1961

An Item Analysis of the Cattell Infant Intelligence Scale

Mary Lou Strassmaier
Loyola University Chicago

Recommended Citation

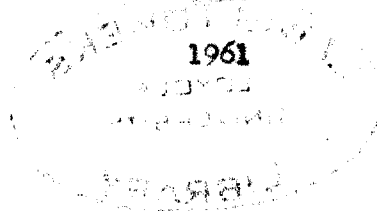AN ITEM ANALYSIS OF THE CATTELL INFANT INTELLIGENCE SCALE

by

Mary Lou Strassmaier

A Thesis Submitted to the Faculty of the Graduate School

of Loyola University in Partial Fulfillment of

the Requirements for the Degree of

Master of Arts

June

1961

LIFE

Mary Lou Strassmaier was born in Eugene, Oregon, July 12, 1936.

She was graduated from St. Francis High School, Eugene, Oregon, June, 1954, and from Marylhurst College in Portland, Oregon, June, 1958, with the degree of Bachelor of Arts.

Graduate studies were begun by the writer at Loyola University in September, 1958. In October, 1958, she began working in the capacity of a trainee in clinical psychology at the Guidance Center of Catholic Charities of the Archdiocese of Chicago.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

The somewhat knotty problem of infant intelligence testing has confronted interested psychologists since the pioneering efforts of Kuhlmann, Gesell, and Buehler in 1922, 1927, and 1935, respectively. Since these early efforts, others have followed and all, with the possible exception of Gesell, have found themselves immersed in essentially the same difficulty, the poor predictive status of the resulting tests. Although various explanations have been offered, the problem continues to be an open one and one within which further investigation seems essential from a theoretical as well as a practical point of view.

Theoretically, the investigation of infant mental development is important in terms of that scientific curiosity which motivates psychologists to study human behavior from the cradle to the grave. The researcher is interested in investigating developmental behavior in order to study its structure and specific function, and to attempt to isolate the emergence and follow the evolution of the more intricate high order mental processes observable in children and adults.

Valid infant testing is of most immediate interest to the practitioner. From the standpoint of positive mental health it is important to detect mental deficiency at as early an age as possible in the interests of

1

parental guidance as well as planning for the care and training of the retarded child. Finally, one of the most extensive uses of infant intelligence tests is made by adoption agencies. It is essential that the mentally retarded be screened out and that the infant's intellectual potential be estimated so that he can be placed with a family of similar capabilities.

In spite of the urgent needs for valid infant tests, validity studies of those tests currently available have yielded disappointing results. A recent study by Bledsoe summarized those investigations considered most indicative from a methodological point of view. While the correlations resulting from these investigations varied from -.20 to .47, all were low when compared to correlations yielded by most intelligence tests used at the childhood and adult levels.[1] Furthermore, longitudinal studies have pointed to considerable long term variability in individual test performance. The Berkeley Growth Studies, a twenty-five year project the primary purpose of which has been to study the development of intelligence as measured by tests, are perhaps the most extensive example of such research. Data published by Bayley as a result of the Berkeley Studies indicate that the most variability occurs during the first eighteen months or two years of life and then begins to taper off as the growth of the child becomes less rapid and more constant.[2] In the light of such early unstable growth careers, one

---

[1]Patricia Bledsoe, "The Predictive Value of the Cattell Infant Intelligence Scale," Unpublished Doctoral Dissertation (Loyola University, Chicago, 1955), p. 18.

[2]Nancy Bayley, "Mental Growth in Young Children," Nat. Soc. Stud. Educ., Part II (February, 1940), p. 29.

would expect poor predictive efficiency of infant tests, especially when retest age extends beyond the infancy period as it does in the studies reported by Bledsoe.[3]

Some explanations of IQ inconstancy in infancy have been summarized by Bayley in various publications:

(a) "Intelligence is fundamentally innate, but...its growth is irregular because it is the resultant of numerous functions that develop both concomitantly and successively, each growing out of previously established abilities but developing in its own way. As a result mental organization changes as new components develop and as older matured ones cease to play an important role in differentiating abilities in all except cases of extreme retardation."[4]

(b) Intellectual growth is inherently stable and could be predicted if adequate tests were used and used correctly. Bayley points out that this hypothesis assumes that intellectual development progresses consistently because it is a unit factor or a "g" factor which could theoretically be isolated from specific factors.

(c) If intelligence is native and composed of a unit factor which increases in efficiency throughout childhood, then the rate of increase cannot be predicted because it is variable for some unknown reason.

---

[3]The reader is referred to Table 3 in the appendix.

[4]Bayley, "Mental Growth in Young Children," p. 30.

(d) Opposed to the former is the assumption that variations in rate of mental growth are the results of different environmental and emotional influences which act on the organism in different ways at different stages of development.[5]

In her discussion of these hypotheses, Bayley considers the first to be the most tenable in the light of the Berkeley Growth Studies' findings. The second seems to be nullified by her failure to find a general factor during the first three years of life, and by the lack of correlation between scores during the first nine months and those at four years of age. In rejecting the fourth hypothesis, Bayley points out that the scores do become stable prior to the time the children are subject to the leveling experience of similar school environments, and that school entrance does not seem to reintroduce instability. While she admits that environment may be operative in changing scores, its effect should be suspected only when there is a wide diversity of home background and opportunities for stimulation, or in the case of children suffering from the effects of accidents or disease.[6] Bayley concludes then, that intellectual growth is composed of a group of developing functions which are, in their early stages, separately timed, inherently unstable, and incapable of prediction except in cases of extreme retardation.[7] Her most recently published position in regard to the value of infant tests is that they are useful

[5] Ibid., p. 30.

[6] Ibid., p. 31.

[7] Nancy Bayley, "On the Growth of Intelligence," Amer. Psychol., Vol. 10 (December, 1955), pp. 806-808.

as a gross evaluation of the normality or abnormality of developmental status, and for research purposes in terms of the relationship of development to learning and ways in which personality structures are formed.[8]

The long term research of Arnold Gesell at the Yale Clinic has, on the other hand, led him to conclusions opposing those of Bayley. Gesell stresses the orderly progress of growth and its predictibility when clinical rather than exclusively psychometric techniques are applied. His studies of thirty children, from infancy to the teens or older, have led him to consider growth as a patterning process which results in behavior which is symptomatic of the maturity of the nervous system. Development, he indicates, is capable of diagnosis because the patterning process is determined by lawful growth forces.[9] "Growth is a progression which proceeds in orderly relation to age from conception to death. A scientific extrapolation of the age factor will inevitably lead to the pattern of future growth. The refinement of this extrapolation is prediction. Growth is lawful. In spite of its bewildering complexity, the growth of the child mind will also be found to be within the realm of law."[10] Gesell continues to point out, however, that his success in prediction does not depend entirely upon psychometric procedure but upon behavioral examinations supported by interview, developmental history, and incidental

---

[8]Nancy Bayley, "Value and Limitations of Infant Testing," Children, Vol. 5 (July-August, 1958), p. 133.

[9]Arnold Gesell and Catherine Armatruda, Developmental Diagnosis (New York, 1941), p. 4.

[10]Arnold Gesell, Infancy and Human Growth (New York, 1928), p. 405.

observations.[11]

As Bayley and others have pointed out, Gesell's limited use of quantitative scoring and statistical methods open to question his conclusions in regard to the predictibility of mental growth.[12] However, it is interesting in this respect to report a study by Simon and Bass which attempted, as did Gesell, to make a clinical evaluation of infant development but with the application of statistical procedures in order to measure the significance of the results obtained. The study, described by the authors as a pilot study, attempted to control biological and life experiential data and the emotional relationship of the infant to the test situation as qualifying test findings. As part of the study, a group of 45 infants were examined with Cattell and Gesell items at an age of less than one year. Later examinations were administered, post-school age, to the same subjects. The initial measure of contingency was .45, with an acceptable level of significance obtained. When the group of records were adjusted for the depressing effects of non-optimal conditions such as poor previous foster home care, long term hospitalization, etc., the degree of association improved to .63 and again attained an acceptable level of significance. The results of this research, then, indicate that an acceptable degree of prediction is possible when clinical rather than strictly psychometric procedures

[11]Arnold Gesell, The Mental Growth of the Preschool Child (New York, 1926), pp. 408-409.

