THE
TWENTY-SEVENTH YEARBOOK

OF THE

NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

NATURE AND NURTURE

PART I

THEIR INFLUENCE UPON INTELLIGENCE

Prepared by the Society's Committee
W. C. Bagley, B. T. Baldwin, C. C. Brigham, F. N. Freeman, Rudolf Pintner, G. M. Whipple, and L. M. Terman (*Chairman*)

Assisted by
Nettie L. Berg, Helen R. Bobo, Barbara S. Burks, Mildred Burlingame, Mary L. Casey, Helen P. Davidson, Lois Doe-Kulmann, Dorothy Durling, A. I. Gates, Arnold Gesell, Florence Goodenough, Katherine Greene, Mattie C. Hardy, Doris I. Harter, Gertrude Hildreth, Carolyn Hoefer, K. J. Holzinger, Harold Jones, T. L. Kelley, Clara Lorenzen, Bertha Luckey, Helen L. Koch, Doris Maddow, Janet Matthew, Katharine McBride, Blythe C. Mitchell, Katharine Murdoch, Joseph Peterson, Agnes Rogers, C. P. Stone, Gladys G. Tallman, E. L. Thorndike, and R. R. Willoughby

Edited by
GUY MONTROSE WHIPPLE

This Yearbook Will Be Discussed at the Boston Meeting of the National Society, February 25 and 28, 1928, 8:00 P. M.

PUBLIC SCHOOL PUBLISHING COMPANY
BLOOMINGTON, ILLINOIS
1928

OFFICERS OF THE SOCIETY
for 1927-8

Board of Directors
(Term of office expires March 1st of the year indicated)

B. R. BUCKINGHAM (1929)
Ohio State University, Columbus, Ohio

WERRETT W. CHARTERS (1928)
University of Chicago, Chicago, Illinois

ERNEST HORN (1930)
State University of Iowa, Iowa City, Iowa

CHARLES H. JUDD (1928)
University of Chicago, Chicago, Illinois

LEONARD V. KOOS, *Chairman* (1930)
University of Minnesota, Minneapolis, Minnesota

HAROLD RUGG (1929)
Teachers College, Columbia University, New York City

GUY MONTROSE WHIPPLE (*Ex-officio*)
Danvers, Massachusetts

Secretary-Treasurer
GUY MONTROSE WHIPPLE (1929)
Danvers, Massachusetts

MEMBERSHIP
OF THE SOCIETY'S YEARBOOK COMMITTEE
ON
THE POSSIBILITIES AND LIMITATIONS OF TRAINING

WILLIAM C. BAGLEY, Professor of Education, Teachers College, Columbia University, New York City, New York

BIRD T. BALDWIN, Research Professor in Educational Psychology, State University of Iowa, Iowa City, Iowa

CARL C. BRIGHAM, Associate Professor of Psychology, Princeton University, Princeton, New Jersey

FRANK N. FREEMAN, Professor of Educational Psychology, The University of Chicago, Chicago, Illinois

RUDOLF PINTNER, Professor of Educational Psychology, Teachers College, Columbia University, New York City, New York

LEWIS M. TERMAN (*Chairman*), Professor of Psychology, Stanford University, Palo Alto, California

GUY M. WHIPPLE, Director of the National Intelligence Tests, and Editor of the Society's Yearbooks, Danvers, Massachusetts

TABLE OF CONTENTS

EDITOR'S PREFACE

The Yearbook on Nature and Nurture has grown out of a suggestion made by Dr. L. M. Terman in December, 1923, that the Society might profitably initiate some research on the limitations of educability, might help to answer the question whether educational effort can or can not make bright children out of dull ones. The Executive Committee of the Society, at a meeting held at Detroit, January, 1924, acted favorably upon the suggestion and requested Dr. Terman, in collaboration with Dr. Bagley, to select a suitable committee to undertake the work. In April, 1924, the plans were formally endorsed and a Committee appointed to produce a Yearbook on the "Possibilities and Limitations of Training." Further development of the plans has been set before the Society already (for example, in the 24th Yearbook, Part II, pp. 386-8, and in the 25th Yearbook, Part II, pp. 254-6). The change of title of the Yearbook itself does not, of course, involve any essential change in the subject matter dealt with by the Committee during the four years the work has been under way.

The Board of Directors appropriated for the expenses of this Yearbook Committee $700 in April, 1924, $600 in October, 1925, and a special fund of $500 in February, 1927. These sums, however, represent only a fraction of the money expended in procuring the data that follow. Four of the major contributions—the Chicago and Stanford investigations of foster children (Chapters IX and X of Part I), the Hollingworth-Cobb investigation (Chapter I of Part II), and the Heilman investigation (Chapter II of Part II)—have been facilitated by grants from the Commonwealth Fund, from Mr. Max Rosenberg, from Stanford University, from the Institute of Educational Research of Teachers College, from the Carnegie Corporation of New York, from the Colorado State Teachers College, and from the Denver Public Schools. The Society expresses its appreciation of these contributions; explicit acknowledgment of them is made elsewhere in these volumes. Many other organizations and institutions have also assisted us, especially through permitting their representatives to devote their time and energy to the gathering of data and preparation of contributions. If these contributions and the personal efforts of the numerous persons concerned could be computed also in terms of dollars and cents, the investment of capital in the production of the Twenty-Seventh Yearbook would indeed be impressive. The book presents a splendid example of the kind of coöperative research which the Board of Directors has aimed to develop through its policy of establishing committees of experts, supplying them directly or indirectly with adequate funds, over a period of several years, and providing a suitable medium for the publication and dissemination of their reports.

The fundamental importance of the topic under discussion in this Yearbook is clearly set forth in Dr. Terman's introductory chapter. Nothing more need be said on that point, but it may be well here to warn the reader that the problem is not only important, but also difficult. It has been impossible in this volume to preserve the relatively simple style of presentation that has characterized many of our yearbooks which were designed to be directly helpful to the classroom teacher. It has been necessary in the present series of studies to employ intricate methods of approach and complex statistical methods of treatment. That could not be avoided, but the Committee has sought to assist the lay reader by prefacing each chapter with a brief orientation, explaining the purpose, method, and outcome of the study in question and its relation to the general problem of the Yearbook.

Those who think of the nature-nurture issue as a fruitful field for rhetorical debate will be disappointed in the pages that follow. No one has tried to be controversial. Every contributor has sought to be straightforward and objective. The aim in this Yearbook has not been to compare opinions (however helpful that may have been in dealing with other educational issues), but to determine facts.

With these explanations in mind, the editor trusts the members of the Society will approve of the content and style of the Twenty-Seventh Yearbook, despite its technical vein and the somewhat abstruse character of many of the contributions.

G. M. Whipple

CHAPTER I

INTRODUCTION

LEWIS M. TERMAN
Stanford University
Chairman of the Yearbook Committee

Background of the 1928 Yearbook. For several years a lively controversy has waged among school people and psychologists with reference to the causes and significance of the large individual differences which intelligence and achievement tests have disclosed. The issues involved are obviously of great consequence for educational theory and practice, since they are part and parcel of the age-old problem as to the relative influence of nature and nurture upon human destinies. If the differences found are due in the main to controllable factors of environment and training, then, theoretically, at least, they can be wiped out by appropriate educational procedures—procedures which it would then become our duty to provide. On the other hand, if they are due primarily to differences in original endowment, then the duty of the school is clearly to provide for differentiated training which will take these native differences into account.

