing; (b) placement in an adoptive home during the first month of life, (in order to avoid retarding effects of early and prolonged institutional care), and (c) full term gestation.

In the final selection, the test records of 158 children who had
been tested with the Cattell Scale within one week of their six month birthday were adjudged suitable for analysis. The age level of six months was selected as the chief focus of the study for several reasons. Infants of six months are considered to be fairly stable, as far as the quality of their test performance is concerned. They are typically very much interested in their surroundings, and their attention to the test objects is probably more quickly and readily elicited than at any other age during the infancy period. These factors increase the possibility of obtaining valid test scores. Further, their attention span for individual objects is sufficiently long to permit adequate observation by the examiner.

Cattell found that six months was the single age level under twelve months to have the highest correlation with later Binet scores. As an additional reason for selecting this age level, since the majority of the infants referred to the Guidance Department for testing are about six months of age, there was a greater number of records for this group available, and it was felt that findings for this age level would be of some value when using the Cattell Scale in the future.

The next step in the study procedure was that of contacting the adoptive families and requesting their cooperation in permitting the second examination of the children, then in the three and a half to six years age range. Accordingly, the Director of the Catholic Home Bureau wrote a letter to each family, explaining the project and enclosing a postal card.

2. The retests were administered from March to October, 1955.
to be returned as an indication of intent to cooperate. Of the total group, 115 families responded, either by returning the card or by telephoning the writer. Twenty additional families were then contacted by telephone. The remaining twenty-three families could not be located. Of the 135 families thus contacted, 111 eventually brought children back to the clinic for the retest; 19 families indicated they were unable to return due to varied circumstances (several families had moved to other states) and five families indicated that they did not wish to bring the children back for testing.

The Revised Stanford-Binet, Form L, was used as the retest criterion. In order to provide a constant physical environment for the testing situation, all of the children were brought to the Guidance Department for their examinations. The writer administered 107 of the 110 tests; the remaining three were given by two other psychologists of the Guidance Department staff. The examinations were administered and scored in accordance with the directions outlined in the manual (65). As an additional check on scoring, the records of the examinations given by the writer were rescored by another psychologist. The Cattell tests had also been administered in the offices of the Guidance Department, so that the physical testing environment had been the same for all of the babies. The infant tests were all administered by the same psychologist. The

3. See Appendix B

4. The record on one girl was removed from the retest group, after it was decided that her infant test had possibly not been valid because of a temporary physical condition.
testing equipment and other requirements had been in accord with those specified in the Cattell manual (20).

Description of the Subjects

As far as could be ascertained, this group of 110 children is representative of all of the children placed in adoption in early infancy by the Chicago Catholic Charities. The group is also probably representative of the children placed in infancy by most adoption agencies in large urban communities. However, this group of children is somewhat select as compared with the general population. In the first place, the policy of the agency precludes early placement of infants for whom adoption is contraindicated by reason of birth injury, serious physical disorder, or background incidence of mental illness. Second, although almost all of the children in the study group were born out of wedlock, the available data on the education of their true parents (which can be employed as a rough index of mental status) indicated that the true parents surpassed the educational level of the general population. Third, the stimulation and opportunities afforded by the adoptive environs have probably been above the average. According to the data on the educational status of the adoptive

5. The Cattell test performances of the group of 158 infants from which the study group was selected have already been studied in a thesis project by the writer. The mean Cattell IQ for the larger group was found to be 112.9. For a discussion of the validity of the examining done by the psychologist who tested the infants, see Appendix C.
parents, as well as the occupational status of the adoptive fathers, this
group of adoptive parents surpassed the average for the country as a whole
in these areas. Agency standards for adoptive homes are high and investiga-
tions are rigorous, with the result that the families are "middle class" in
character and a genuine desire for a child can be assumed after careful
screening. 6

The mean retest age of the 110 children in the study group was
four years, eight months; the median retest age was four years and
eleven months. The age range was three years, one month, to five years,
eleven months, the standard deviation 1.63 months. Forty-one children
were within the age range of five years to five years, eleven months,
fifty children were within the age range of four years to four years, eleven
months and nineteen children were between the ages of three years and three
years and eleven months.

One hundred and nine of the children were white; one child,
adopted by a physician and his wife, a trained nurse, was Negro. No serious
health problems had been discovered in any of the children. One boy had a
mild congenital heart murmur; one girl had a visual defect serious enough

6. Some reference should be made at this point to the fact that
the agency attempts, in so far as is possible, to "match"
the child to his prospective parents, in terms of race,
nationality, general coloring and education of the true and the
adoptive parents. This practice is known as "selective place-
ment". Reference has already been made (p. 51) to the conclusion
by Leahy that selective placement operates to an unknown extent
in resemblances found between children and adoptive parents.
Thirty-three of the children had no siblings. Sixty-seven had one sibling, also an adopted child; of this group, 29 had an older brother or sister, thirty-eight had younger siblings. One child had five brothers and sisters who were all adopted or placed with the family on a boarding care basis. Nine of the children had either one or two siblings who were natural children of the adoptive parents.

Eighteen of the children in the five to six year age group had some kindergarten experience. The average attendance was about four months. None of the younger children had had consistent nursery school experience of any duration.

Status of the True Parents

The information about the true parents of the children was taken from the case records of the true mothers, compiled at the time the mothers were under active supervision of agency caseworkers while awaiting the birth of the baby. As a general rule, the mothers are known to the agency for some months prior to delivery. Contacts with the caseworkers are frequent and the information gained about the backgrounds of the mothers is thought to be reliable. The information about the true fathers is second-hand and hence may not be as reliable. In a few of the 110 cases, the information given in the records was not complete.

Status of the True Mother

The mean age of the true mothers at the time of the birth of the
child was twenty-four years and four months; the standard deviation was 5.9 years. The age range was sixteen to forty-one years.

Information on the educational status was available for 101 of the mothers. As far as could be determined, the grades reported in the case records represented grades completed. The mean grade completed by the true mothers was 11.43 years, the median was 12.14 years and the standard deviation was 1.74. Nineteen girls, or 19 percent, had had some schooling beyond the high school level; one was a college graduate and five were graduate nurses. Of the remaining 82 girls, 44, or 44 percent, were high school graduates, 29, or 29 percent, had nine, ten or eleven years of schooling, eight girls, or eight percent, had completed grammar school, and one girl had completed seven elementary grades.

Table III gives the distribution of the occupations of 107 of the true mothers.