[12]Nancy Bayley, "Mental Growth During the First Three Years of Life: a Developmental Study of Sixty-one Children by Repeated Tests," Genet. Psychol. Monogr., Vol. 14 (July, 1933), pp. 75-76; Bledsoe, p. 3; Anne Anastasi, Psychological Testing (New York, 1954), pp. 282-283.

are utilized.[13]

In spite of the disagreement as to the possibility of validly measuring mental ability as early as the infancy period, infant tests continue to be in rather widespread use. For example, the Child Welfare League of America reported that 62 of its 96 member agencies used infant testing to help determine adoptability and the selection of adoptive families.[14] It would seem essential to continue research with infant tests in order to explore every possibility for their valid use. A number of investigators, rather than taking the clinical approach to the problem as did Gesell, Simon and Bass, have turned their attention to the construction of infant tests with the hope that more adequate item selection would improve over-all validity. Underlying these efforts can be found agreement with Bayley's second hypothesis mentioned above, that mental ability can be measured as early as infancy if only the instruments of measurement were more precise. While a number of studies, to be discussed in the next chapter, have made some item analyses, the Cattell Infant Intelligence Scale, considered to be one of the most satisfactory instruments available,[15] has never to the writer's knowledge been treated in this manner.

The writer has at her disposal 110 Cattell Scale records of six month old

[13]Abraham J. Bass and Libby G. Simon, "Toward a Validation of Infant Testing," Amer. J. Orthopsychiat., Vol. 26 (April, 1956), pp. 340-350.

[14]Child Welfare League of America, Adoption Practices, Procedures, and Problems (New York, 1952), p. 25, cited by Simon and Bass, p. 340.

[15]Anastasi, p. 283.

infants who were tested for adoptive placement by the Chicago Catholic
Charities. The group was retested with the Revised Stanford-Binet,
Form L, as a result of a study done by Patricia Bledsoe in partial ful-
fillment of the requirements of Loyola University for the degree of
Doctor of Philosophy in clinical psychology. The purpose of Bledsoe's
study was to investigate the predictive value of the Cattell Scale. As
a result of her investigation, Bledsoe concluded that the Cattell Scale
is unsuccessful as a predictive 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...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."[16]

While the results of the above research are disappointing from the
point of view of total score, the foregoing discussion suggests value in
investigating the predictive strength of the individual items. It would
be ultimately hoped that studies similarly designed on different infant
tests could explore and compare the behavior and predictive potential of
individual items, thereby providing at some later date an item pool
from which could be drawn a more adequate test than those currently in
existence.

---

[16]Bledsoe, p. 51.

Accordingly, the writer's research will be concerned with the extent of agreement between the performance of the infant sample on the individual Cattell items and their later total test performance on the Stanford-Binet. A statistical analysis of such a comparison may be expected to shed some light on the predictive aspect of the items studied. Specifically, the null hypothesis, that no relationship exists between Cattell Scale items administered at six months of age and later Stanford-Binet performance, will be investigated. The results of this research will then be compared with those of similar investigations in order to observe and discuss similarities and differences.

CHAPTER II

REVIEW OF THE LITERATURE

Because this research purports to explore the validity of test items
currently available in the Cattell Infant Intelligence Scale, it was deemed
necessary to discuss the problems inherent in finding items appropriate to
infant testing, to investigate the methods of item selection utilized by
the authors of infant tests, to discuss the types of behavior measured
by test items, and finally, to report studies similar to that of the writer
which have attempted to evaluate the predictive efficiency of infant test
items.

Difficulties of Infant Testing

The period of infancy usually considered to extend from birth up to
eighteen months,[1] is perhaps one of the most interesting periods of life
to study due to the rapid rate of growth which is observable. On the other
hand, the infant presents formidable difficulties as the subject of
psychological testing. Because the infant relates to his surroundings in
a primitive, nonverbal manner, it is necessary for the examiner, as
Anastasi points out, "to set the stage" so that the desired response is

---

[1] Anne Anastasi, Psychological Testing (New York, 1954), p. 278.

10

stimulated.[2] Such a procedure necessitates test items which, because of their intrinsic interest, will capture and maintain the infant's attention long enough to provoke a scoreable reaction. As Anastasi also indicates, the scoreability of the response poses an added difficulty due to the absence of an objective criterion of performance.[3] To complicate the procedure even more, the nature of the infant response, considerably more than that of the child or adult, appears to lie on a continuum ranging from an emphatic negative to a perfect plus. Rather than being strictly dichotomized on a pass-fail basis, most individual responses seem to lie somewhere between the positive and negative poles of this continuum.

Bledsoe has suggested that the physical and experiential immaturity of the infant narrows his range of responsiveness considerably, thereby limiting that observable behavior from which test items may be drawn. The absence of language in particular prevents the tester from tapping those aspects of general intelligence, such as the abilities to abstract and relate, which may be presumed to be at least potentially present.[4] As a result, a survey of currently available infant test items reveals an abundance of sensorimotor tasks, especially at age levels prior to one year, which have been shown to have little relationship to later intellectual status.[5]

----

[2]Ibid., p. 287.

[3]Ibid., p. 287.

[4]Bledsoe, p. 3.

[5]Nancy Bayley,"Consistency and Variability in Growth of Intelligence from Birth to Eighteen," J. genet. Psychol., Vol. 75 (December, 1949), p. 178.

Methods and Criteria for Item Selection

One of the first attempts at infant testing comprised a special revision and downward extension of the Binet scale. In 1922, Kuhlmann's revision of the Binet extended the test down to the third month level.[6] In 1939, Kuhlmann revised his scale further using four criteria for item selection and placement:

(a) The increase in median raw score on a test between adjacent age levels was reported to be the chief criterion of item validity.

(b) Increase at successive age levels in percentage passing was also considered important; each test was placed at the exact level at which 50 per cent of the subjects passed.

(c) Variability in raw score was also considered a valuable criterion, high variability being regarded as undesirable on the assumption that it might prove to be an index of unreliability and/or validity.

(d) A measure of internal consistency was obtained by correlating each item with total test score.[7]

A survey of those infant test manuals now available indicates that most authors have relied heavily on items originally developed by Gesell and to some degree those initiated by Buehler and Wolfe. The Gesell Developmental Schedules, first published in 1925, and most recently revised in 1947, were the result of long term observations of infant behavior at

[6]Anastasi, p. 283.

[7]F. Kuhlmann, Tests of Mental Development (Minneapolis, 1939), pp. 18-22.

the Yale Clinic of Child Development. Items deemed to be representative of four behavioral areas, motor, language, adaptive, and personal-social, were selected and placed by determining that age at which an aspect of behavior could normally be expected to make its appearance.[8]

Buehler's Babytests were constructed on the basis of 24 hour laboratory observations of infant behavior. Attempts were made to include behaviors which were natural and characteristic of infants at particular ages.[9] Buehler reports that items were eliminated under the following conditions: for technical reasons, because they did not fit in with the infant's interest or capacities of attention, when the desired response was not stimulated, when there was undue environmental influence, when there were excessive individual differences in performance, and when results did not comply with statistical requirements. Statistical selection and placement of items was carried out by determining the age level at which 66 per cent of all subjects passed a task.[10]

Another early scale, mainly utilizing Gesell items, was constructed by Linfert and Hierholzer and published in 1928. The authors report that items were eliminated which failed to show satisfactory correlation between age in months and percentage passing. The resulting test extended from one to twelve months. Age norms were to be calculated from a table indicating percentage of successes on the various tests administered. The test authors

[8] Arnold Gesell and Catherine Armatruda, Developmental Diagnosis (New York, 1941), pp. 6-7.