Now any proposal to apply in practice the doctrine of the biological inequality of human beings, whether in politics or education, impinges upon some of our deepest-lying prejudices and challenges our fundamental philosophies of life. It was only natural, therefore, that all too frequently the controversy should have descended from the level of calm discussion of the pertinent scientific evidence to that of acrimonious asseveration and emotional appeal. Rightly or wrongly, some have felt that educational democracy is at stake, and any threat, fancied or real, against the democratic ideal of public education arouses the same antagonism as a threat to the democratic principle of political control. Catch phrases, like "educational determinism" or "democracy and the I.Q.," have become as charged with emotion as were once such slogans as "states rights," "abolition," or "taxation without representation." The belief in native differences and in the validity of intelligence tests as a method of measuring them is looked upon by some of the environmentalists as belonging to the same category as the theological doctrine of infant damnation. Unfortunately, the dogma of certain of the environmentalists has too often been answered by the dogma of hereditarians. Too many on both sides have shown more interest in finding and phrasing arguments to support a theory than in investigating the problem by the methods of science.

Not that there has not been ample reason in the past why speculations rather than facts have tinged our thinking upon this question. As a scientific problem, this one is exceedingly young. Until Galton (without proving whether the fact was due to nature or to nurture) showed, when he published his *Hereditary Genius* in 1869, that high ability has a strong tendency to cluster in family lines, no one had ever made an extensive investigation in this field. Although *Hereditary Genius* was the stimulus for a few lesser studies which followed it closely in time, it was not until the beginning of the twentieth century that concerted resources were devoted to unravelling the problem. Even then the early attempts were of necessity tentative and incomplete, partly because the problem itself was as yet merely sensed rather than clearly formulated, partly because a large amount of preliminary exploratory work was essential before the lines of attack most likely to be fruitful and crucial could be appreciated, and partly because refined methods of measuring intellectual abilities were not yet available.

Possibly one misconception (now fortunately clarified by recent contributors to the nature-nurture field) has clouded the pertinent issues as much as any single other. This is the notion that a 'nature' theory implies a type of 'glass bottle' mental development which is quite independent of any stimulation from environment. No idea could be more misleading, and no phenomenon more impossible. The theoretical as well as the practical nature-nurture problem that faces us deals with human beings as they are—not as an experimenter might rear them in a nurture-proof laboratory. We are interested in the child or adult as he comes to us—with his unique complex of ancestry, associates, home training, schooling, and physical and moral attributes. We are interested, not in finding out how he would have developed if he had had no environment at all; rather we wish to discover whether or not he can be made a more intelligent individual or a more learned one by improving the conditions of his *milieu* within the limits found in reasonably good social communities. More generally, we wish to find the relative potency of all types of human environment to add to, or to detract from, human endowment, and to know the limits placed upon achievement by endowment.

Inception of the Present Yearbook Project. The situation set forth in the preceding paragraphs was discussed in the February, 1924, meeting of the Board of Directors of the National Society for the Study of Education, and following this discussion an agreement was reached to devote an early yearbook to the topic, "The Possibilities and Limitations of Training." Shortly afterward I was invited by the Board of Directors to form a committee for the preparation of a yearbook devoted to this subject. I agreed to do so on condition that the project be made a research undertaking rather than a spectacular combat of educational gladiators. In view of the large amount of heated discussion that had taken place, it seemed best to devote this yearbook almost entirely to the dispassionate exposition of the results of new investigations bearing on the nature-nurture problem, with only a minimum of space devoted to orientational and interpretative contributions.

A committee was at once formed which was favorable to such a program. The Committee was selected so as to represent all shades of opinion on the issues involved, for it seemed very important to avoid the suspicion of making the Yearbook the special mouthpiece of either side to the controversy. Whatever faults may be found with the outcome of the Committee's work, I think it can be said that all of its members have shown throughout an objective-minded, investigational attitude. The sole concern of the Committee has been to bring together all the new evidence it was possible to secure on either side of the question. This method of approach once agreed upon, one finds that it is possible for persons of the most antagonistic beliefs to coöperate effectively in a program of research.

Activities of the Committee. The Committee has had only two meetings, one in Chicago in October, 1925, the other in Philadelphia in December, 1926. The Chicago meeting was fully attended and occupied an entire day. The Philadelphia meeting was briefer and could be attended only by Baldwin, Brigham, Terman, and Whipple. Fortunately, the Committee chairman (Terman) and the Yearbook editor (Whipple) were able to hold a series of conferences at Stanford University in May and June, 1927, and to agree upon various matters with reference to the selection and arrangement of material.

Previous to the first meeting of the committee, the general plan of the Yearbook had been determined upon by correspondence. This plan included the following specifications:

1. The Yearbook should be devoted to the nature-nurture problem, with special reference to the relative influence of these factors upon scores in intelligence and achievement tests.

2. The major part of the space should go to the exposition of new investigational results.

3. To this end, effort should be made to secure the coöperation of as many scientific workers as possible in the conduct of researches bearing on various aspects of the problem.

4. Investigations tending to emphasize *either* the *possibilities* or the *limitations* of training were to be equally welcomed.

5. Contributors would be expected to confine themselves primarily to the exposition of factual data and to the brief statement of their conclusions based thereon.

In October, 1925, an appeal for coöperation was mailed to nearly two hundred persons throughout the country, and this was followed by two published appeals, one in the *Journal of Educational Research,* May, 1926, the other in *School and Society,* March 27, 1926. Systematic personal enquiries were made for the purpose of getting in touch with investigations already under way which might have a bearing on the Yearbook's problem. Other investigators were urged to undertake special researches for the purpose. The two large and meaty volumes here offered are evidence enough of the response secured. The table of contents for the two volumes shows a wide variety in the special phases of the problem which were attacked. The part devoted to factors influencing intelligence-test scores covers such broad and basic topics as "home environment," and such special ones as "coaching." The part dealing with achievement test scores presents critical data upon a number of factors, such as school expenditures and school attendance, whose efficacy to produce educational results we have always taken for granted in the past. The section on special mental traits offers upon several important problems material which is virtually pioneer work. It would not be easy to estimate the total labor cost which the investigations of these two volumes represent. In all probability it would bulk to more than $100,000 if bought and paid for in the market. Several of the studies would not have been made at this time, had it not been for the stimulus offered by the Yearbook project.

Acknowledgments. In behalf of the Committee I wish to thank all those who have contributed so liberally of their time and effort to make this Yearbook a success. In many cases the contributions have been made at considerable personal sacrifice. The Committee appreciates not only the contributions as such, but also the nonpartisan and scientific spirit in which they have been made. There has been little or no rhetorical straining for effect in the exposition or interpretation of results. Few treatments of the nature-nurture problem have consistently maintained so high a level of open-mindedness and intellectual honesty.

Grateful acknowledgment is made to the Commonwealth Fund for a grant of $15,000, divided equally between Chicago and Stanford Universities, in support of the foster children studies. The Stanford grant was supplemented by a contribution of approximately $2,500 from the Thomas Welton Stanford Fund for Psychological Research and by a gift of $1200 from Mr. Max Rosenberg, of San Francisco. The State Teachers College of Greeley, Colorado, and the public school officials of Denver coöperated helpfully in making possible the extensive and laborious investigation by Dr. Heilman. Many other educational institutions and many individuals, not all of whom it is possible to mention by name, have contributed in important ways to the carrying out of the investigations herein reported.

Special acknowledgment is made of the invaluable assistance rendered by Barbara Stoddard Burks, now Mrs. Herman Ramsperger, of Stanford University. Besides carrying out the Stanford foster-children investigation and preparing the masterly summary of nature-nurture literature, Miss Burks served throughout the period of the Yearbook's preparation as my research and secretarial assistant. In this capacity she not only conducted the major part of the routine correspondence, but in addition gave valuable as-

sistance to several contributors, both in planning their researches and in reporting their results. Following her suggestions, several chapters of the Yearbook were largely rewritten and many statistical procedures were modified. Probably no other single individual deserves more credit for whatever merits the Yearbook possesses.