It can be seen from Table III that the occupational status of the true mothers as a group closely resembled that expected in the general population, for women in the same age group of approximately 24 years.

Status of the True Fathers

The mean age of the true fathers at the time of the child's birth was twenty-eight years and six months; the standard deviation was 5.0 years. The age range was from eighteen to fifty years.

Information on educational status was available for seventy-two of the true fathers. The mean grade completed by this group was 12.48 years;
Table III
Occupational Status of the True Mothers

<table>
<thead>
<tr>
<th>Job Classification</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional *</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Clerical Workers **</td>
<td>40</td>
<td>37</td>
</tr>
<tr>
<td>Skilled Trades</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Minor Retail Workers ***</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Students</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Factory Workers</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Waitresses</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>Domestic and Housewives</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Unemployed</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>107</td>
<td>100</td>
</tr>
</tbody>
</table>

* Teacher, laboratory technician, five nurses.

** Bookkeepers, office personnel and switchboard operators.

*** Shop clerks and cashiers.
the median grade was 12.63 years, and the standard deviation was 2.33.
Eighteen, or twenty-five percent, of the fathers had some work beyond the high school level; eight were college graduates and three had some graduate training. Forty-three, or sixty percent, were high school graduates. Of the group remaining, eight fathers, or eleven percent, had nine, ten, or eleven years of schooling, two were grammar school graduates, and one father had had five years of schooling.

The occupational status of the true fathers, as compared with that of the group of adoptive fathers and the general U.S. population, can be found in Table IV. In addition to the eighty-three employed fathers listed in Table IV, six fathers were in the military service and four were college students.

The Adoptive Parents

Pertinent background information about the adoptive parents was taken from the case records of the Catholic Home Bureau, adoption agency of the Chicago Catholic Charities. Occupational status of the adoptive fathers was verified at the time of the retest.

Status of the Adoptive Mother

The average age of the 110 adoptive mothers at the time of the retest was thirty-seven years and six months; the standard deviation was 4.3 years. The age range was from thirty to forty-six years. The adoptive mothers had been in the thirty-one to thirty-four mean age range at the time the children were given to them as infants. This is several years over the average age
at maternity, but is probably typical of an adoptive mother population.

The mean grade completed by the 110 adoptive mothers was 11.90 years; the median grade was 12.48 years and the standard deviation was 2.21. Thirty-five, or thirty-two percent of the group, had some schooling beyond the high school level; ten were college graduates and one adoptive mother had had some graduate training. Thirty-nine, or thirty-five percent, were high school graduates, twenty-five mothers, or twenty-three percent, had had nine, ten, or eleven years of schooling, and eleven, or ten percent, had graduated from grammar school.

Status of the Adoptive Fathers

The mean age of the 110 adoptive fathers at the time of the retest was thirty-nine years, and four months; the standard deviation was 4.9 years. The age range was thirty to fifty-four years. As a group, they were in the thirty-three to thirty-six year age range when the children were placed with them as infants.

The mean grade completed by the 110 adoptive fathers was 12.75; the median grade completed was 12.80, and the standard deviation was 2.59. Of the total group, forty-eight, or forty-four percent, had had some schooling beyond the high school level; eighteen were college graduates and seven had graduate training. Thirty-five, or thirty-two percent, were high school graduates, twenty-one adoptive fathers, or twenty-percent, had nine, ten, or eleven years of schooling, and one adoptive father had completed seven grades.
Table IV

Minnesota Occupational Scale Distribution of 110 Adoptive Fathers and 83 True Fathers

<table>
<thead>
<tr>
<th>Group</th>
<th>True Fathers</th>
<th></th>
<th>Adoptive Fathers</th>
<th></th>
<th>U.S. Males Percent (1950 census)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
<td>Number</td>
<td>Percent</td>
<td></td>
</tr>
<tr>
<td>I: Professional</td>
<td>3</td>
<td>4</td>
<td>10</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>II: Semi-professional, managerial</td>
<td>9</td>
<td>11</td>
<td>24</td>
<td>22</td>
<td>11</td>
</tr>
<tr>
<td>III: Clerical, retail business, skilled</td>
<td>24</td>
<td>29</td>
<td>44</td>
<td>40</td>
<td>18</td>
</tr>
<tr>
<td>IV: Farmers</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>V: Semi-skilled</td>
<td>18</td>
<td>22</td>
<td>26</td>
<td>24</td>
<td>29</td>
</tr>
<tr>
<td>VI: Slightly skilled</td>
<td>26</td>
<td>31</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>VII: Laborers</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>14</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>83</strong></td>
<td><strong>100</strong></td>
<td><strong>110</strong></td>
<td><strong>100</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

The distributions of the occupations of the 110 adoptive fathers and eighty-three of the true fathers are presented in Table IV, based on the Minnesota Scale of Paternal Occupations (43). This scale is frequently employed as an index of socio-economic status. Table IV also presents for comparative purposes the occupational distribution in percentages of the
The distributions of the occupations of the 110 adoptive fathers and eighty-three of the true fathers are presented in Table IV, based on the Minnesota Scale of Paternal Occupations (43). This scale is frequently employed as an index of socio-economic status. Table IV also presents for comparative purposes the occupational distribution in percentages of the general United States male populations, based on 1950 United States census figures.

It can be seen from Table IV that the group of adoptive fathers is quite superior in occupational and social status to the group of true fathers and to the general employed male population. The majority of the adoptive fathers (seventy-one percent) are in the first three classifications, and the group contains no representatives of the unskilled working class. While there is some overlapping of the status of the adoptive and the true fathers (forty-five percent of the true fathers were in the first three classifications), the majority of the true fathers were in the semi-skilled and the slightly skilled labor groupings. The contrast between the socio-economic status of the adoptive fathers and that of the true fathers is interesting in view of the fact that the mean educational standings of the two groups are very similar.

Table V presents a summary of the educational standings of the true and the adoptive parents, based on the years of schooling completed. According to the figures on educational status contained in the 1950 United States census report, the median number of school years completed by white males and females twenty-five years of age and over and residing in urban
areas were 10.3 and 10.6 years respectively. It can be seen from Table V that both adoptive and true parents are superior to the national levels in educational standing. The two groups of parents are similar to each other in educational status; the adoptive parents are slightly advanced over the true parents but not to a significant degree. The mean mid-parent educational status (the average of the school years completed by both parents) for the group of 110 adoptive parents was 12.71 (S.D. was 2.08); mean mid-parent education for seventy-two sets of true parents was 11.97 years (S.D. was 1.64).