[9] Charlotte Buehler and H. Hetzer, Testing Children's Development from Birth to School Age (New York, 1935), pp. 25-26.

[10] Ibid., pp. 84-87.

reported that the total point scores indicated a linear increase with age.[11]

Bayley published the California First Year Mental Scale in 1933. Borrowing largely from Gesell items, she also included some tasks from other available infant tests. Tentative scales were administered to 61 infants at monthly intervals starting at approximately one month of age. Items were finally selected on the basis of the following criteria: (a) their occurrence in all or most of the infant subjects; (b) increase in percentage passing from one age level to the next; (c) internal consistency with the total behavioral criterion; (d) their apparent relevance as intellectual or adaptive functions.[12] The resulting scale extended from one to eighteen months with the items placed in ascending order of difficulty by means of the Thurstone Method of Absolute Scaling.[13]

Fillmore published the Iowa Tests for Young Children in 1936, as a result of a study made on 643 children from one to forty months of age. Initially the tests were selected and placed on the basis of percentage passing; tests were discarded when a high percentage passing demonstrated their facility. This method was discarded in favor of Thurstone's Method of Absolute Scaling. Fillmore was the only author surveyed by this writer who carried out tests of item validity using external criteria as part of her standardization procedure. Her procedure and results will be discussed

---

[11]Harriette Linfert and Helen Hierholzer, A Scale for Measuring the Mental Development of Infants During the First Year of Life, Catholic University of America Studies in Psychology and Psychiatry, Vol. I (Washington D.C., 1928), pp. 4-19.

[12]Nancy Bayley, "Mental Growth During the First Three Years," p. 12.

[13]Ibid., p. 50.

below.[14]

Gilliland, in 1949 and 1951, published the Northwestern Infant Intelligence Scale. In their 1949 publication, Gilliland and Shotwell proposed four item criteria for standardization:

(a) Age progression: younger infants should be expected to fail more difficult items upon which older infants succeed; progress in mastery from elementary to more complex stages of a function should parallel age progress.

(b) Variability of success at any one age level: an item should be considered correctly placed if 75 per cent of the infants at a given age level pass it.

(c) Internal consistency of items: each item should agree with total score.

(d) Item validity: correlation of each item with IQ's secured later from other standardized tests.[15]

Published data reveal no evidence that the fourth criterion was ever tested.[16]

The most recent infant test was published in 1954, by the British psychologist, Ruth Griffiths. The test is divided into five separate scales: Locomotor, Personal-Social, Hearing and Speech, Eye and Hand, and

---

[14]Eva Fillmore, Iowa Tests for Young Children, University of Iowa Studies in Child Welfare, Vol. XI, No. 4 (Iowa City, 1936), pp. 11-16.

[15]A.R. Gilliland and Anna Shotwell, "A Preliminary Scale for the Measurement of the Mentality of Infants," Child Develpm., Vol. 14, (September, 1943), p. 168.

[16]Ibid., pp. 167-177; Bledsoe, p. 10; Anastasi, p. 286.

Performance. Each scale yields a separate score although a general quotient may be obtained as an indication of total performance. Griffiths has also drawn from the Gesell Schedules. Items were placed on the basis of percentage passing.[17]

The Cattell Scale, which is the instrument utilized in the writer's study, will be described in detail in the next chapter, however, it is appropriate here to report Cattell's methods of item selection and placement. During the first three years of the Harvard study, which stimulated the construction of the Cattell Scale, Gesell tests were given all children under two years of age.[18] Later the less satisfactory items were eliminated and additional items were included from other sources.[19] Cattell's criteria for item elimination were as follows:

(a) Items which showed insufficient increase in percentage passing from one age group to the next.

(b) Items which increased irregularly in number of passes from age to age, and those which showed plateaus or failed to approach the 100 per cent mark at any age.

(c) Items which were difficult to administer or which required unduly subjective judgment on the part of the examiner.

(d) Items which failed to hold the attention of the child.

(e) Items requiring cumbersome materials.

---

[17] Ruth Griffiths, The Abilities of Babies (New York, 1954), p. 33.

[18] Psyche Cattell, The Measurement of Intelligence in Infants and Young Children (New York, 1940), p. 15.

[19] Ibid., p. 23.

(f) Items thought to come under appreciable influence of home training.

(g) Items which appeared to test control of the large muscles.

(h) Items which seemed to test similar abilities to those already
     tapped at the same age level.

(i) Items at age levels for which an adequate number of more or
     equally satisfactory items were available.[20]

Selected items were placed at age levels according to percentage passing.
In order to insure close comparability of scores with the Stanford-Binet,
certain groups of standardization subjects were retested at three years of
age with Form L of the Revised Stanford-Binet, and the Cattell item place-
ments were adjusted so as to produce approximately the same median IQ as
that later obtained by the same subjects on the Stanford-Binet.[21]

Item Rationale and Factor Analysis

A number of test authors have attempted some rationale of the nature
of the behavior underlying their various test items. Gesell sees growth
as a "continuum complex" of elementary components.[22] He analyses these
components into the four fields of behavior previously enumerated, and his
descriptions of these are perhaps the most vivid to be found in the
literature. The motor area comprises both gross bodily control and fine
muscle coordinations, examples of which include postural reactions, head
balance, sitting, standing, creeping, walking, prehensory approach to an

---

[20]Ibid., p. 27.

[21]Ibid., p. 43.

[22]Arnold Gesell, Infancy and Human Growth (New York, 1928), p. 404.

object, and grasp and manipulation of an object. The adaptive field of
behavior is composed of finer sensorimotor adjustments to objects and
situations such as coordination of eyes and hands in reaching and manipula-
tory activities, the use of the motor equipment appropriately in the solution
of practical tasks, the ability to initiate new adjustments in the presence
of simple problem situations, and resourcefulness. Under the heading of
language are included visible and audible forms of communication: facial
expression, gesture, postural movements, vocalizations, words, phrases,
sentences, mimicry, and the comprehension of the communications of others.
The final division, personal-social behavior, generally includes reactions
to the social culture in which the infant or young child lives: bladder
and bowel control, feeding abilities, sense of property, self-dependence
in play, cooperativeness, and responsiveness to training in social
conventions. As can be seen from the foregoing and as Gesell himself has
pointed out, these areas of behavior are closely related and in many
specifically observed activities of the infant would be found to overlap
one another.[23] For example, a test item requiring the infant to obtain a
ring by pulling on the attached string would require adaptivity but also
considerable fine motor proficiency.

Similarly to Gesell, Buehler and Griffiths have also attempted item
classifications. In Buehler's first presentation of her infant tests, each
item was labeled according to the aspect of development it was supposed to
examine: bodily control, mental ability, manipulation of objects, and social

---

[23]Gesell and Armatruda, Developmental Diagnosis, pp. 4-6.

development.[24] Griffiths' division, as was indicated above, was according to the five categories which make up the five separate scales of her test.

Bayley's studies brought her to the conclusion that mental growth is a composite of a number of maturing functions, a viewpoint similar to that of Gesell. However, rather than considering the functions to be simultaneous or parallel, Bayley believes that they are successive but overlapping to the point that they are not, for the most part separable into discrete classifications. Bayley attempted a number of classifications which were unsatisfactory and ultimately presented two broad categories of behavior, sensorimotor and adaptive. Bayley's research indicated that items of the California Scale measured behavior which was largely sensorimotor during the first six or eight months, while truly adaptive behavior was measured by items beyond that period.[25] In a more recent publication, Bayley has suggested that, if intelligence is a complex of developing functions as her work seems to indicate, then factor analysis of test items may contribute to an understanding of its nature.[26]

Bayley's suggestion has been followed by two studies reported in the published literature. Hofstaetter analyzed Bayley's set of correlations between mental ages from birth to maturity by means of Thurstone's method into three statistically independent factors. It was found that the California Scale undergoes changes during the first three years of life

[24]Charlotte Buehler, The First Year of Life (New York, 1930), p. 195.