Suggestions for Use of the Yearbook. The amount of real service which these volumes will render to educational theory and practice will depend in no small degree upon the spirit in which their content is approached. In the first place, readers who do not sufficiently appreciate the inherent difficulties and complexities of the problem, particularly the difficulty of securing factual data that are at once crucial and unambiguous, will perhaps be disappointed in the lack of perfect consistency in the results of various contributors. For certainly it must be admitted that no final answer to the nature-nurture question has been attained or even approximated. The most that can justly be claimed is that the bounds of our knowledge have been in some measure extended. Anyone who has reasonable regard for the fundamental importance of the problem will be grateful for any advance in our knowledge, however slight, which the contributions of this Yearbook have accomplished. It is conceivable that the elusive nature of the problem is such as to preclude for a long time to come, if not forever, a complete and final solution. One needs to cultivate patience and the faculty of suspending judgment.

Secondly, in his study of the investigations reported the reader is urged to divest himself, in so far as it is humanly possible, of his emotional bias and predilections. Because of unconscious factors that are so likely to play a rôle, one can not always be sure whether one has or has not attained to a state of complete intellectual detachment, but the effort should nevertheless be made. How absurd, after all, is the attitude of mind which would force truth into any kind of preconceived mold, when only truth that is genuine and undistorted can give us the control over human nature and human destinies that our various institutions, including the school, are intended to exercise! Surely, anyone who is imbued at all with the spirit of modern science should be less concerned about *what* is true than about knowing what *is* true. In order to counteract the inevitable human tendency to bias, the reader will do well to make a conscientious effort to find merit in those investigations which lend least support to the beliefs he has previously been inclined to favor. It is so easy, as Darwin long ago warned us, to forget the evidence that does not happen to agree with our theories!

Finally, I wish to express the profound hope of the Committee that numerous readers of the Yearbook will be stimulated by the very shortcomings and imperfections of the investigations reported to undertake researches which will be better planned, better controlled, more extensive, and therefore more conclusive in their results. It is by this standard, ultimately, that the success of our efforts must be judged.

PREFATORY NOTE TO CHAPTER II

Any attempt to investigate the relative significance of nature and nurture through modern avenues of investigation is almost certain to lead to more or less elaborate statistical treatment of the numerical data that are assembled. While the science of statistics has within recent years supplied the investigator with various novel weapons of attack, they are often two-edged weapons, whose use is not without hazard to the user. The purpose of this chapter is to describe some of these statistical hazards.

At the request of the Yearbook Committee, Miss Burks has drawn up the presentation of the hazards which constitutes the first part of Chapter II. Her presentation was then referred to Professor Kelley, whose interest in this particular matter is attested by his *Influence of Nurture upon Native Differences,* with the request that he make such supplementary comments and criticisms as he deemed wise. Professor Kelley's contribution constitutes the second part of the chapter.

Readers who have no familiarity with modern statistical methods will probably find portions of this chapter too technical for easy perusal; however, we suggest that they will be repaid if they take the trouble to skim the chapter, omitting the technical portions, because many of the fallacies described therein have been altogether too prevalent in supposedly scientific, not to mention popular, discussions of heredity and environment. We shall never get anywhere in such discussions unless we are perfectly clear as to the fallacies that are likely to beset our thinking.—*Editor*.

CHAPTER II

STATISTICAL HAZARDS IN NATURE-NURTURE INVESTIGATIONS

BARBARA STODDARD BURKS
AND
TRUMAN L. KELLEY
Stanford University, Palo Alto, California

I. CHIEF CONSIDERATIONS

1. Selection

'Selection' is given first place among the hazards because it is so persistent, so widespread, and often so hard to recognize. A practical definition of selection as used here would be: the systematic operation of one or more factors that prevent a group of individuals from being what they are assumed to be. It is found, for example, in attempting to determine how much native difference exists between the mental levels of various races, that in the higher school grades negro children are closer to the level of white children than is the case in the lower grades. On the face of it, this might appear to mean that schooling had wiped out the early difference between negro and white children. If the white and negro children in the higher grades were typical of children of their age, this would indeed be the case. But if it turns out that only the ablest negroes continue at school, it may be that nurture has had no effect at all in narrowing the gap between the abilities of the chosen samples of the two races.

More subtle selective factors may sometimes be at work. If, in an investigation of the mental resemblance between parents and children, 40 percent of the families approached refused to lend themselves as subjects for the experiment, it is evidently possible that the families refusing coöperation might be those in which the children resembled the parents least. Bright parents might be ashamed of children less bright than themselves, and the occasional dull parents who have bright children might shrink from exposing their relative backwardness. Many other examples of selective errors could be cited. The scientific worker must be constantly alert to avoid them.

2. Inextricable Causes

There are many types of study in which this hazard is inherent. Under some conditions the only way to obviate the difficulties is to find a new approach to the problem that will extricate the 'causes' by experimental means.

Let us consider for a moment the correlations of .40 to .50 between the intelligence of siblings or between that of parents and offspring which have been reported by many different investigators. To what are these due? "To nature, to similar germ plasm," answers the hereditarian. "To nurture, to the molding influences of home training and similar educational opportunities," answers the environmentalist.

Either answer is consistent with the observed facts, yet neither answer can be established through the facts immediately at hand. The hypothesis that family resemblance may be due to the *combined* forces of nature and nurture could also explain the observed facts. It therefore behooves us to defer interpretaton until data from studies using different methods of attack untangle the real causes of family resemblance.

Analogous situations could be found in nearly every phase of the field under discussion. Are the correlations of .60 to .80 usually found between intelligence and school achievement due to an influence of intelligence upon achievement, or to an influence of schooling upon intelligence? Burt,[1] arguing from a regression equation for predicting mental age from educational age and other variables, concludes that mental ability as measured on his revision of the Binet Test is the product of schooling to the extent of about 50 percent. Courtis,[1] on the other hand, arguing from a regression equation for predicting educational age from mental age and other variables, concludes that nearly 90 percent of the pupil's school achievement is conditioned by his intelligence. All question aside as to the validity of the statistical methods by which the actual numerical estimates were obtained, it is an interesting, though somewhat disturbing, fact that two such opposing conclusions can be drawn from sets of very similar data simply through the *a priori* assumptions underlying the reasoning in each case.

---

[1] See the summaries of studies by Burt and by Courtis and others in Part II of this Yearbook.

3. Use of Partial and Multiple Correlation

There has been a tendency to assume that the techniques of partial and multiple correlation offer a precise means for evaluating the relative contributions of certain 'causal' factors to a criterion, as for example 'intelligence' or school achievement. For a more extended discussion of the topic than that given here the reader is referred to an earlier paper[2] by the writer. A few excerpts from the paper may suffice to indicate the type of danger that besets the user of partial and multiple correlation techniques.

We may approach the problem inductively by first examining a hypothetical instance of the misapplication of partial correlation so extreme that it would scarcely be approached by any real situation. We wish to investigate, let us say, the effect of age of entering first grade upon subsequent rate of progress through the grades. We may take for our subjects an unselected group of 12-year-old children. After noting for each child (*a*) the age at which he entered school; (*b*) his present grade; and (*c*) his rate of progress as measured by the average length of time it has taken him to complete each grade, we may find a substantial correlation between (*a*) and (*c*). Such a correlation would be reasonably expected if bright children tend both to enter school very young and to receive extra promotions.