Statistical Methodology

As a preliminary analysis, the frequency distributions of the Cattell and the Stanford-Binet IQ's of the one hundred and ten children were made, and the mean IQ's and the standard deviations of the performances on both scales were obtained. Before applying correlations statistics to the two sets of IQ's, it was determined that the Stanford-Binet scores of the one hundred and ten children could, if desired, be treated as one homogeneous group in comparing IQ's since there was no significant difference found between the measures of variability in two smaller age groupings within the broader age range. For this, the formula for determining the standard error of the difference between standard deviation was used and the standard t test was applied to determine the significance of the difference.

As the main part of the study, the extent of agreement between Cattell and Stanford-Binet IQ's was determined by the use of the Pearson product-moment correlation. This procedure was applied to the larger group
Table V

Educational Status of the True Parents and the Adoptive Parents

<table>
<thead>
<tr>
<th>Years of Schooling</th>
<th>True parents</th>
<th></th>
<th></th>
<th></th>
<th>Adoptive parents</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Father</td>
<td>Percent</td>
<td>Mother</td>
<td>Percent</td>
<td>Father</td>
<td>Percent</td>
<td>Mother</td>
</tr>
<tr>
<td>17 - 20</td>
<td>3</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>7</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>13 - 16</td>
<td>15</td>
<td>21</td>
<td>19</td>
<td>19</td>
<td>41</td>
<td>37</td>
<td>34</td>
</tr>
<tr>
<td>9 - 12</td>
<td>51</td>
<td>71</td>
<td>73</td>
<td>72</td>
<td>56</td>
<td>51</td>
<td>64</td>
</tr>
<tr>
<td>5 - 8</td>
<td>3</td>
<td>4</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>Mean</td>
<td>12.14</td>
<td>11.13</td>
<td>12.75</td>
<td>11.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>12.63</td>
<td>12.14</td>
<td>12.80</td>
<td>12.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>2.33</td>
<td>1.74</td>
<td>2.59</td>
<td>2.21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

of one hundred and ten scores, irrespective of chronological age, and to two smaller chronological age groupings with mid-points of four and five years respectively. In order to determine the extent of agreement between the child's relative status in the total group of the Cattell and the
Stanford-Binet Scales, the IQ's of all of the children on both tests were then converted to standard scores (T scores) and the Pearson product-moment correlation was used to investigate the relationship.

As an additional method of comparing the test and retest scores, the amount of change in IQ points for each of the one hundred and ten cases was found and the mean change in IQ points for the total group of children on the second test as compared with the first test was determined. The cases showing an increase in IQ and the cases showing a decrease in IQ were examined in order to find the extent of such cases, as well as the mean increase and the mean decrease in IQ points.

The resemblances between the intelligence of the children and the following factors were determined by the use of the Pearson product-moment correlation: true father education, true mother education, true mid-parent education, and adoptive father's occupational status, as determined by the Minnesota Scale of Paternal Occupations. Both Cattell and Stanford-Binet relationships to these factors were determined. As a further investigation of the relationship of environmental factors to an adopted child's intelligence, the average Stanford-Binet IQ's of children of adoptive fathers of differing and socio-economic status (as determined by occupational status) were obtained and compared to one another, as well as to findings from other investigations.
CHAPTER V

ANALYSIS OF RESULTS

When tested with the Cattell Infant Intelligence Scale at six months of age, the group of 110 children obtained a mean IQ of 114.4; their median IQ was 114.6 and the standard deviation was 9.2. When the group was retested with the Revised Stanford-Binet, Form L, at a mean age of four years and eight months, the mean IQ was 115.4, the median IQ was 115.7 and the standard deviation was 12.4.

The mean and median scores obtained by these children when tested with the Cattell Scale at six months of age were above the scores expected of a group of infants representing the general population. They were also superior to the scores earned by the six-month-old infants in Cattell's standardization group; Cattell did not report a mean IQ but her median IQ obtained at this age level was 108 (N was 103). Since the one hundred and ten adopted children have maintained their above-average group status on the Stanford-Binet Scale administered at a mean age of four years and eight months, it can be concluded that the early high scores resulted from the quality of the infant performances, rather than from lenient scoring by the examiner or the inclusion of too many easy infant items.

The mean Stanford-Binet IQ of 115.4 obtained with this group of
adopted children is in agreement with the findings of Burks (18), Leahy (55), and Skodak and Skeels (62) who reported above average IQ's for adopted children who had been placed in their adoptive homes in early infancy.

At six months of age only six children had IQ's below 100, and only one of these scored below 90. At the mean age of four years and eight months, only eleven children scored below 100, three of these below 90. There were no IQ's below the dull-normal range on either test. A wider spread in scores was found for the test at the preschool age level; the Stanford-Binet IQ range was 81 to 115, as compared with a range of 84 to 133 on the six month test.

The restricted variability noted for the infant scores is in general agreement with the findings of Bayley (10) who found that variability decreases during infancy as the end of the first year is approached (the standard deviation for the Bayley group at twelve months was seven IQ points), and then increases again steadily during the preschool and school age levels. However, Bayley reported a standard deviation of 13.2 for her group of children when tested at six months as compared with the standard deviation of 9.2 found for the present study group. The mean IQ of Bayley's infants at this age level was 99.1 and some low scores were included. The lesser variability noted for the present group, on the other hand, is probably related to the fact that it is a more highly selected group; the mean IQ of 115.4 and the absence of scores below the dull normal range testify to the success of the child placing agency in its efforts to avoid
placing children of inferior mentality for adoption. This also probably accounts for the fact that the standard deviation of 12.4 IQ points for the three to six year age range is lower than the mean standard deviation of 15.6 points, reported by Terman and Merrill (65) for the age range of three to six years in their standardization sample.

The 110 children ranged in age from three years, one month to five years, ten months. In order to provide for differences in variability within the group due to age changes over this comparatively wide age range, it had been planned originally to divide the Stanford-Binet results of the larger group of 110 children into at least two separate age groups before applying the correlation procedures. However, when 94 of the 110 children were divided into two age groups with mid-points of four and five years, it was found that the variability of scores within the two groups did not differ significantly. The mean Stanford-Binet IQ of the 53 four-year-old children was 118.2, with a standard deviation of 12.8 IQ points; the mean Stanford-Binet IQ of the forty-one five-year-old children was 112.9, with a standard deviation of 11.4. It was therefore concluded that the 110 children could be treated as one homogeneous group.