[25]Bayley, "Mental Growth During the First Three Years of Life," pp. 53-63.

[26]Bayley, "On the Growth of Intelligence," p. 608.

but remains fairly constant thereafter. The first factor, sensory alertness, predominates up to 20 months and from 40 months on contributes almost nothing. From 20 to 40 months, factor II, persistence, was found to be dominant. From 40 months on the third factor, manipulation of symbols, accounts for most of the variance of intelligence test scores.[27] Hofstaetter's results seem to agree with Bayley's conclusions that the California Scale is largely sensorimotor in character during the first year of life and that adaptive behavior is not significantly measured until after that period.

Another factor analysis was carried out by Nelson and Richards on sixth month Gesell items. The writers intercorrelated seventeen items at this level in the middle range of difficulty (25 per cent to 75 per cent passing) and found that three factors seemed to account for most of the variance in behavior of the six month old infants studied. The factors were designated as testability or halo effect, alertness, and motor ability.[28] As can be seen, these results are comparable to those of Hofstaetter in that the sensorimotor component seems to dominate at this early age.

Predictive Item Validity

Four studies of item analysis have been carried out in attempts to

[27]P.R. Hofstaetter, "The Changing Composition of Intelligence: A Study of Technique," J. genet. Psychol., Vol. 85 (September, 1954), pp. 159-164.

[28]Virginia L. Nelson and T.W. Richards, "Studies in Mental Development: II. Analysis of Abilities Tested at the Age of Six Months by the Gesell Schedules," J. genet. Psychol., Vol. 52 (June, 1958), pp. 327-331.

ascertain predictive efficiency. As part of the standardization procedure
for the Iowa Tests, Fillmore correlated performances on her items with
Stanford-Binet and Kuhlmann IQ's obtained from one to two years after initial
testing. The biserial technique was utilized in processing the data
statistically. Results indicated that successes on the items of the infant
test were not markedly related to the child's later mental status as shown
by his IQ on the Kuhlmann and Binet tests. Fillmore attempted to explain
her results on the basis of a number of considerations:

(a) A single item could be expected to be less valid than a battery
of items well selected.

(b) The criterion IQ's themselves were shown to be inconsistent so that
a selected IQ, as used in the study, may not have been wholly
reliable.

(c) The one to two year lapse of time between infant and criterion
testing may have allowed for some stimulation of mental
development.

(d) The Iowa Tests and the Stanford-Binet and Kuhlmann tests seemed to
be measuring intelligence in different ways so that lack of corre-
lation between them does not necessarily prove lack of validity.

(e) The child's openness to the examiner's verbal stimulation at the
upper age levels could account for better performance at this
point and contribute to low correlations between infant item
performance and later IQ.[29]

[29] Fillmore, pp. 23-31.

Another study by Nelson and Richards investigated the interrelationships between performances on the sixth month items of the Gesell Developmental Schedules and the Merrill-Palmer test at 24 months, and the Stanford-Binet at 36 months. Twenty-three subjects were initially examined at or near their sixth month birthdays and then re-examined at the two ages referred to above. The data obtained was manipulated in several ways: (a) biserial correlations were computed to show the relationships between performances on the sixth month items of the Gesell and the total Schedule score, or "Developmental Quotient"; (b) biserial correlations were computed to show the relationships between performances on the sixth month items and the Merrill-Palmer test administered at 24 months; (c) biserial correlations between the sixth month items and the 1916 revision of the Stanford-Binet, using mental age values at three years of age, were computed. The results of the first manipulation indicated that those items which correlated best with the total number of items passed are of a distinctly more grasping and manipulative nature than are those which correlated least. The authors interpreted these results to mean that, at the sixth month level, Gesell items primarily involve grasping, reaching and manipulative behavior, with such features as posture, locomotion, and perception occupying a secondary position. Correlations between the sixth month Gesell items and 24 month Merrill-Palmer performances indicated a tendency for items designated as "awareness of distance perception" by the authors to increase in their relative rank of importance, while motor items tended to decrease. The highest coefficient obtained by this manipulation (.58) was deemed somewhat low in that its forecasting efficiency was computed to be 19 per cent. The authors

conclude that the sixth month Gesell items may not be considered to measure more than 50 per cent of that which the Merrill-Palmer measures at two years. As in the case of the Merrill-Palmer correlations, the strictly motor items of the Gesell tended to have lower correlations with the Binet than those designated as "awareness of distance perception". Other results of the third manipulation suggested that as much as 60 per cent of the variance in mental age at three years is sampled by the sixth month Gesell items. Results of multiple correlation procedures between the Gesell items and the Binet indicated that some items were better predictors of later intellectual status than the total schedule. When these items (five in number which correlated about .54 with the Binet mental ages) were combined, a correlation of .80 resulted.[30]

Anderson attempted to construct a more predictively efficient infant scale by means of the item selection method using five year old mental age status on the Stanford-Binet (1916 revision) as the criterion. Items initiated by Gesell, Buehler, Linfert and Hierholzer, as well as some devised by the author of the study were administered. A total number of 91 infants were examined by this improvised scale at the ages of 3, 6, 9, 12, 18, and 24 months; the Stanford-Binet was administered at five years of age. The criterion data was then dichotomized into two groups: those obtaining the 15 highest and 15 lowest Stanford-Binet IQ's. Each item of the infant scale was examined with respect to the number passing and failing in each of the dichotomized groups at each age. The most diagnostic items were then

[30]Virginia L. Nelson and T.W. Richards, "Studies in Mental Development: I. Performance on Gesell Items at Six Months and its Predictive Value for Performance on Mental Tests at Two and Three Years," J. genet. Psychol., Vol. 52 (June, 1938), pp. 303-325.

selected on this basis and correlated with five year test status. Odd-
even reliability coefficients were also calculated. Less significant items
were then added until the largest predictive coefficients and highest odd-
even reliabilities were secured. It was found that the number of items at
each age which had some significance when checked against the criterion
groups was low. A marked characteristic of the discriptive items at the
earlier age levels (three and six months) was that they involved alertness
to external environmental stimulation. The author summarized the study by
stating that the item analysis technique did effect more predictive
efficiency.[31]

The most recent study to be found in the literature involving item
analysis was reported by Bayley. In a preliminary study of the items of
the California First Year Mental Scale, six children were selected at each
extreme of ability as measured by the 14 to 16 year California tests. It
was noted the age at which each of these twelve children passed each of the
first year items. Results indicated that the high scoring teenagers, as
infants, had passed 31 items two months or more in advance of the low
scorers. Bayley states that most of these items occurred in the second
half of the year and that they were an "odd assortment" revealing little
rationale for their discriminative quality. The scores for the total
Berkeley Growth Study sample were then computed on the 31 item scale for
ages six, nine, and twelve months. The correlations of these scores with

[31]L. D. Anderson, "The Predictive Value of Infancy Tests in Relation
to Intelligence at Five Years," Child Develpm., Vol. 10 (September, 1939),
pp. 203-212.

the mean intelligence sigma scores at ages 16, 17, and 18 years were .09
at six months, .32 at nine months, and .30 at twelve months. None of the
correlations were significant.[32]

It is noteworthy that the above studies vary in their optimism in
regard to the efficacy of item analysis. The results of Bayley and Fill-
more seem to indicate that little can be contributed to the overall validity
of infant testing as a result of the low correlations obtained between the
individual items and criterion performance. On the other hand, the results
of Nelson and Richards, and Anderson seem more optimistic particularly when
a number of items with relatively high correlations are combined when
correlated with the criterion. In view of the apparent conflict, a discus-
sion of this writer's results in comparison to the above would seem warranted
and will be attempted in Chapter IV.

---

[32]Bayley, "Growth of Intelligence," p. 807.