Now let us render "present grade" constant by partialling it out. Then the correlation between age of entrance and time required to complete a grade necessarily falls to negative unity, since we are dealing now with children whose early entrance to first grade is exactly balanced by slow progress (*i.e.*, more than normal time spent in each grade), or whose late entrance to first grade is exactly balanced by rapid progress. The condition imposed that all children shall be in the same grade means that age of entrance must completely determine rate of progress. Yet it does not follow that the relationship between age of entrance and progress is absolute, irrespective of the "effect" of present grade upon progress. Indeed, present grade could have produced no *effect* whatever upon the rate of progress preceding it, since it is itself the direct *resultant* of age of entrance and rate of progress. The application of the partial correlation technique has thus been meaningless in so far as we are interested in getting at any true interdependence of age of entrance and rate of progress.

Next let us examine a less extreme situation. In an attempt to determine how much effect differences in environment have upon the development of children's intelligence we might collect data from an unselected group of families giving the intelligence-test scores of a child

[2] *Jour. Educ. Psych.*, November and December, 1926.

and both his parents, and the family cultural status as measured on a specially devised scale. If the results checked with the trends indicated by various investigators, the correlations computed might not be far from the following:[3]

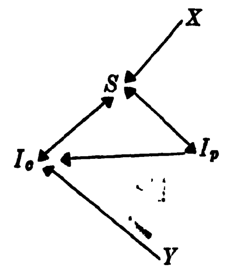
Mental age of mid-parent and intelligence of child......	.60
Mental age of mid-parent and cultural status..........	.77
Intelligence of child and cultural status...............	.48

Using these hypothetical figures, let us calculate the partial correlation between intelligence scores of mid-parent and child. In current terminology we "hold constant" the cultural status, or "eliminate the effect" of cultural status.

Substituting in Yule's formula,

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r^2_{13})(1 - r^2_{23})}} = .42.$$

Should we then be justified in concluding that the real relationship between the intelligence of parents and their children, after similarities induced by similar cultural surroundings had been discounted, was measured by a correlation coefficient of only .42? Surely not, because we know that by the nature of partial correlation we have eliminated not only those portions of parents' and children's intelligence that may *result* from differences in cultural status, but those portions which *contribute* to cultural status as well. The situation is represented diagrammatically below. I_p represents intelligence of mid-parent; I_c, intelli-



gence of child; S, cultural status; X, factors other than I_p or I_c contributing to S; and Y, combinations of genes not showing in the measurement of I_p contributing to I_c. It is readily seen that in applying the partial correlation formula to this situation we have "partialled out" too much, and that *in any study of causation we are partialling out*

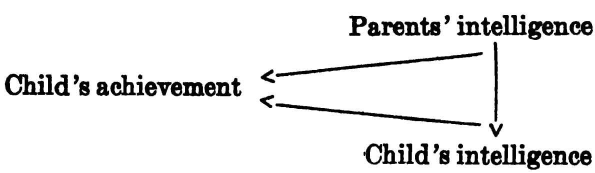
[3] The correlation figures have been changed from the hypothetical values appearing in the quoted article to actual values (corrected for attenuation) computed later from experimental data collected by the writer.

too much when we render constant factors which may in part or in whole be caused by either of the two factors whose true relationship is to be measured, or by still other unmeasured remote causes which also affect either of the two isolated factors.

The same generalization, of course, applies to studies of causation . . . in which mutiple correlations are computed with different factors under investigation successively dropped out to see how much each 'contributes' toward estimating the criterion.

The question . . . . arises as to whether partial or multiple correlation can ever be fully defended as instruments in the study of causation. Surely, as Kelley has pointed out,[4] there is no more justification for inferring causation in partial correlation than in raw correlation. Nevertheless, there are situations in which a variable is indisputably a cause rather than an effect. In one obvious type of situation, chronological age is such a variable. The many studies which have employed partial correlation technique to eliminate the contribution of maturity to correlated measures are apparently on safe ground, provided they have published the precise age ranges of their subjects. However, the caution sounded by Yule[5] in his chapters on partial correlation and normal correlation should not be lost sight of: namely, that except in the relatively rare situation of normal correlation, the partial coefficient cannot be taken as a measure of the correlation between variable 1 and 2 for every constant value of variable 3, but only as a sort of average correlation.

An additional consideration not dealt with in the earlier article suggests a further limitation of the conditions under which the partial correlation is a valid measure of causation. To justify its use as such a measure it is not sufficient merely that the factor partialled out should stand in the relation of cause to the two variables whose interdependence is to be measured. If the variable partialled out is entirely irrelevant to the problem at hand (as we took chronological age to be in our example) the partial correlation does give us what we seek. However, what would be the proper procedure to apply if we had a situation represented by the diagram shown herewith?



[4] *In* Rietz, H. L., et. al., *Handbook of Mathematical Statistics*, 1924, pp. 139 ff.

[5] *Introduction to the Theory of Statistics*, 1924.

The variable to be partialled out, parents' intelligence, is by hypothesis a cause and is not itself caused by any of the other variables entering the problem. The arrows in the diagram indicate the directions which the influences of cause to effect are assumed to take. Since teachers often attribute good class work in part to home training received from superior parents, and likewise attribute poor class work to the lack of encouragement received from unintelligent parents, it would be pertinent to establish, if possible, the relationship between the child's school achievement and his intelligence when the influence of parental intelligence is eliminated.

If we follow the obvious procedure of partialling out parental intelligence, we indeed succeed in eliminating all effect of parental intelligence. But here, as in the first illustration used, we have partialled out more than we should, for the *whole* of the child's intelligence, including that part which can be predicted from parents' intelligence as well as the parts that are due to all other conditioning factors, properly belongs to our problem. We are interested in the contribution made to school achievement by intelligence of a normal range of variability rather than by the narrow band of intelligence that would be represented by children whose parents' intelligence was a constant. The partial-correlation technique has made a clean sweep of parental intelligence. But the influence of parental intelligence that affects achievement indirectly *via* heredity (*i.e.*, *via* the child's intelligence) should stay; only the direct influence should go. Thus, the partial-correlation technique is inadequate to this situation. Obviously, it is inadequate to any other situation of this type.

4. Partial Regression Equations

The use of partial regression equations in studies of causation is subject to some of the same dangers already mentioned under (2) and (3). Causes and effects are likely to be confused or interchanged if there is not some *a priori* basis outside the data themselves for defining which variables are causes and which effects. When interpreting the significance of regression coefficients, even though a rational hypothesis regarding the direction of causal influence to effects has been made, still another type of error sometimes occurs. It is commonly assumed that the regression weights of the so-called 'independent' variables in the regression equations are directly proportional to the actual contributions of those variables to our prophecy of a criterion or 'dependent' variable. It has frequently been said of such a type equation as the following,

$$X = .24A + .61B + .15C,$$

(where *X, A, B,* and *C* are measured in standard scores), that *X* consists of 24 parts *A* to 61 parts *B* to 15 parts *C*; or that in estimating *X, A, B,* and *C* contribute in the proportions 24, 61, and 15. It will now be shown, by means of a simple numerical example, that such an interpretation is not justified, and that the regression coefficients, or 'weights,' can be conceived as conversion factors for putting our independent variables into the same units as those employed in measuring the criterion.

Let the variability in the four variables *X, A, B,* and *C* be due to a number of uncorrelated variable factors as follows:

$$\begin{aligned} X &= a + b + c + d + e + f \\ A &= a \\ B &= b + c \\ C &= d + e + f \end{aligned}$$

We may assume without loss of generality that the variables and their component factors are measured as deviations from their own means, and that the component variables have standard deviations which are all equal to one. We can then easily compute the intercorrelations between *X, A, B,* and *C* by the method which has frequently been used by Spearman and by Thomson. The *r* in each case is $\frac{N}{\sqrt{N' N''}}$ where N is the number of factors common to two variables, and N′ and N″ are the total number of factors in each of the respective variables.[6]

$$\text{Thus } r_{AX} = \frac{1}{\sqrt{6}} = .4082, \text{ etc.}$$

Applying the formula to all the pairs of variables, we get this table of correlations:

	X	A	B	C
X	1	.4082	.5774	.7071
A	.4082	1	0	0
B	.5774	0	1	0
C	.7071	0	0	1

[6] A simple derivation of this procedure was presented by the writer in the article mentioned.