Consistency of the Cattell and Stanford-Binet IQ's

When the Pearson product-moment method of correlation was applied

7. When the formula for computing the standard error of the difference between standard deviations was applied to the data for these two age groups, the critical ratio was found to be .01, indicating that the difference in variability between groups was insignificant.
to the results of the two intelligence scales, the coefficient of correlation for the total group of 110 subjects was found to be .07. When 94 of the children were divided into the two age groups with mid-points of four and five years, the correlations between Cattell and Binet IQ's for both groups were found to be .03. These coefficients are not statistically significant.

It is evident from these findings that performances on the Cattell Scale at six months of age had little or no predictive value for performances on the Stanford-Binet Scale when the children were in the three to six year age range. The extent of agreement between the infant and the preschool scores for this present group of children is significantly less than that reported by Cattell, the test author; she had found a correlation of .34 between six month Cattell scores and Stanford-Binet IQ's obtained at thirty-six months (N was 49). The correlations reported in this present study are similar to those of Bayley who reported a correlation of .10 between the averages of the sigma scores on the California Scale at four, five and six months and twenty-seven, thirty and thirty-six months (5), and Anderson who found a correlation of .08 between six month performances on a battery of Gesell and Buhler test items and the 1916 Binet at five years. (3)

In order to reduce the effect of changes in scores resulting from age changes in variability, Bayley transformed her infant test scores into standard scores before applying correlation procedures. In this manner, the extent of agreement between the child's relative status
in the group on two or more tests can be determined. Therefore, the Cattell and Stanford-Binet IQ's for the present group of children were converted into standard scores (T scores). The correlation between the standard scores was found to be .16 for the total group of 110 children, indicating better agreement than that found between IQ's but still below any acceptable level of significance.

Another frequently used method of comparing test-retest scores to determine IQ constancy is to find the number of points the IQ has changed on the second test as compared with the first. This information for the one hundred and ten adopted subjects is given in Table VI.

It is clear from Table VI that there were some wide variations in individual scores. Fifty percent of the children obtained retest scores within a ten point range of their Cattell Infant scores; on the other hand, seventeen of the children showed changes of twenty or more IQ points, and approximately one-third of the group changed fifteen points or more.

A closer examination of IQ point changes revealed the fact that fifty-two cases showed a gain in IQ points, fifty-two cases showed a decrease in IQ points, and six children obtained the same scores on both tests. The average increase in IQ points was found to be greater than the average decrease. There was a mean increase of 13.1 IQ points for fifty-two cases and a mean decrease of 11.2 IQ points for fifty-two cases. The amount of increase ranged between one and thirty-five points; the amount of decrease ranged between one and thirty-three points. The mean change is IQ points for the total group, disregarding signs, was 11.5 IQ points.
In order to determine whether the Cattell Scale could detect deviations from the norm "in the direction of both feeblemindedness and superiority" as Cattell had claimed (21), all Stanford-Binet scores under 90 and above 120 were examined in terms of the ability of the infant scale performances to predict them. Forty of the children obtained Stanford-Binet IQ's of 120 or above, and three children had Stanford-Binet IQ's below 90. Of the forty children with superior Stanford-Binet IQ's, only ten had also obtained superior ratings on the Cattell Scale. In other

<table>
<thead>
<tr>
<th>IQ Points</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 5</td>
<td>32</td>
</tr>
<tr>
<td>6 - 10</td>
<td>24</td>
</tr>
<tr>
<td>11 - 15</td>
<td>15</td>
</tr>
<tr>
<td>16 - 20</td>
<td>22</td>
</tr>
<tr>
<td>21 - 25</td>
<td>12</td>
</tr>
<tr>
<td>26 - 30</td>
<td>2</td>
</tr>
<tr>
<td>31 - 25</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>110</strong></td>
</tr>
</tbody>
</table>
words, judging from the present data the odds are only one in four that a superior Stanford-Binet performance can be predicted with the Cattell Scale when the subject is in the six months age group. Of the three low scores (Stanford-Binet IQ's were 84, 85 and 89), two had received infant scores which were above average for this group of children (Cattell IQ's were 117 and 118), and one had been superior (Cattell IQ was 122). The one child who had received an IQ below 90 on the infant scale (Cattell IQ was 84) obtained an IQ of 103 on the Stanford-Binet Scale at a CA of 5-2.

**Intelligence and External Variables**

In Table VII are presented the correlations between the mental test scores of the children and the educational and socio-economic factors of their adoptive environments and their natural backgrounds.

It will be seen from Table VII that there was no relationship between the six month Cattell scores obtained by this group of adopted children and the educational status of their true and adoptive parents, nor with the socio-economic level of their adoptive environments (based on paternal occupation). These results are in agreement with the findings of Furfey (30), Gilliland (40), and Bayley and Jones (13) who also reported a lack of relationship between test scores during the infancy period and the socio-economic and educational levels of the environment for home reared children.

On the other hand, it will be noted further from Table VII that some statistically significant relationships were in evidence when the preschool-age test was administered. The children's scores were found to correlate
Table VII

Correlations Between True and Adoptive Parent Education, Socio-economic Status, and the Child's IQ

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number</th>
<th>Cattell IQ's (6 months)</th>
<th>Stanford-Binet IQ's (mean - 4 yrs., 8 mos.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True father education</td>
<td>72</td>
<td>.03</td>
<td>.31*</td>
</tr>
<tr>
<td>True mother education</td>
<td>101</td>
<td>-.05</td>
<td>.16</td>
</tr>
<tr>
<td>True mid-parent education</td>
<td>72</td>
<td>-.05</td>
<td>.29**</td>
</tr>
<tr>
<td>Adoptive father education</td>
<td>110</td>
<td>-.02</td>
<td>.12</td>
</tr>
<tr>
<td>Adoptive mother education</td>
<td>110</td>
<td>-.07</td>
<td>.06</td>
</tr>
<tr>
<td>Adoptive mid-parent education</td>
<td>110</td>
<td>-.01</td>
<td>.11</td>
</tr>
<tr>
<td>Adoptive father occupational status</td>
<td>110</td>
<td>-.15</td>
<td>.14</td>
</tr>
</tbody>
</table>

* Significant at the .01 level.
** Significant at the .05 level.
most highly with the education of their true fathers (coefficient of correlation was .31) and with the true mid-parent educational status (coefficient of correlation was .29). These values are significantly greater than zero at the one percent and the five percent levels of confidence, respectively. The next highest correlation was .16 between the true mother's education and the child's Binet IQ, but this value is not statistically significant. The coefficients of correlation between educational status of the adoptive parents and the child's IQ were lower, ranging closer to zero. Perhaps the correlations of .31 and .12 with true and adoptive father educational status can be taken as the most representative of the existing relationships between child's IQ and parent education, since the fathers would be more likely to remain in school longer than the groups of mothers, thereby more closely fulfilling their educational potentials. A check of the background information revealed that several true and adoptive mothers left school at early ages to obtain employment or to marry.