CHAPTER III

PROCEDURE

The Cattell Infant Intelligence Scale[1]

The Cattell Infant Intelligence Scale was published in 1940, as one
result of a longitudinal study of child development conducted at Harvard
University. Items used by Cattell were adapted largely from those included
in Buehler's infant tests and the Gesell Developmental Schedules. Stanford-
Binet items were interspersed from the twenty-second to the thirtieth month
levels. Cattell presents her test as a downward extension of the Stanford-
Binet and, as was previously pointed out, Cattell Scale items were rearranged
on the basis of a retest study so that the median IQ's for each age level
were as close as possible to the median IQ of 106 obtained on the Stanford-
Binet at 36 months.

The scale extends from 2 to 30 months with five items and one or two
alternates being included in each age level. The age levels are placed
one month apart for the first year and two months apart thereafter.
Administration and scoring are similar to the Stanford-Binet except that
serial testing is permitted in order to secure the interest and cooperation

_____

[1] Information given under this heading is from Psyche Cattell, The
Measurement of Intelligence of Infants and Young Children (New York, 1940),
pp. 11-49.

of the infant. A basal age is established at the level at which successes are obtained on all five items and testing is continued up to the level at which all five items are failed. In computing the mental age, additional credits are added to the basal age. Since there are five items placed one month apart throughout the first year level, two-tenths of a months credit is given for each success beyond the basal age. The chronological age of the subject is also estimated in tenths of months and the IQ may be computed in the conventional manner.

After 1346 examinations of 274 children were administered, items were allocated to age levels on the basis of percentage passing. The subjects included in the standardization sample were examined at the ages of 3, 6, 9, 12, 18, 24, and 30 months. While each was not examined at every age level, the subjects averaged five examinations each. Items were allocated to age levels between the standardization ages (2, 4, 5, 7, 8, 10, and 11 months during the first year) on the basis of percentage passing at the adjacent standardized age levels. Retests by Cattell indicated close agreement between the IQ obtained at the ages at which the test was standardized and the between ages which were estimated.

Cattell reported that the criteria for enrollment in the Harvard study, as well as the retest Stanford-Binet mean IQ, suggest that her sample was socially and intellectually somewhat above the general population. Enrollment requirements included good physical health and normal delivery of the infant, primarily North European national origin of the parents, more or less permanent employment of the father, and the mother's willingness to cooperate with the demands of the study over a period of years. While a

few parents had professional backgrounds, most were employed in such positions as policeman, clerk, and storekeeper.

Using the prediction of 36 month Stanford-Binet scores as a criterion, Cattell found that her scale becomes increasingly more valid after 12 months and less valid prior to that age. The validity coefficient at 6 months was .10, at 9 months it was .34, at 12 months .56, at 18 months .67, and at 24 months .71. Odd-even reliability coefficients were reported by Cattell to be .56 at 3 months, .88 at 6 months, .86 at 9 months, .89 at 12 months, .90 at 18 months, .85 at 24 months, and .71 at 30 months.

Selection and Description of the Subjects[2]

The data utilized in the writer's investigation was gathered as a result of a validity study of the Cattell Infant Intelligence Scale completed by Patricia Bledsoe in 1955. The subjects were infants who had been placed for adoption during the first month of life from Saint Vincent's Infant Asylum and Misericordia Hospital, two maternity and infant homes which are conducted under the auspices of Catholic Charities of the Archdiocese of Chicago. Agency policy in regard to adoptive infants requires a six months probationary period during which time the infants are in the adoptive homes under the supervision of a social worker. Approximately one month prior to the completion of legal adoption the infants are brought to Catholic Charities Guidance Center for a psychological examination. Bledsoe's records were available as a result of this procedure.

Three phases of sample selection were reported in Bledsoe's study.

---

[2]Information given under this heading is from Bledsoe, pp. 33-36.

From the several hundred Cattell Scale records available, all administered
by the same psychologist, Bledsoe's general aim was to obtain a sampling of
valid tests administered at the same age level. Furthermore, she attempted
to obtain a sample which was representative of all children placed by the
agency in early infancy. Accordingly, selection was initially restricted to
the records of infants who had been tested with the Cattell Scale between
January 1, 1950, and June 30, 1952. This method allowed two previous years
of supervised infant testing experience by the administering psychologist.
The second step involved examining the 1950-52 records according to the
following criteria: (a) examinations within one week of the sixth month
birthdays of the infants, (b) reasonable indication, in the light of the
examiner's opinion and the test behavior as recorded in the report accompa-
nying each test record, that the examination was valid, (c) placement in
the adoptive home during the first month of life, (d) full term gestation.
The records of 158 infants were found to conform to the above requirements.

The final phase of selection involved securing the cooperation of the
adoptive families in returning their children for retesting with the Stanford-
Binet. The children were, at the time of retesting, between three and one-
half to six years of age. Bledsoe reported that 115 families responded to
a letter explaining the project and requesting their cooperation, 20 addi-
tional families were contacted by telephone, and the remaining 23 families
could not be located. Of the 135 families contacted, 110 eventually
brought their children to the clinic for retesting. Bledsoe indicated that
12 families stated that they were unable to return due to a number of reasons,
and 5 responded that they did not wish to participate in the study.

Following the selection of the sample, the Revised Stanford-Binet, Form L, was administered by Bledsoe to 107 children; three examinations were given by two other psychologists of the Guidance Center Staff. In each case the tests were administered and scored in accordance with the directions outlined in the manual, and all scoring was checked by a psychologist other than the administrator. The Cattell Scales were also administered in accord with the requirements outlined in the Cattell manual. In order to provide a constant physical environment, both Binet and Cattell examinations were administered to all subjects in the offices of the Guidance Center.

Bledsoe ascertained that her sample of 110 children is representative of all the children placed in adoption in early infancy by the Chicago Catholic Charities. She stated that the group is also probably representative of the children placed in infancy by most adoption agencies in large urban communities. As a result, however, of strict agency policies in regard to adoptive infants, it was suggested that the study group is somewhat select as compared to 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 back-
> ground incidence of mental illness. Second, although almost
> all of the children in the study group were born out of wed-
> lock, 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 average. According to the data on the educational
> status of the adoptive fathers, this group of adoptive
> parents surpassed the averages for the country as a whole
> in these areas. Agency standards for the adoptive homes are
> high and investigations 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.[3]

Bledsoe reported the mean retest age of the sample to be four years, eight months, with a range from three years, one month to five years, eleven months. Within the range of five years to five years, eleven months were 41 children; 50 children were between the ages of four years and four years, eleven months; 19 children were between the ages of three years and three years, eleven months. Of the 110 subjects, one was Negro and the remainder were caucasion. No serious health problems were noted in the case histories of the children. One boy was reported to have had a mild congenital heart murmer and one girl had a visual defect which required glasses.

No siblings were reported in the families of 33 subjects; 67 children had one sibling who was in each case also adopted. Of the latter group, 29 subjects had an older brother or sister, and 38 had younger siblings. One child had five brothers and sisters who were either adopted or placed with the family on a boarding care basis. In 9 cases the siblings were the natural children of the adoptive parents.

The educational backgrounds of the sample were in all cases on the preschool level. In the five to six year age group, 18 children had some kindergarten experience; the average attendance was reported to have been about four months. No consistent nursery school experience was indicated for any of the younger children.

------

[3]Ibid., p. 35.

Description of True and Adoptive Parents[4]

Information regarding the true parents of the study group was obtained by Bledsoe from the case records compiled by agency caseworkers while the true mothers were awaiting delivery of the infants. In a few cases the information given in the case records was incomplete and it was pointed out that information gained about the true fathers was second-hand and therefore may be of limited reliability.

The mean age of the true mothers at the time of birth of the subjects was 24 years, 4 months; ages ranged from 16 to 41 years. The mean age of the true fathers at the time of birth of the subjects was 28 years, 6 months; true father ages ranged from 18 to 50 years.