The intercorrelations have been tabled thus because the matrix of the determinant Δ which is used in computing multiple correlations and multiple regression coefficients is composed of the intercorrelations arranged in this way.

Computing $R_{X.ABC}$ by the ordinary formula, we find R equal to unity. This, of course, might have been foreseen, since all the factors in *X* are contained either in *A, B,* or *C.*

Now, computing the regression equation for estimating *X,* we get

$$\frac{\overline{X}}{\sqrt{6}} = .4082 \frac{A}{1} + .5774 \frac{B}{\sqrt{2}} + .7071 \frac{C}{\sqrt{3}}$$

where X, A, B, and C are all measured as deviations from their own means. Thus, we see that in the simple case in which *A, B,* and *C* are independent of one another the weights to be attached to the standard scores of each to predict *X* are equal simply to the correlations of each with *X.*

This does not mean, however, that *X* is composed of *A, B,* and *C* in the proportions .4082, .5774, and .7071; nor even, what might seem more reasonable, that proportions .4082 of *A,* .5774 of *B,* and .7071 of *C* summed together give a composite that is equal to *X.* By hypothesis *all* of *A, B,* and *C* must be summed to give a composite equal to *X.* If we reduce our regression formula above, we see that condition fully met; for, multiplying the left and right members of the equation by $\sqrt{6}$, we get $\overline{X} = A + B + C$, in which the scores of all the variables are measured in the same system of units.

Of course we seldom get results of such simplicity in actual practice because it is only in an artificial situation that we could find a group of variables whose gross scores were all measured in the same units. Even a group of variables which *apparently* employ the same units often suffers from Professor Aikins' "jingle fallacy," as for example in Burt's much buffeted regression equation for predicting Binet mental age. Achievement age, "reasoning" age, chronological age, and Binet mental age were all measured in *years.* But the underlying common trait was *level of intelligence,* and the four variables expressed it in four sets of units which were rendered comparable only by the conversion factors of the regression equation.

A partial analogy from another science may serve to clarify this point. Through the chemical association of sodium and chlorine, ordinary table salt is obtained. What will be the result in terms of salt if we combine a large number of sodium atoms, say 26×10^{18}, with a given volume of chlorine gas, say .48 c.c.?

$$26 \times 10^{18} \text{ sodium atoms} + .48 \text{ c.c. of chlorine} = ?$$

It is obviously impossible to tell, unless the amounts of both sodium and chlorine can be expressed in common units which can be used also to express the salt product. Knowing the weight of an atom of sodium and the specific gravity of chlorine, we can express both in terms of milligrams, thus obtaining the equation for 'predicting' table salt as follows:

$$1.0 \text{ mg. sodium} + 1.51 \text{ mg. chlorine} = 2.51 \text{ mg. salt.}$$

Since regression coefficients are not valid to ascribe relative efficacy to different variables in estimating a criterion, the question arises whether some other device might not be. Since we are dealing with deviations from means (not with absolute levels, *i.e.*, total scores), we are justified in holding that a variable is effective in estimating a criterion in proportion to the amount of variance remaining in the criterion when all other factors are held constant but the variable in question varies as much as before.

In our first example, this is very easy to determine in the case of *A*, *B*, and *C*, since they are independent of one another.

$$\sigma^2_{X.A} = \sigma_X^2 \, (1 - r^2_{XA})$$

The holding of *A* constant thus reduces the full value σ^2_X by a proportion r^2_{XA}.

Proceeding similarly with *B* and *C*, we find for the relative efficacy of the variables in estimating *X*,

$$\begin{aligned} r^2_{AX} &= (.4082)^2 = .1667 \\ r^2_{BX} &= (.5774)^2 = .3333 \\ r^2_{CX} &= (.7071)^2 = .5000 \\ & \qquad\qquad\quad \overline{1.0000} \end{aligned}$$

It thus appears that the *squares* of the regression coefficients (which in this case are also the squares of the correlation coefficients) are proportional to the contributions to estimate. This, like some of our previous results, also might have been predicted, since the fractional expressions for the values of the r's squared are $\frac{1}{6}$, $\frac{1}{3}$, and $\frac{1}{2}$, respectively, and are at once seen to be equal to the proportional number of factors in *X* which are contributed by *A*, *B*, and *C*.

Sewell Wright[7] has shown that in the general case in which the 'causal' variables are correlated with one another, instead of uncorrelated, the squares of the regression coefficients (which he terms "path coefficients") still represent the proportional contributions made by the different variables to estimating the criterion. The reader is referred to Wright's articles in which the use of these path coefficients is explained. Wright's caution that his method offers no basis for assuming causal relationships, though providing the means for evaluating numerically causal relationships already known to exist, should be borne in mind whenever the technique is used.

5. Nygaard's "Percentage Equivalent for the Coefficient of Correlation"

In a recent article by Nygaard[8] a statistical hazard has been pointed out which concerns the numerical interpretation of the correlation coefficient. But in attempting to develop a method by which a "direct and understandable interpretation may be made of the amount of relationship indicated by a coefficient of correlation" Nygaard has unwittingly fallen into another pitfall.

Assuming (1) that trait *C* depends for its value entirely upon traits *A* and *B*, (2) that *A* and *B* are uncorrelated, and (3) that a weight, k, of *A*, and a weight, h, of *B* combine to give *C*, the author makes the statement that "the ratio of dependence, or percentage of dependence, of *C* upon *A*, will be $\frac{k}{k+h}$, and upon *B*, $\frac{h}{k+h}$."

This statement would be perfectly true if *A* and *B* were measured in identical units, but not otherwise. If John's first helping consisted of one quarter of a pie, and his second helping consisted of one sixth of a pie, his entire dessert would have a "percentage of dependence" upon the first helping of $\frac{\frac{1}{4}}{\frac{1}{4}+\frac{1}{6}} = \frac{3}{5} = 80$

[7] *Jour. Agric. Research*, 20:1921, 557-585; also *Genetics*, 8:1923, 238-255.
[8] *Jour. Educ. Psychol.*, February, 1926.

percent, and upon the second helping of $\frac{1/6}{1/4 + 1/6} = 2/5 = 20$ percent. But if John's dinner consisted of a stick of celery and half of a fried chicken, the mistake would be immediately obvious if we attributed $\frac{1}{1 + 1/2} = 2/3$ to the celery and $\frac{1/2}{1 + 1/2} = 1/3$ to the chicken.

Since it is almost never that the variables used in predicting a criterion are measured in the same units, the formulas for "percentage equivalents" derived by Nygaard have a very much narrower applicability than he believed them to have.

6. Correlations in Populations of Various Ranges

In view of the numerous discussions that have appeared in the literature regarding the dependency of correlation values upon the range of talent of the populations from which they are derived, it is not necessary to treat this topic at length here. Yet in spite of the comments and formulas for correction that have been published, the fact that in experimental work comparisons between test groups and control groups are occasionally still made without taking any account of possible differences in range indicates that the importance of this hazard is not universally appreciated. Great care is sometimes taken to see that two populations differ by a measurable amount in a 'controlled' factor, such as social status, health, intelligence, etc., yet the control and experimental subjects may be chosen in such a way that a very significant difference obtains in the range of the ability or talent investigated. Sometimes, also, data collected by different experiments may fail to agree simply because the range of the subjects is different in each case. Under such conditions conclusions may be drawn that are conflicting, misleading, or false.