When the data from the present study are compared with the findings of other studies of adopted children, general agreement can be noted in so far as there is little evidence of measurable influences on the child's intelligence by the educational level of the adoptive environment, and there is some evidence of a tendency for the adopted children to resemble their true parents, when true parent education is used as a rough index of parent mental ability. The only other study to investigate relative influences of true and adoptive parent characteristics on the adopted child's
that by Skodak and Skeels (62) - found that the true parent-child relationships were still increasing at thirteen and one-half years of age. It is of course impossible to state whether a similar trend will obtain for the present group of children. The correlations from this present study closely resemble those obtained by Skodak and Skeels, provided that true father education is substituted for true mother education; when the one hundred adopted children studied by Skodak and Skeels were tested at a mean age of four years and three months, the correlations between the child's IQ and the education of the adoptive father and mother were .02 and -.04, and the correlation with the true mother's education was .31. It is not apparent why the correlation with true mother education is lower in this study than that obtained by Skodak and Skeels.

In summary, the data on educational status obtained in the present investigation agrees with the results of other investigators who found that (1) agreement between background and environmental characteristics and the intelligence of the child appears only after the infancy period, and (2) in the case of adopted children, there is greater agreement with true parent characteristics than with adoptive parent characteristics, which, if parent education is taken as a rough index of parent intellectual ability, suggests the relatively greater strength of hereditary forces in shaping the mental development of adopted children, but (3) the extent of agreement between the IQ's of adopted children and the education of their true parents is not as great as that found for children of similar ages reared with their own parents (the correlations found by Bayley and Honzik
between mid-parent education and test scores at similar ages were .58 and .34).

Table VII also gives the comparative agreement between the adoptive father's occupational status and the IQ's of the children at six months of age and at a mean age of four years and eight months. It can be seen that whereas there was a negative correlation between this factor and IQ's at six months, the correlation coefficient has increased to .14 at the preschool age. This correlation is higher than those obtained between adoptive parent educational status and child's IQ, a finding also noted by Burks and Leahy, but it is still too low to indicate a positive environmental influence due to socio-economic status. This coefficient of .14 is also lower than the extent of agreement found by Bayley and Honzik between father's occupation and the IQ's of children reared with their own parents, another indication of the relative freedom of the intelligence of adopted children from measurable environmental influences.

It was pointed out in Chapter III that two studies of children reared by their own parents and one study of adopted children had found that when children were grouped according to the occupational levels of their fathers, their mean IQ's tended to relate positively to the paternal occupational level. For example, Skeels (61) who studied the adopted children, employed the Minnesota Scale of Paternal Occupations as a basis for classification and found that the mean IQ of children whose fathers were in Groups I, II and III was 112.1, as compared with a mean IQ of 98.3 obtained with children whose fathers were in Groups IV, V, VI and VII.
Honzik had found means of 118 and 105 for similar groupings and Goodenough reported a mean IQ of 113.7 for Groups I and II, and 96.0 for Group VI.

Table VIII presents the mean IQ's for the 110 adopted children, when they were grouped according to the occupational status of their adoptive fathers, using the Minnesota Scale of Paternal Occupations as the basis for classification.

An analysis of the findings presented in Table VIII failed to reveal any significant relationships between socio-economic status and intelligence, when the 110 adopted children were grouped according to the occupational levels of their adoptive fathers. When the mean IQ of 116.5, found for children whose adoptive fathers were classified in Groups I, II and III, was compared with the mean IQ of 114.3 for Groups V and VI, it was determined that the means did not differ significantly; the t ratio was 0.76. Further, while the mean IQ of 119.8 for children whose fathers were in Group I was higher than for any other single group, it was not significantly higher than, for example, the mean IQ of 114.5, found for children whose adoptive fathers were in Group VI; employing techniques for small samples, the t ratio was found to be 0.89.

It is not clear why these results should differ from those reported by Skeels for a group of 73 adopted children, since the occupational levels of the adoptive fathers in that study were comparable to those in the present study and the mean IQ of 115.3 for the total group of children examined by Skeels (age range was twelve to sixty months) closely approximated the mean IQ of 115.4 obtained for the present study group.
Table VIII

IQ's of 110 Adopted Children in Relation to Adoptive Father's Occupation

<table>
<thead>
<tr>
<th>Father's Occupation</th>
<th>Number (N - 110)</th>
<th>Mean Stanford-Binet IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>10</td>
<td>119.8</td>
</tr>
<tr>
<td>Group II</td>
<td>24</td>
<td>114.4</td>
</tr>
<tr>
<td>Group III</td>
<td>44</td>
<td>115.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>78</strong></td>
<td><strong>116.5</strong></td>
</tr>
<tr>
<td>Group V</td>
<td>26</td>
<td>114.2</td>
</tr>
<tr>
<td>Group VI</td>
<td>6</td>
<td>114.5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>32</strong></td>
<td><strong>114.3</strong></td>
</tr>
</tbody>
</table>

But at any rate the findings from the present study in relation to the influence of occupational status on the child's IQ are in closer agreement with the total picture in which measurable environmental factors prove to be unrelated to the intelligence of adopted children.
CHAPTER VI
SUMMARY AND CONCLUSIONS

Because of similarity in construction to the Revised Stanford-Binet Intelligence Scale, Form L, and convenience in administration and scoring, the Cattell Infant Intelligence Scale, published in 1940, is now in fairly common use in the clinic. It is used mainly as an aid in evaluating the suitability of young infants for adoption; hence its clinical value depends almost entirely upon its ability to predict intelligence at later ages. Very little research on the Cattell Scale beyond the original work of the author has been reported in the literature, and there have been no published studies of the predictive value of the scale in which detailed analyses of validity have been made available.