The educational status of 101 true mothers was reported in the case records. The grades reported in the records seemed, as far as could be determined, to represent grades completed. The mean grade completed was 11.43 years, the median was 12.14 years, and the standard deviation was 1.74. Nineteen of the true mothers had continued their education beyond the high school level: one was a college graduate and five were graduate nurses. Of the remaining 82 mothers, 44 were high school graduates, 29 had had 9, 10, or 11 years of schooling, 8 girls had completed grammar school, and one girl had completed seventh grade.

The educational backgrounds of the true fathers were obtainable in 72 cases. The mean grade completed was 12.38 years, the median grade completed was 12.63 years, and the standard deviation was 2.33. It was reported that

[4] Information given under this heading is from Bledsoe, pp. 36-40.

18 true fathers had continued their educations beyond the high school level;
8 were college graduates and 5 had received some graduate training. Forty-
three of the true fathers were high school graduates, 8 had had 9, 10, or 11
years of schooling, 2 were grammar school graduates, and one had had 5
years of schooling.

Bledsoe reported that the occupational status of the true mothers
approximated that expected in the general population for women of about
24 years of age. The majority of the true fathers were classified as semi-
skilled or slightly skilled laborers. Forty-five per cent fell into the
professional, semi-professional, and clerical categories.

It was stated in Bledsoe's study that background information pertinent
to the adoptive parents was obtained from the case records of the Catholic
Home Bureau, the adoption agency of the Chicago Catholic Charities. The
occupational status of the adoptive fathers was reported to have been
verified at the time of retesting.

The mean age of the 110 adoptive mothers at the time of retesting was
37 years, 6 months, with a range in years from 30 to 46. At the time the
infants were placed, the adoptive mothers were in a mean age range from
31 to 34 years. While this is several years in advance of the mean maternity
age, Bledsoe points out that it is probably typical of an adoptive mother
population.

At the time of retesting, the mean age of the adoptive fathers was
39 years, 4 months, with a range from 30 to 54 years. The group ranged
from 33 to 36 years when the children were placed for adoption as infants.

The mean grade completed by the 110 adoptive mothers was 11.9 years,
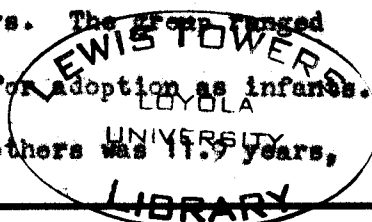
the median grade was 12.48, and the standard deviation was 2.21. Thirty-five of the adoptive mothers were educated beyond the high school level; 10 were college graduates, and one had some graduate training. Of the remainder of the group, 59 were high school graduates, 25 had completed 9, 10, or 11 years of schooling, and 11 had graduated from grammar school.

The mean grade completed by the adoptive fathers was 12.75, the median was 12.80, and the standard deviation was 2.59. Some education beyond the high school level was received by 48 of the adoptive fathers; 18 were college graduates, and 7 had graduate training. Thirty-five were high school graduates, 21 had 9, 10, or 11 years of schooling, 5 had completed 8 grades, and one had completed 7 grades.

Few occupations were reported for the adoptive mothers, probably due to the fact that most of them were functioning as housewives at the time of retesting. The adoptive fathers were reported to be superior in occupational and social status to the true fathers and to the general employed male population. Seventy-one per cent were classified in professional, semi-professional, and clerical groupings. None were reported to be in the unskilled working group. The reader can refer to Tables 4 and 5 in the appendix in order to compare in detail the educational and occupational status of the true and adoptive fathers. Summarily, it may be stated that, while the educational backgrounds of the two groups were similar, the adoptive fathers were superior to the true fathers in occupational status.

Statistical Procedure

Bledsoe's data was used by the writer to compile data sheets for each of the 110 subjects investigated. The data sheets contained chronological

ages, mental ages, and IQ's resulting from the administrations of the Cattell Scales and Stanford-Binets. The subject's performance on each item of the Cattell Scale was also recorded just as it was marked on the record by the administering psychologist. Each data sheet was then compared to the original record by the writer and an assistant in order to insure accuracy.

Each item of the fifth, sixth, seventh, and eighth month levels was subsequently analyzed in terms of percentage of subjects passing and failing it. While a few subjects were administered the third, fourth, tenth, and eleventh month items, the number of cases was considered too small to submit to statistical analysis: 4 cases were administered third month items, 19 were administered fourth month items, 15 subjects were administered items at the tenth month level, and only 2 cases were included at the eleventh month level.

An item was considered to have been passed by the subject when the record indicated a plus sign after it. Items were considered failed when either a minus or a question mark was recorded. Question marks were considered to denote failures due to the fact that such items were consistently given no credit in the computation of the mental age. In a few cases, there were no marks to indicate the subject's performance on an item, even though the item may have occurred prior to the maximal age level. In such cases, the subject's performance on that particular item was not included in the statistical analysis because it was impossible to ascertain whether the item had been omitted in administration or whether the examiner had forgotten to record it. In no case of this kind did the omission of an item make the mental age questionable, for the examiner credited only those

items marked with a plus sign.

Biserial correlations were computed between Cattell Scale items and Stanford-Binet IQ's. In computing the biserial correlations, Cattell Scale performance was treated as a dichotomous variable on a pass-fail basis, and Stanford-Binet performance was treated as a continuous variable. Only those Cattell items which 90 per cent of the subjects passed or failed were included in this analysis. It is pointed out by McNemar that one can place little confidence in a biserial correlation when the dichotomies are more extreme than .10 or .90.[5] In such situations, the resulting correlations may be considered to be more an indication of the difficulty level of the item than of its predictive validity.

As a final procedure, the sampling errors of all but the minus biserial correlations were computed. The null hypothesis, that no correlation exists, was then tested by reference to the normal curve, as is indicated when samples are as large as that used by this investigation.[6]

---

[5] Quinn McNemar, _Psychological Statistics_ (New York, 1955), 2'ond ed., p. 195.

[6] _Ibid._, p. 195.

CHAPTER IV

RESULTS

The sample group, consisting of 110 adopted children, obtained a mean IQ of 114.3 when tested at six months of age with the Cattell Infant Intelligence Scale. The standard deviation at this age level was 8.9, and the IQ's ranged from 84 to 135. When the sample was retested with the Stanford-Binet at ages ranging from 3 years, 1 month to 5 years, 10 months, the mean IQ was 115.34 and the standard deviation was 11.8. The retest IQ's ranged from 85 to 145. Both the Cattell mean IQ and the Stanford-Binet mean IQ obtained from this sample are somewhat above the level expected from the general population. Such a result is understandable in the light of the rigid selection procedures carried out by the adoption agency which placed the infants in their adoptive homes. The higher than average mean IQ's indicated in this study are in conformity with the findings of other investigators who have studied adoptive children.[1]

The percentages of children passing each Cattell Scale item from the fifth to the ninth month levels are tabulated on Table 1. It can be seen from this presentation that all items on the fifth month level and items

---

[1]Patricia Bledsoe, "The Predictive Value of the Cattell Infant Intelligence Scale," Unpublished Doctoral Dissertation (Loyola University, Chicago, 1955), p. 42.

one, two, and five on the sixth month level were too easy for this sample in that 100 per cent, 99 per cent, and 91 per cent of the group passed them. This seems to be explained by the higher than average mean IQ obtained on the Cattell Scale. Had the mean IQ been around 100, one would have expected the percentage passing to have been lower on these items.

When these results are compared to those of Cattell with six month infants, it can be seen that the items were generally easier for the writer's sample. A comparison of central tendencies between the two studies suggests that the writer's group was somewhat superior to the Cattell sample, as well as to the general population. Cattell reported a median IQ of 108 at six months, no mean IQ was available for her group.[2]

Biserial correlations between Cattell items which less than 90 per cent of the infants passed or failed are presented on Table 2. Of the eleven items which survived the 90 per cent cut-off point, six resulted in positive correlations ranging from .25 to .01, and only one of these, a language item placed at the eighth month level, could be expected to predict later intellectual status better than chance. The remaining items resulted in negative correlations ranging from -.02 to -.16.