Kelley[9] has tabled the values that correlation coefficients between two variables would take as the range of one variable was extended. If the original r is .60 for example, it becomes .707 if the ratio of the original S.D. to the S.D. for the extended range is .75; the r becomes .832 if the ratio is .50; it becomes .949 if the ratio is .25, and .991 if the ratio is .10, etc.

[9] *Statistical Method*, 1924, p. 225.

7. Incommensurability of Results from Different Tests

There has been a more or less prevalent tendency for investigators to treat test results as though the *names* of the tests accurately defined the functions measured, and as though the scores on any two tests—even if they employ rather different material—are comparable provided only the tests are called by the same name. Thus, for example, we find studies that purport to measure the effect of language handicap on verbal intelligence tests scores by comparing the mental ages of foreign children earned on verbal and on non-verbal intelligence tests. The mental ages of children of certain low-testing nationalities commonly turn out to be closer to the norms of American children when measured on non-verbal tests than when measured on verbal ones. But in as much as verbal and non-verbal test scores, even for American children, seldom correlate with one another higher than .6 or .7, it is obvious that, although both types of tests are called 'intelligence' tests, they each measure about as much not held in common as they measure of what is held in common. Hence, it is not legitimate to infer from such data alone that language handicap accounts for the low scores of the foreign children on verbal tests. It would be only a little less defensible to argue that because certain national groups averaged close to American norms in some such trait as height (which has been shown to have a slight correlation with intelligence), their deficiency on intelligence tests was therefore proved to be due to language handicap.

8. Spurious Index Correlation and Spurious Mutual Correlation with Age

Two sources of error which are often overlooked are discussed by Thomson and Pintner.[10] The first of these, spurious index correlation, results when the paired scores of a correlation table are divided by the same figure, provided this figure varies from pair to pair, and provided the resulting scores show a negative correlation with the figures by which the original paired scores were divided. This situation is quite commonly met when I.Q.'s are computed for mental tests of the kind *for which I.Q. and*

[10] *Jour. Educ. Psych.*, Oct., 1924.

chronological age are negatively correlated. The correlation between I.Q.'s on two such tests is higher than the value representing the real similarity in the functions measured by the tests, owing to the fact that each pair of scores has undergone distinctive treatment.

The other type of error discussed by Thomson and Pintner is that of spurious mutual correlation with age. In the words of the authors: "It is quite possible for two tests to have no organic connection with one another, and yet for the M.A. found by either to correlate highly with chronological age up to even .7 in extreme cases. In such a case there might be no correlation between I.Q.'s, and yet, if the cases were spread well over a long range of chronological age, there might be a very high correlation of M. A.'s between the tests." A striking example of the same type of spurious correlation is that between M.A. and height if a wide age range of children is used. This correlation nearly disappears if children only of a single age group are retained.

It seems appropriate to mention also the type of misinterpretation that is likely to result if I.Q.'s and E.Q.'s are compared on various tests that show *different correlations with chronological age*. It is often assumed that if a child is working 'up to capacity,' his E.Q. will equal his I.Q. Often, likewise, the supposed evenness or unevenness of his various abilities is gauged by the fluctuation in his E.Q.'s for different school subjects. For very rough practical purposes, this procedure may suffice. But it would be perfectly possible for a pupil who had the same percentile ranking for his age in intelligence, reading, and handwriting, for example, to have very different 'quotients' in these traits, simply because the traits showed different amounts of overlapping with chronological age.

The 'standard score' has been proposed and used in preference to the E.Q. by some writers. This gives a pupil's deviation from his age norm in terms of the variability of his age group and would seem to be the most meaningful measure obtainable, providing real age norms are secured. All too often, however, the experimental literature provides instances in which the mean score for a year age-range is used as the norm for a chronological age span of twelve months. Such a method as the one advocated by De Voss[11]

---

[11] In *Genetic Studies of Genius*, Vol. I, 1925, Chapter XII.

for obtaining an *achievement profile* in terms of 'standard scores' penalizes the children who have just had a birthday, and gives an undue advantage to those closely approaching a birthday. Moreover, these penalties and advantages have different amounts for different tests of an achievement battery, depending upon the correlations of the separate tests with chronological age. The writer has calculated that for certain pairs of tests from the Stanford Achievement battery, discrepancies greater than a standard deviation of the score distribution would occur in children whose *true ability* was the same on the two tests, if these children were measured by norms nearly six months away from their actual ages. The only help for this difficulty would be to secure separate standard score norms for each month of age increment, instead of for year increments only.

9. Confusion Between Variability and Absolute Level

Occasionally, one encounters statements like this: "The ultimate achievement of any given individual is due to his original ability, probably to the extent of 60 to 90 percent, and to actual differences in opportunity or external circumstances only to the extent of 10 to 40 percent.[12] The meaning of such statements is usually far from clear. When we say that an ability owes so much to nature and so much to nurture, do we generally have in mind:

(1) Level of ability of every human being (which the quotation immediately above implies)?
(2) Human level of ability on the average?
(3) Deviations of every human being from the mean of the general population?
(4) Differences among human beings on the average?

Of these four possibilities, (1) and (3) contradict the observed facts, which are that nature and nurture are variable influences by no means perfectly correlated with one another. Consequently, the level of ability or the deviation in ability of different individuals is determined now by one proportion of nature to nurture and now by another proportion.

[12] Starch, D., *Educational Psychology*, 1919, p. 94.

The second possibility would perhaps offer a pertinent line of investigation if any way could be found to carry it out. The writer has never seen any adequate means proposed for making such a determination, however, and strongly doubts whether adequate means will ever be available. To do so would apparently require a knowledge of the effect of nature alone without the aid of nurture, and the effect of nurture alone without the aid of nature. The problem would thus break down, since no development whatever would be possible without the contributions of both nature and nurture.

There remains, then, the fourth possibility as the only practicable and unambiguous concept to use. When the relative contributions of nature and nurture are discussed, no doubt should be left in the mind of readers that contributions to *variability* are being considered.

This comment does not imply, of course, that pertinent and valuable studies cannot be made of the *direct* effects of a variety of influences upon abilities. It is important to know how many points hookworm or an illiterate home can depress the I.Q. or how many points a certain number of hours of study per week or a certain brand of textbook can raise the E.Q. Such investigations probably have more practical bearing than the classical type of study on the proportional contributions of nature and nurture, and indeed constitute the greater bulk of the chapters appearing in this *Yearbook.*

10. Ambiguity of Correlations Between Averages

It is not uncommon to encounter material which presents the correlations obtaining between average scores for a number of groups, each one of which contains a large number of individuals. Bagley, for example, presents correlations of the average Army Alpha ratings of men from the different states matched against average ratings of the states on various educational influences, such as schools, magazine circulation, etc. The correlation coefficients are extremely high—chiefly above .90. These results do not mean that the relationship between intelligence and schooling or magazine reading is as high as .90. On the contrary, studies reporting correlations between the intelligence of *individuals* and the number of years of schooling received by them generally yield coefficients around .60. The correlation coefficients become inflated when averages are employed because a great many factors that ordinarily keep the correlation between intelligence and education from being perfect cancel out (*i.e.*, they are approximately the same, taking the averages of entire states). Consequently, these 'inflated' coefficients cannot be interpreted as ordinary correlation coefficients can be. In fact, it is almost impossible in most cases to give any definite interpretation to them whatever.

11. Manipulation of Scores Without Providing Measures of Group Dispersion

Few things are more exasperating to investigators who wish to make use of previous work in their field than to come upon studies which present increases in certain test scores in terms of points or of percents without indicating the significance of these increases in terms of group variability. Do 10 points or 35 percent average increase on card-sorting scores and 10 points or 35 percent average increase on spelling scores mean the same thing or two different things? In one case such an increase might be equivalent to a jump from the 25th to the 75th percentile of 'unselected' ability, and in the other case, equivalent to a jump only from the 50th to the 51st percentile. It is evident that some measure of group dispersion should always be given in connection with reports of score improvement, and that the nature of the group having such dispersion should be very carefully defined.