In her standardization data, Cattell (20) reported a correlation coefficient of .34 to exist between performances of six-month-old infants on the Cattell Scale and on the Stanford-Binet Scale at thirty-six months, which is a comparatively high estimate of infant scale validity. Escalona (25) has indicated that the predictive value of the Cattell Scale was better than that found for other infant scales, but she has never published data to support her statement. On the basis of some very limited research with the Cattell Scale, MacRae (58) suggested that predictive efficiency could
be increased by employing categorical ratings of performance, such as "superior", "average" and so forth, in preference to exact IQ's. When considered together, these three reports could afford some optimism in using the scale for clinical work. But the contrast between these reports and the findings of Nancy Bayley, who has conducted the most thorough investigation of infant intelligence test performances in conjunction with her longitudinal growth studies (5, 8, 10, 11, 12), supports the need for a more thorough analysis of the validity of the Cattell Scale. Bayley has discovered that there is little or no agreement between infant test scores obtained on the California First Year Mental Scale and test scores obtained at later ages by the same children.

The main purpose of this present paper has been to study the predictive value of the Cattell Scale by comparing the performances of a group of one hundred and ten children on the Cattell and the Revised Stanford-Binet Intelligence Scales. The Cattell Scale was administered to the children when they were within one week, plus or minus, of their six-month birthdays; they were retested with the Revised Stanford-Binet, Form L, when within the age range of three to six years.

The one hundred and ten children who served as subjects for the study had been placed for adoption during their first month of life, and some information concerning their true and adoptive parents was available. Therefore, it was decided as a related investigation to duplicate some research done with other groups of adopted children - especially the work done by Skodak and Skeels (62) - and study the relative agreement between
certain true and adoptive parent characteristics and the intelligence of the children.

Accordingly, by employing appropriate statistical procedures, the extent of agreement between Cattell and Stanford-Minst scores was determined, as well as the extent of relationship between the children's mental test scores at the infant and preschool age levels and the following factors: true father education, true mother education, true mid-parent education, adoptive father education, adoptive mother education, adoptive mid-parent education, and adoptive father occupational status. On the basis of these comparisons, the following results were obtained:

1) The total group of one hundred and ten children performed on a better than average level on both the infant and the preschool intelligence tests. In an earlier study done by the writer of the Cattell performance of a larger group of six-month-old infants, from which this present study group was drawn, it was concluded that the above average performance of the group on the Cattell Scale could have resulted from lenient scoring by the examiner or from the inclusion of too many easy items for that age range on the scale. The fact that a representative group of those infants have maintained their above average status on another intelligence scale three to five years later suggests that the high mean score obtained by the group resulted from the quality of the infant performances, rather than from inadequate scoring or defects within the infant scale. We cannot, therefore, conclude on the basis of infant performances alone that the Cattell Scale is in need of restandardisation.
2) The high mean Stanford-Binet IQ obtained by this group of adopted children is in agreement with the findings of other investigators who studied groups of adopted children placed in their adopted homes in early infancy. As nearly as could be determined, this group of children was representative of all children placed as young infants by child-placing agencies functioning in urban settings. The above-average intellectual status of this group, together with the fact that none of the one hundred and ten children scored below the dull normal level and only three children scored within the dull normal range on the Stanford-Binet Scale, testifies to the success of the child-placing agency in its efforts to avoid placing children of inferior mentality for adoption.

3) A comparison of the Cattell and the Stanford-Binet IQ's of the one hundred and ten subjects gave zero-order Pearson correlations, indicating that the Cattell Scale administered at the chronological age of six months had little or no value in predicting Stanford-Binet IQ's for this group of children when in the three to six year age range. When the IQ's were converted into standard scores, thereby overcoming the differences in variability on the infant and preschool tests and permitting comparisons of relative status within the group on the two tests, the correlation was increased to .16 but still remained below any acceptable level of significance.

4) Some wide variations in individual scores were noted. The mean change in IQ for the total group of one hundred and ten children was 11.5 points. Fifty percent of the children scored within ten points of their Cattell IQ on the Stanford-Binet Scale, but approximately one-third
of the group changed fifteen IQ points or more. Three of the children showed changes in IQ points of close to three sigma.

5) An equal number of cases (fifty-two) showed increases and decreases in IQ. However, the mean increase of 13.1 IQ points was greater than the mean decrease of 11.2 IQ points.

6) On the basis of the present data, the odds are only one in four that the Cattell Scale administered at a chronological age of six months can predict a superior Stanford-Binet performance at the preschool age level. There was not a sufficient number of low scores in the present group of children to determine the efficiency of the scale in predicting inferior Binet performances, but the child who obtained the lowest Stanford-Binet IQ in the group had scored above the mean on the Cattell Scale.

7) There was no relationship found between mental test scores of the one hundred and ten adopted children at six months of age and the educational and socio-economic status of the adoptive parents, nor between the six month IQ's and the educational standing of the true parents. This is in agreement with the results of other investigators (30, 13, 40), who also reported a lack of correspondence between infant intelligence test performances and these factors.

8) When scores on the preschool test were compared to these socio-economic variables, some significant correspondence between the child's intelligence and true parent characteristics was noted. This was especially true of agreement between the child's IQ and true father's education, and the child's IQ and true mid-parent schooling. Conversely, correlations
between the intelligence of the children and socio-economic factors in the adoptive environment remained low. These findings are in agreement with those of the only other study to investigate the relative influences of true and adoptive parent characteristics, provided true father education is substituted for the true mother education. The reason why the child's IQ should show a lower correlation with the true mother's educational level than with that of the true father for this group of children is not immediately apparent, unless of a larger proportion of these true mothers left school before fulfilling their educational potential.

9) A comparison of the mean IQ's of children whose adoptive fathers were in professional, semi-professional, managerial and skilled labor classifications with the mean IQ of the children whose adoptive fathers were in the semi-skilled and unskilled laboring classes failed to reveal any significant difference between the intellectual levels of the children. Since this is in contrast to findings with children reared by their own parents, it can be taken as an added indication of the relative freedom of the intellectual development of adopted children from measurable influences of their adoptive environments.

Conclusions

It is evident from these findings that the Cattell Scale is unsuccessful as a predictive clinical instrument. The data points clearly to the fact that there were such wide variations in scores among the children studied that in only one out of every two infant examinations could the
examiner hope to come reasonably close to estimating later status. Further, the present results disprove the claim by Cattell, the test author, that the Cattell Scale can detect extreme variations from the norm even at early ages. These findings render the scale of little value in clinical practice in predicting for individual children.