While the positive correlations indicate a very minor relationship to the criterion, there seems to be a trend suggesting slightly greater conformity between test and retest results as higher age levels are approached. It is notable in this respect that the only significant item was placed at the eighth month level. This tendency conforms to the results

---

[2]Cattell, p. 30.

Table 1

Percentage of Subjects Passing Cattell Items

Item	Writer's study		Cattell's study[a]	
	N	% passing	N	% passing
5 months				
Turns to bell	77	99	172	89
Attains ring	77	96	167	89
Transfers object from hand to hand	78	96	68	87
Regards pellet	77	95	176	84
Picks up spoon	77	100	66	85
6 months				
Attains cube	110	100	181	79
Lifts cup	110	99	177	79
Manipulates mirror	110	87	167	80
Unilateral reaching	108	85	174	63
Approaches second cube	110	91	26	62
7 months				
Attempts pellet	110	85	149	58
Pats and smiles at mirror	109	62	163	50
Inspects ring	110	57	84	49
Takes two cubes	110	82	66	45
Exploits paper	109	83	63	40
8 months				
Pulls ring by string	105	17	163	06
Interested in string	106	56	79	20
Says "mama" or "dada"	105	23	180	09
Rakes pellet	104	31	175	10
Interested in details of bell	106	00	63	01

Table 1 continued

Percentage of Subjects Passing Cattell Items

Item	Writer's study		Cattell's study[a]	
	N	% passing	N	% passing
9 months				
Scissors grasp of pellet	86	01	175	00
Looks for spoon	77	06	170	09
Rings bell	70	06	76	04
Adjusts to gestures	81	00	76	03
Adjusts to words	72	03	76	00

[a] Cattell, pp. 50-54.

Table 2

Correlations Between Cattell Item Performance

and Later Stanford-Binet IQ's

Item	N	rb	P
6 months			
Manipulates mirror	110	-.16	——
Unilateral reaching	108	.08	.29
7 months			
Attempts pellet	110	-.02	——
Pats and smiles at mirror	109	-.01	——
Inspects ring	110	-.05	——
Takes two cubes	110	.08	.29
Exploits paper	109	.11	.25
8 months			
Pulls ring by string	105	-.08	——
Interest in string	106	.02	.44
Says "mama" or "dada"	105	.25[a]	.008
Rakes pellet	104	.11	.19

[a]Significant at better than the .01 level.

of Bayley, who, in her item analysis found that the most diagnostic items
occurred during the second half of the first year.[3]

The singular significance of the language item is in agreement with the
findings of Anderson, whose correlational analysis and analysis of signif-
icant items reflected the major role played by the progress of language
development in the prediction of later intellectual status.[4]

Another conclusion of Anderson, as well as the findings of Nelson and
Richards, are however, not born out by the writer's investigation. Anderson
reported that items which involved the infant's positive reaction to
immediate environmental happenings with a beginning interest in contac-
tual exploration of objects, were also important as early indicators of
intellectual development.[5] Nelson and Richards found that items involving
a more or less perceptual reaction to specific stimulus correlated better
with Merrill-Palmer and Stanford-Binet results than items defined as motor
and postural-locomotor.[6] The writer's results are more in agreement with
those of Bayley who found little rationale behind the few significant
items her study revealed.[7] In some cases in the writer's analysis where two
items could be expected to measure the same aspect of behavior, one
correlated positively and the other negatively with later Stanford-Binet

---

[3]Bayley, "On the Growth of Intelligence," p. 807.

[4]Anderson, p. 207.

[5]Ibid., p. 207.

[6]Nelson and Richards, "Performance on Gesell Items at Six Months,"
p. 321.

[7]Bayley, "On the Growth of Intelligence," p. 807.

performance. The difference in correlations between the two items which involve the pellet will serve as an example: "rakes pellet" correlated .11, while "attempts pellet" correlated -.02. with the criterion.

A comparison of the awareness items which Nelson and Richards found to be good predictors to similar Cattell items suggests further discrepancies. Nelson and Richards' "reacts to mirror image", which correlated .30 with later Stanford-Binet performance, and "regards pellet", which correlated .32, are similar to three Cattell items studied by the writer: "attempts pellet", "manipulates mirror", and "pats and smiles at mirror".[8] It can be seen from Table 2 that the Cattell items, which correlated -.02, -.16, and -.02 respectively, are significantly lower than the results reported by Nelson and Richards.

The conflicting findings of these studies are somewhat difficult to evaluate due to the fact that different tests with different administration and scoring procedures were used. The conflict may be due to the latter. In the case of Nelson and Richards' investigation, it may be due to the fact that their sample was bimodal in character.[9] In such cases, the point biserial technique may have yielded lower quantities. It should be noted in relation to Anderson's study that the criterion for significant items was not, as far as could be determined, a statistical one. Therefore, re-evaluation on a statistical basis could reveal somewhat different results.

---

[8] Nelson and Richards, "Performance on Gesell Items at Six Months," p. 320.

[9] Ibid., p. 818.

While Anderson was able to find a number of items at earlier age levels (3, 6, and 9 months) which could be combined by the multiple correlation technique into a more predictive scale, such a procedure does not look promising for the Cattell Infant Intelligence Scale at the earlier age levels because of the numerous low positive and negative correlations. Anderson's results are explainable because he began his investigation from a relatively large item pool contributed to by a number of available infant tests. Nelson and Richards also attempted the multiple correlation technique with six month Gesell items with apparent success. However, one would question whether an adequate number of highly correlated items would have been available if the point biserial procedure had been used in their study.

CHAPTER V

SUMMARY AND CONCLUSIONS

The investigation of infant mental development has long been of concern to psychologists. This study derives its importance not only from scientific curiosity concerning the genesis of those mental processes which are manifest in the child and the adult, but also as a result of the relatively wide practice of infant intelligence testing as an aid in the selection and placement of adoptive children.

Since the first efforts of Kuhlmann, Gesell, and Buehler to devise infant tests, the question of the predictive validity of these scales and those which have followed has become a point of extensive investigation and considerable disagreement. The conflict has resolved itself into two opposite camps led primarily by Nancy Bayley and Arnold Gesell. As a result of her participation in the Berkeley Growth Studies, Bayley has come to the conclusion that the satisfactory prediction of later intellectual status as a result of infant tests is impossible due to the extensive organizational changes which take place during the early phases of intellectual development. She regards infant tests to be valuable primarily as aids to a gross evaluation of developmental normality and abnormality, and for research purposes.

Gesell's studies at the Yale Clinic have brought him to a different conclusion. He maintains that, in spite of its bewildering complexity, intellectual growth is an orderly and lawful progression of patterned processes. Because of its lawful nature, mental development is amenable to diagnosis and prediction providing the total life picture of the infant can be assessed by the examiner, preferably over a period of time.

In spite of Gesell's optimism, the results of validity studies of the various infant tests have been disappointing. In efforts to investigate as many avenues of approach as possible, some investigators have turned their attention from correlating the total score with a criterion to an analysis of the individual items which make up the tests. Results of some of these attempts have been somewhat more encouraging.

While the Cattell Infant Intelligence Scale, published in 1940, is widely used in clinical practice, it has never, to the knowledge of the writer, been investigated by means of item analysis. Accordingly, the purpose of the writer's study has been to investigate the extent of agreement between the performance of 110 adoptive infants on the individual Cattell items and their later total test performance on the Revised Stanford-Binet, Form, L. Biserial correlations were computed from the infants' performances on the Cattell item and their later Stanford-Binet IQ's. The item performances of the subjects were dichotomized on a pass-fail basis, while Stanford-Binet performances were treated as a continuous variable. The resulting biserial correlations were submitted to the appropriate test of statistical significance.