12. Probable Errors when Individuals Are Used More than Once in a Scatter

This hazard is met when coefficients of family resemblance are computed in which several offspring are correlated against the same parents or several siblings per family are paired in all possible ways. Owing to the cumulation of errors of sampling when any score enters a correlation table more than once, the probable error of the resulting correlation coefficient is greater than the value given by the ordinary P.E. formula (in which N is taken to be the actual number of entries in the correlation table). If most of the scores enter the table a considerable number of times, the ordinary P.E. formula may be quite inapplicable. Formulas have been derived for determining the probable errors of coefficients so obtained when each score enters the table the same number of times.[18] Unfortunately, the occasion does not often arise when cases available for use can conveniently enter a table the same number of times, if the number of times is greater than one. When it is important to have an accurate determination of the P.E. of r (and it usually is important), either it should be shown empirically that to use the lower or upper limit of N yields approximately the same P.E., or enough cases should be discarded so that every score enters the correlation table the same number of times. If the former can be done, the lower limit of N would, of course, be the number of entries that could be made without using the same case twice. The upper limit would be the total number of entries in the correlation scatter.

13. Assumed Cumulative Effect of Environment

A pageful of citations could be presented in which some statement such as this is made: "It is fair to assume that the longer environment acts, the greater is its effect." By the use of this basic assumption, elaborate 'proofs' are sometimes built up to show that environment can or cannot account for such and such observed facts. The assumption may or may not be true; again, it may be true under some conditions, false under others; it is far from axiomatic. Thus, in some situations it would seem at least as reasonable to postulate that environment quickly accomplishes its maximal effect, and if constant thereafter, is powerless further to add or detract.

14. Over-Simplification of the Mendelian Theory

The attempts made during the last twenty years to fit to the observed facts of family resemblance in mental traits a theory which would account for the mechanics of hereditary transmission are probably premonitions of immensely important and serviceable developments in the eugenics of the future. It may truthfully be said, however, that these attempts as yet have done little more than scratch the surface of the problem, and that the chief value of

---

[18] Smith, K., *Biometrika*, 14:1922, 1-22.

some of them is in stimulating controversies that may result in further work along more productive lines.

The studies that have aroused so much criticism are of the type which have investigated family resemblance on some highly complex, continuously graded trait like intelligence, insanity,[14] nomadism, or musical talent, have attempted to categorize it as present or absent, and have made counts of its appearance in several generations to see whether this can be made to fit any of the simple Mendelian ratios.

The surprising thing about the results of such experiments is that the traits often *appear to behave almost as though they actually were Mendelian unit characters,* although 'unit character' and 'continuously graded trait' would seem to constitute an emphatic contradiction in terms. The agreement between experimental data and simple Mendelian ratios is never perfect, but is sometimes well within errors of sampling, and is often marked enough to make a simple Mendelian hypothesis appear quite reasonable were it not for the contradiction just noted.

As a matter of interest, I have performed some calculations to ascertain whether a hypothesis of cumulative Mendelian genes (which certainly would be more in accordance with observed distributions of mental traits than a 'unit character' hypothesis) might not be made to give as good approximations to simple Mendelian ratios as those reported in investigations of mental traits by Goddard, Davenport, Hurst, Rosanoff and Orr, Cobb, Peters, *et al.* It is not asserted that the hypothesis utilized in the following calculations is a correct one; it is presented merely to show that a more probable mode of hereditary transmission than that considered by some of the 'simple Mendelian' advocates can account for the ratios of two arbitrarily designated types of offspring about as well as can the over-simplified theory, and at the same time account for trait distributions in a general population far better.

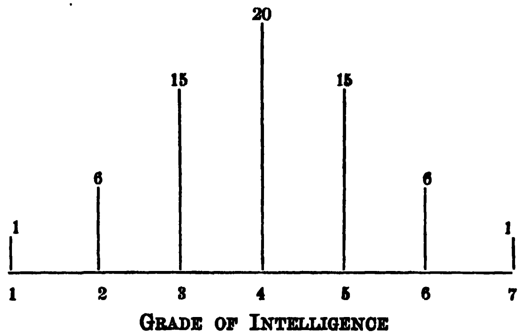
Assume (1) that intelligence is due to three cumulative pairs of genes, Aa, Bb, Cc, showing neither dominance nor epistacy, (2) that the two phases of these genes have equal incidence in the population, (3) that all genes make equal contributions to the trait

---

[14] Strangely enough, all forms of insanity have sometimes been lumped together.

intelligence, and (4) that mating is at random. (The latter assumption was probably unnecessary to make as far as final 'Mendelian ratios' were concerned; it was made simply because the calculations would otherwise have become very unwieldy.)

The resulting population showed a distribution of seven grades of intelligence in the proportions shown in the following figure.



Assume now that the two lowest grades of intelligence correspond to 'feeble-mindedness' and the five other grades to 'normality.' The numbers of offspring of various genetic formulas were computed for parent combinations covering the entire gamut of random mating, assuming four offspring to the mating; and the probability was allowed for that in a definite proportion of families for which the expectation would be three 'normals' to one 'feeble-minded,' half-and-half, etc., all four offspring would turn out to be normal or feeble-minded. Goddard's scheme, described in his book *Feeble-Mindedness,* was used for determining whether or not parent 'formulas' in certain matings should be called simplex, or indeterminate for 'normality.'[15]

Goddard's notation for designating types of mating is used, and a few figures from his study on *Feeble-Mindedness* are compared with our hypothetical ones.

NN designates the duplex normal individual, NF the simplex normal, and FF the nulliplex, or feeble-minded.

---

[15] A normal parent is called 'simplex' if his mating results in any feeble-minded offspring, and 'indeterminate' if his mating results in no feeble-minded offspring, since in the latter case he might conceivably be either simplex or duplex.

Type of Mating	Offspring Feeble-Minded (Percent)	Offspring Normal (Percent)
FF — FF		
Artificial population	81.6	18.4
Goddard	98.7	1.3
Simple Mendelian expectation	100.0	0.0
FF — NF		
Artificial population	47.3	52.7
Goddard	57.3	42.7
Simple Mendelian expectation	50.0	50.0
NF — NF		
Artificial population	28.1	71.9
Goddard	31.9	68.1
Simple Mendelian expectation	25.0	75.0

The only place where the ratios of 'feeble-minded' and 'normal' offspring in the writer's artificial population differ radically from Goddard's factual data is in the FF — FF type of mating where the artificial parents seem to produce too many normal offspring. If the number of genes and line of demarcation between normal and feeble-minded postulated for the artificial population should be altered by trial, it is probable that better agreement could be found at this point. The thing to be emphasized here is that the right combination of two feeble-minded parents can occasionally produce normal offspring if cumulative genes account for intelligence, whereas this would not be possible by a simple Mendelian theory. Goddard's 1.3 percent normal offspring, which he attempted to explain through occasional inaccuracies in the estimated intelligence of individuals entering his count, may thus be the necessary consequence of the actual conditions of inheritance.

II. Supplementary Considerations

It might not be out of place to suggest what a few needed avenues of approach to the nature-nurture problem appear to be, and to indicate as succinctly as possible some facts and possibilities that could well be kept in mind in planning new work.