The reason why the Cattell Scale should have such low predictive efficiency cannot be readily explained by a study such as the present one. Such conclusions are more appropriately drawn from longitudinal studies in which children are tested at all key ages along the growth span; from longitudinal findings individual growth curves can be charted and general group trends in mental development can be noted and evaluated. But it is possible to conjecture from the present data, particularly in view of the fact that the findings are in fairly close agreement with those from Bayley's longitudinal study.

It was pointed out earlier in this paper that low predictive efficiency of an infant scale could be the result of inadequate construction of the infant scale, causing spurious estimates of ability during the infancy period, or it could derive from the nature of infant behavior itself, in so far as infant activities may not yet be related in any externally observable or measurable way to the more abstract activities which we consider to be intelligent behavior in the adult. It was also pointed out earlier that this latter alternative has been Bayley's conclusion.

There is no reason to conclude from the present data that the Cattell Scale does not permit an adequate evaluation of current infant
status. It is of course possible that the value of infant scales, such as the Cattell Scale, can be increased to some extent by refining our present knowledge of infant behavior and by locating other aspects of infant development which are more closely related to later intellectual functioning than the activities we now consider important. But the limited range of infant behavior does not permit much optimism in this direction.

Rather, changes in the measurable aspect of behavior as the transition is made from infancy to later ages seems to this writer to be the more logical explanation of the poor agreement between infant and later test scores. In the first place, the correlations are very low, and it is unlikely that current infant scales which are fairly adequate in covering all aspects of infant behavior, could be so wide of the mark in locating the abilities that are measured at later ages. In the second place, according to the findings of the present study and the research by Furfey, Bayley and Jones, and Gilliland, those aspects of mental development which are most prominent during the infancy period are relatively independent of hereditary and environmental influences, at least in so far as we are able to measure such influences; but the fact that there is a tendency noted toward increasingly positive relationships to these factors as the child matures gives some support to the conclusion that infant behavior is not yet representative of higher order "intelligent" activities, and therefore would not show the same characteristics as that of the school age child and the adult.

This does not necessarily imply that there are qualitative
differences between the behavior of the infant and the adult. Undoubtedly the elements of intelligent behavior are present in some form in the infant. But it is logical to assume that as mental development proceeds through gross to fine stages, a refinement of intellectual activities occurs and some shifting of emphasis on various functions likewise takes place. For example, during infancy, a period of very rapid growth, the physical maturation of the organism is most prominent, but this is not the case at later ages after bodily postural control has been acquired and visual-motor coordination is established. Further, physical and intellectual development, while closely related, need not be equivalent in strength nor in developmental rate; in the absence of a direct relationship, an infant who shows rapid and precocious physical development need not show parallel development in intellectual functioning at a later age.

The possibility that such age changes in the measurable aspects of behavior do occur provides a pessimistic outlook for predictive infant intelligence testing. If this conclusion is correct, the best that could be expected of an infant scale would be an accurate appraisal of the infant's developmental rate at the particular age tested. There is some limited evidence that multiple correlation techniques can improve currently available infant test groups to some extent; this approach to the problem of infant testing has not been well explored.
REFERENCES


69. Williams, Judith R., & Scott, B. Growth and development of Negro infants. Child Develop., 1953, 24, 103-121.
INFANT INTELLIGENCE SCALE

Record Form
Copyright 1940 by Psyche Cattell

No. ........................................... Name .............................................................................................................................. .

Examiner ........................................ Race ........................................... Age ...........................................

Referred by .................................... Date of Exam. ..................................... M.A. ..................................

Test Satisfactory ................................ Date of Birth ..................................... I.Q. ..................................

Remarks:

<table>
<thead>
<tr>
<th>Willingness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Self-confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social-confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

2nd Month
1. Voice, attends (supine)
2. Inspects environment (supine)
3. Ring, follows, horizontal (supine)
4. Follows moving person (supine)
5. Babbles or coos
   a. Ring, follows vertical (prone)
   b. Head, lifts (prone)

3rd Month
1. Ring, follows in circle (supine)
2. Feeding, anticipates (bottle)
3. Cube, regards (sitting)
4. Spoon, regards (sitting)
5. Fingers, inspects (supine)
   a. Chest, lifts by arms (prone)
   b. Head erect and steady

4th Month
1. Fingers, manipulates (supine)
2. Hands, open
3. Ball, follows (sitting)
4. Voice, turns to (sitting)
5. Activity increased at sight of toy (supine)
   a. Rattle, recovers from chest (supine)
   b. Rattle, active play (supine)

Month

<table>
<thead>
<tr>
<th>PSYCHOLOGICAL CORPORATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>522 Fifth Avenue,</td>
</tr>
<tr>
<td>New York 18, N. Y.</td>
</tr>
</tbody>
</table>
5th Month

1. Bell, turns to (sitting)
2. Ring, attains (supine)
3. Transfers object from hand to hand (supine)
4. Pellet, regards (sitting)
5. Spoon, picks up (sitting)

   a. Rattle, attains at shoulder (supine)
   b. Ring, pulls down (supine)

(After 5 months all items are given in the sitting position)

6th Month

1. Cube, secures
2. Cup, lifts
3. Mirror, manipulates
4. Reaching, unilateral
5. Reaching, persistent
   a. Cube, approaches 2nd

7th Month

1. Pellet, attempts
2. Mirror, pats and smiles
3. Ring, inspects
4. Cube, takes two
5. Paper, exploits
   a. String, grasps
   b. Peg, pulls out

8th Month

1. Ring, pulls by string
2. String, manipulates
3. Says "dada," etc.
4. Pellet, secures
5. Bell, interest in details
   a. Hand preference
   b. Spoon, bangs

9th Month

1. Pellet, scissor grasp
2. Spoon, looks
3. Bell, rings
4. Adjusts to gesture
5. Adjusts to words
   a. Imitates sounds

10th Month

1. Toy, uncovers
2. Cup and cube, combines
3. Third cube, attempts
4. Spoon-rattle, hits outside
5. Peg board, fingers holes
   b. Spoon-cup, spoon first

11th Month

1. Pellet, plucks
2. Cube under cup, secures
3. Box and stones
4. Words, one
5. Cube in or over cup
   b. Doll, squeaks
12th Month
1. Spoon, imit, beating
2. Cubes, in cup, one, No..............
3. Pencil, marks
4. Spoon-rattle
5. Words, two (list)
a. Doll, hits in imitation

13th and 14th Months
1. Words, three (list)
2. Cube, unwraps
3. Glass, frustration
4. Pellet-bottle, imitates
5. Peg, out and in
   a. Cube, takes third
   b. Box, opens

15th and 16th Months
1. Formboard, round block
2. Words, five (list)
3. Beads in box
4. Pellet-bottle, solves
5. Round box, closes
   a. Pegboard, urges No. placed..............
   b. Scribble in imitation

17th and 18th Months
1. Cubes, 10 in cup, No..............
2. Doll, one part
3. Formboard, Rd. hole rev., a.............. b.............., 1
4. Pencil, scribble
5. Picture, points to one
   a. Asks with words. Examples..............
   b. Pegboard A. No. placed..............