The results of the writer's study indicate only one item, a language

item placed at the eighth month level which was significantly predictive
of later Stanford-Binet status. There was also some tendency for positive
correlations to occur more often on the seventh and eighth month levels than
on the sixth month level. However, because of the preponderance of
insignificant relationships, it is difficult to draw any confident conclu-
sions from this tendency.

The writer's study was in agreement with the findings of Bayley in
that there seemed to be little explanation why some items resulted in
positive correlations and others in negative correlations. The findings of
Nelson and Richards, and Anderson, that certain types of behavior such as
awareness of distance perception were more significant as predictors were
not confirmed by this investigation. The results of the writer's study also
cast some doubts on the feasibility of using a multiple correlation
technique with the early age levels of the Cattell Scale in order to com-
bine items into a more adequately predictive test.

On the basis of the writer's findings, Bayley's pessimism in regard to
the impossibility of adequate prediction of later intellectual status by
means of tests administered in early infancy seems to be greatly streng-
thened. One could conclude that language seems to be the first clue to the
extent of intellectual potential which could be expected to actualize as
the infant develops. It seems probable that other clues must remain out
of the reach of the tester until language has developed to a more
appreciable degree than the first utterances of "mama" and "dada".

If further research involving the evaluation and prediction of
mental ability in infancy is to take place, it seems valid to conclude

that the psychometric approach cannot offer a singular satisfactory method. It was noted in the survey of the literature that some investigators have found more comprehensive clinical techniques to be of value. More studies utilizing developmental and environmental histories together with repeated observations by means of tests may realize fruitful results. Since this is the practice which is used most in clinics, it would seem valuable to submit it more extensively to the rigors of scientific investigation in order to evaluate its efficacy.

BIBLIOGRAPHY

I. PRIMARY SOURCES

A. BOOKS

Buehler, Charlotte. The First Year of Life. New York, 1930.

Buehler, Charlotte, and Hildegard Hetzer. Testing Children's Development from Birth to School age. New York, 1935.

Cattell, Payche. The Measurement of Intelligence of Infants and Young Children. New York, 1940.

Gesell, A. The Mental Growth of the Preschool Child. New York, 1926.

————. Infancy and Human Growth. New York, 1928.

————. Infant Development: The Embryology of Early Human Behavior. New York, 1952.

Gesell, A., and Catherine Armatruda. Developmental Diagnosis. 2'ond ed. New York, 1941.

Griffiths, Ruth. The Abilities of Babies: A Study of Mental Measurement. New York, 1954.

Kuhlmann, F. Tests of Mental Development. Minneapolis, 1939.

Mc Nemar, Q. Psychological Statistics. 2'ond ed. New York, 1955.

B. ARTICLES

Anderson, L. D. "The Predictive Value of Infancy Tests in Relation to Intelligence at Five Years," Child Development, V (September 1939), 302-312.

Bayley, Nancy. "Mental Growth in Young Children," 39'th Yearbook of the National Society for Studies in Education. XXXIX, Part II (February

1940), 11-79.

————. "Mental Growth During the First Three Years," Genetic Psychological Monographs, XIV (July 1933), 1-92.

————. "Consistency and Variability in the Growth of Intelligence from Birth to Eighteen Years," Journal of Genetic Psychology. LXXV (December 1949), 165-196.

————. "On the Growth of Intelligence," American Psychologist. X (December 1955), 805-818.

————. "Value and Limitations of Infant Testing," Children. V (July-August 1958), 129-133.

Fillmore, Eva. Iowa Tests for Young Children. University of Iowa Studies in Child Welfare, Vol. XI. Iowa City, Iowa, 1936.

Gesell, A. "The Stability of Mental Growth Careers," 39'th Yearbook of the National Society for Studies in Education. XXXIX, Part II (February 1940), 149-160.

Gilliland, A., and Anna Shotwell. "A Preliminary Scale for the Measurement of the Mentality of Infants," Child Development. XIV (September 1943), 167-177.

Hofstaetter, p. "Changing Composition of Intelligence: A study of the T Technique," Journal of Genetic Psychology. XXCV (September 1954), 159-164.

Linfert, Harriette, and Helen Hierholzer. A Scale for Measuring the Mental Development of Infants in the First Year of Life. Catholic University Studies in Psychology and Psychiatry, Vol. I. Washington D. C. , 1928.

Nelson, Virginia, and T. Richards. "Studies in Mental Development: I. Performance on Gesell Items at Six Months and its Predictive Value for Performance on Mental Tests at Two and Three Years," Journal of Genetic Psychology. LII (June 1938), 303-325.

————. "Studies in Mental Development: II. Analysis of Abilities Tested at the Age of Six Months by the Gesell Schedules," Journal of Genetic Psychology. LII (June 1938), 327-331.

Simon, A., and Libby Bass. "Toward a Validation of Infant Testing," American Journal of Orthopsychiatry. XXVI (April 1956), 340-350.

C. UNPUBLISHED MATERIAL

Bledsoe, Patricia. "The Predictive Validity of the Cattell Infant Intelligence Scale." Unpublished Doctoral Dissertation. Loyola University, Chicago, 1955.

II. SECONDARY SOURCES

Anastasi, Ann. Psychological Testing. New York, 1957.

Table 3

Retest Correlations from Published Validity Studies of

Six-Month-Old Infant Test Performances[a]

Investigator	Test	N	Criterion	Retest Age	r
Furfey and Muhlenbein	L-H Scale	42	S-B (1916)	4 yrs. 2 mos. (median)	-.20
Fillmore	Iowa Tests	?	S-B (1916)	2 or 3 yrs.	.32
Anderson	Buehler & Gesell items	91	S-B (1916)	5 yrs.	.08
Nelson and Richards	Gesell	31	S-B (1916)	36 mos. (MA)	.47
Bayley	California Scale	61	Cal Preschool Scale	27, 30, & 36 mos.	.10
Cattell	Cattell Scale	42	S-B (1937)	36 mos.	.30

[a]Patricia Bledsoe, "The Predictive Value of the Cattell Infant Intelligence Scale," Unpublished Doctoral Dissertation (Loyola University, Chicago, 1955), p. 18.

Table 4

Minnesota Occupational Scale Distribution of

110 Adoptive Fathers and 85 True Fathers[b]

Group	True Fathers Number	Per cent	Adoptive Fathers Number	Per cent	U.S. Males Per cent (1950 census)
I: Professional	3	4	10	9	8
II: Semi-professional, managerial	9	11	24	22	11
III: Clerical, retail business, skilled	24	29	44	40	18
IV: Farmers	2	2	---	---	10
V: Semi-skilled	18	22	26	24	29
VI: Slightly skilled	26	31	6	5	10
VII: Laborers	1	1	---	---	14
Totals	85	100	110	100	100

[b]_Ibid._, p. 39.

Table 5

Educational Status of the True Parents and

the Adoptive Parents[c]

Years of Schooling	True parents				Adoptive parents			
	Father		Mother		Father		Mother	
	Number	Per cent	Number	Per cent	Number	Per cent	Number	Per cent
	72	100	101	100	110	100	110	100
17 - 20	5	4	—	—	7	6	1	1
13 - 16	15	21	19	19	41	37	34	31
9 - 12	51	71	73	72	56	51	64	58
5 - 8	3	4	9	9	6	6	11	10
Mean	12.44		11.43		12.75		11.90	
Median	12.63		12.14		12.80		12.48	
Standard deviation	2.33		1.74		2.59		2.21	

[c]Ibid., p. 40.

APPROVAL SHEET

The thesis submitted by Mary Lou Strassmaier has been read and approved by a board of three members of the Department of Psychology.

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

The thesis is therefore accepted in partial fulfillment of the requirements for the Degree of Master of Arts.

May 29, 1961
_____
Date

Frank J Kobler
_____
Signature of Adviser