19th and 20th Months
1. Tower of three
2. Formboard, square
3. Stick, attains object
4. Doll, commands, two
5. Doll, points to three
   a. Selects box containing toy
   b. Pegboard B

21st and 22nd Months
1. Square box, covers
2. Words, combines
3. Formboard, solves (small)
4. Pictures, points to two
5. Doll, commands, 3
   a. Doll, points to 5
   b. Identifies object by name, 2

Pencil
Marks, 12
Imitates, 16
Scribble, 18
Stroke, 27
H-V line, 30
Stroke-circle, 30

Tower
1st trial..............
2nd trial..............
3rd trial..............
Other..............

Pegboard
Pulls out, 7
Fingers, 10
Out and in, 14
Urged, 16
A, 18
B, 20

Formboard
Rd. block, 16
Rd. Rev., 18
Square, 20
Solves, 22
Solves Rev., 30

Words spoken
Dada, 8
1, 11
2, 12
3, 14
4, 5, 16
6, 7, 8, 9, 10,
Est. No..............
Ask with words, 18
Combines words, 22

Doll-Chair
Chair
Drink
Nose

Doll, points
Hair
Mouth
Ears
Hands
Eyes
Nose
Feet
23rd and 24th Months

1. Identifies objects by name, 4
2. Paper, attempts fold
3. Watch, incomplete, 3rd
4. Stanford-Binet commands, 2
5. Names objects, 3
   a. Picture vocabulary, 3
   b. Cubes, replace in box

3rd year 1st quarter (25th, 26th and 27th Month)

1. Train, blocks in row
2. Egg beater
3. Pencil, imitates stroke
4. Picture vocabulary, 7
5. Pictures, points to 6
   a. Names objects, 4
   b. Digits, 4–7, 6–3, 5–8, 1

3rd year 2nd quarter (28th, 29th and 30th Month)

1. Tower-bridge
2. Pencil, H–V…… S–C…………………, 2
3. Formboard, rotated, 1
4. Paper, folds definitely
5. Identifies by use, 4
   a. Pictures, points to, 7
   b. Cube just one

3rd year 2nd half (S–B, III)*

1. Stringing beads (4+) (2 min.) No. ..........................
2. Pict. voc. (12+) No. ..........................
3. Block bridge
4. Pict. mem. (1+) a………… b…………
5. Circle (1+) a………… b………… c…………
6. Three dig. (1+) 641…… 352…… 837……
Alt. Form board: rotated (2+)

4th year 1st half (S–B, III–6)*

1. Simple commands (3+) a………… b………… c…………
2. Pict. voc. (15+) No. ..........................
3. Compar. sticks (3 of 3, or 5 of 6)
4. Pict. I (2+) a………… b………… c…………
5. Ident, by use (5+)
6. Compre. I (1+) a………… b…………
Alt. Cross

4th year 2nd half (S–B, IV)*

1. Pict. voc. (16+) No. ..........................
2. Obj. from mem. (2+) a………… b………… c…………
3. Pict. compl.: man (1 point)
4. Pict. ident. (3+) No. ..........................
5. Forms (8+) No. ..........................
6. Compre. II (2+) a………… b…………
Alt. Sent. mem. I (1+) a………… b…………

* The items for these ages are copyright, 1937, by Houghton Mifflin Company and used by permission of and special arrangement with, the publishers, Houghton Mifflin Company.
Dear Mr. and Mrs. __________________

The Guidance Department of the Catholic Charities in cooperation with the Department of Psychology of Loyola University is conducting a survey to determine the validity of psychometric tests on small children. Since your child _________________ was tested by the Guidance Department some years ago, they are anxious to do a retest for purposes of comparison. We are asking you to cooperate in this program because it will help our future adoptive work and will be a contribution to educational values.

Will you return the enclosed card indicating your attitude toward this project? It will mean some inconvenience to you since the child will have to be brought in for the test. But we do feel that this is an important phase of our work and we most strongly urge you to comply with our request.

Sincerely yours,

CATHOLIC HOME BUREAU

By ___________________________ Director

Enc.
APPENDIX C

NOTE ON THE VALIDITY OF EXAMINER SCORING OF CATTELL SCALE PERFORMANCES*

The six-month Cattell test performances of these 158 children were studied in a thesis project by the present writer. The mean IQ for the group was found to be 112.9. The infant tests were not administered by the writer, although they were all administered by the same psychologist. The mean IQ of 112.9 is significantly higher than the mean IQ of 100 expected in a random sample of the general infant population. Three alternatives can be offered in explanation of this significant difference. (1) Our sample is not representative - i.e., our group of 158 infants was actually of better intelligence than the general population. (2) The Cattell Scale itself was too easy for this age group, thereby enabling the infants to obtain better than expected mean scores. (3) The examiner was too liberal in scoring test performances.

The first two alternatives can be affirmed or denied only when we have the results of the administrations of the Stanford-Binet Scale, which is of course considered to be well established in terms of its validity. In an effort to come to some conclusion about the third possible explanation, the writer examined the results of 100 Cattell tests which she administered during her second year of infant testing. Like the infants in the study group, these 100 infants were all in the six months (plus or minus one week) age group. The mean IQ for this group was 110.9. The difference between the two means was not statistically significant.

* Taken from the dissertation outline submitted by the writer.
The dissertation submitted by Miss Patricia Bledsoe has been read and approved by five members of the Department of Psychology.

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated, and that the dissertation is now given final approval with reference to content, form, and mechanical accuracy.

The dissertation is therefore accepted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

October 8, 1952

[Signature of Adviser]