1. Further Studies on Specific Effects of Specific Influences

We need more studies of the type which this Yearbook contains: accurate determinations of the effect of diseases or of various physical conditions upon mental development and the ability to do school work; careful evaluation of the effect of specific elements of home or school environment as differentiated from general excellence of either; measurement of racial differences with the possible effects of environment eliminated through experimental control; studies, employing rigorously matched control groups, to measure the week-by-week improvement on mental tests and achievement tests of foreign-speaking subjects as they learn English; the effect of subtle factors of personality upon the I.Q. and upon the use to which the I.Q. is put, etc. Such investigations would offer a life time of work to a corps of the best research workers that psychology, biology, education, and sociology could produce.

Of all the influences enumerated above, perhaps that of home environment is the most important to understand and evaluate. Professor Terman has proposed an experiment which would give very clear and conclusive data upon the effect of home environment on the I.Q. and would be entirely feasible, though very costly. Briefly, the experiment is: "From several hundred families of the grade whose offspring ordinarily yield a mean I.Q. of 80 (say, families of low-grade unskilled laborers), take 500 children as soon as they are born, and, after subjecting them to ten years of superior educational and cultural influence, compare their I.Q.'s with those of 500 other children from the same families who have not had these advantages. The mean I.Q. difference found would measure the combined effect of the environmental opportunities enjoyed by one group and denied to the other."

Eventually, too, an elaborate synthesis will have to be made. It must be determined whether or not the effects of various factors, singly studied, combine additively, or whether some effects are swallowed up in the presence of greater ones. For example, if poliomyelitis should lower the I.Q. an average of 15 points, and if lethargic encephalitis should lower it 10 points, would subjects who had suffered from both diseases be retarded 25 points, or would all the possible damage have been done by the first disease?

2. Nature-Nurture Contributions to Other Traits

It is probably not putting it much too strongly to say that we know next to nothing about the way character traits, interests, attitudes, ambitions, and special ability traits of all kinds are conditioned. Farnsworth, for example, in a summary appearing in Part II of this *Yearbook,* shows what an unsatisfactory state our knowledge of the inheritance of musical talent is in, and this in spite of the fact that more studies have undoubtedly been made which seek to uncover the determiners of this talent than of any other kind of quality just enumerated.

The *Yearbook* offers in Part II exploratory studies by May and Hartshorne on honesty, and by Anderson on mechanical ability. It is to be hoped that these two chapters may help to stimulate an extensive program of research in the field of personality and of special ability traits. There are many psychologists and practical educators who would put temperament and special bent even beyond I.Q. or E.Q. in importance to the individual and to society.

3. Mechanics of Mental Heredity

If one of our ultimate goals is the accurate control of mental endowment through eugenics, it is not enough to know the proportional contributions of nature and nurture to ability, nor even to know the specific effects of given amounts of various influences, such as parental intelligence, home environment, schooling, etc., upon ability. To be able to predict with any assurance what type or types of offspring to expect from any combination of parents, it is necessary to get in some way at the genetic constitution of the parent generation, to find out whether or not mental traits are transmitted by Mendelian factors, and if so, through interminable research, slowly to identify these factors, locate them in the chromosomes, and determine their degrees of dominance.

Practically the only studies which have approached mental heredity from the point of view of genetics are of the type mentioned in an earlier section of this paper where it was pointed out that a view of Mendelian heredity far too narrow to fit the observed facts had been employed.

However, a few theoretical studies deserve mention in which the attempt has been made to develop broad generalized hypotheses of Mendelian inheritance in 'continuous' traits and to compare the 'expected' correlations between relatives with the correlations found experimentally as a check upon the hypotheses. Although these studies have employed data from physical, rather than from mental traits, they supply the cues for investigations of mental heredity.

Pearson,[16] turning in 1904 to the newly rediscovered Mendelian theory, and assuming that a trait was due to N cumulative genes showing perfect dominance, calculated that under random mating the correlation between parent and offspring would be .33. As this value was out of harmony with the coefficients that had been found for parent-child measurements on a large number of continuous traits, Pearson definitely turned away from the theory which he had found inconsistent with experimental results, and never again made any serious attempt to interpret his own data by means of it.

It was subsequently pointed out by Yule,[17] and later by others, that parental correlations higher than .33 and quite consistent with actual values would follow from a generalized Mendelian theory if complete dominance in genes were not assumed.

It remained for Fisher[18] to work out a scheme by which the correlations between parents and offspring, between siblings, and between fathers and mothers are used to infer a coefficient of environment, "*the ratio of the variance* [of an unselected population on a continuous trait] *with environment absolutely uniform to that when difference of environment also makes its contribution.*" Fisher shows that when a trait is due to Mendelian factors, "the effect of dominance is to reduce the fraternal correlation (*i.e.*, its genetic value of one half) to only half the extent to which the parental correlation is reduced," and that this effect "is independent of the relative importance of different factors or of their different degrees of dominance." Making the important assumption that *environment works in a random manner*, thus reducing rather than raising correlations between relatives, and reducing fraternal correlations to the same extent as parent-child correlations, Fisher then utilizes the differences actually found between fraternal and parent-child correlations to distinguish between the effects of dominance and those of environment, finally arriving at the coefficient of environment defined above.

---

[16] *Phil. Trans.*, 203 A, 1904, 53-87.
[17] 1906 Conference on Genetics, *Horticultural Society's Report.*
[18] *Trans. Royal Soc. Edinburgh*, 52:1918, 399-433.

The difficulty of applying Fisher's scheme to family correlations in mental traits is that the effects of environment cannot here be assumed to be random. Demonstrably, environment works here to increase family resemblances rather than to decrease them, and it is probable that this increase is higher for siblings than for parents and offspring. Consequently, Fisher's method could not be applied to problems in mental heredity without fundamental modifications.

Another possibility not considered by Fisher (and as far as the writer knows, not by any other pioneers in the theory of hereditary transmission) is that *environment may have different degrees of influence when the endowment for a given trait is of larger or smaller amount.* It would seem reasonable to suppose that, when a large amount of a trait was present in the genotype, the possibilities of somatic fluctuation would be greater than when only a small amount was present.

This supposition is rendered quite probable in the light of results from a study of parent-offspring correlation by Davenport.[19] Although Davenport's study is concerned with measurements of height, rather than with mental measurements, it provides one of the best series available of family data on a complex trait and is considered here in the absence of adequate data upon any mental trait.

Davenport finds regression almost nil in the offspring of two very tall parents, but quite pronounced in the offspring of two very short parents, and intermediate for 'mid-parent' heights in between. He interprets these results as meaning that tallness is due to the *lack* of inhibiting factors, while shortness is due to inhibiting factors, and that the inhibiting factors are dominant, thus allowing for greater variability in the genotypes of very short parents than of very tall ones. This hypothesis would account for the greater amount of regression of offspring of short parents, but would be inconsistent with the nearly symetrical distribution of height universally found in unselected populations.

If, instead of Davenport's assumptions, a differential effect of environment were postulated, so that the somatic variability of the genotype increased in proportion to the number of inhibiting fac-

[19] *Eugenics Record Office Bull.* No. 18, 1917.

tors present, the phenomena of more variability and greater regression in offspring from short parents could be explained without the untenable condition that the dominant phases of all genes for stature are in the direction of shortness. The greater variability, under the writer's hypothesis, would be due to the fact that individuals from a number of variable genotypes would wander into any given phenotype more frequently and from more distant classes at the short end than at the tall end of the distribution. The offspring would then tend to revert to the original genotypes. Regression of offspring on mid-parent would occur, not because there was any 'urge' within the germplasm to return to mediocrity, but because more genotypes would wander into any phenotype from the direction of the mean than from the direction of the extreme of the population, simply because there *are* more and more individuals, the closer we approach the mean of any normally distributed population.[20] Using the 'artificial population' described earlier in this chapter, the writer computed how much regression in offspring would occur when certain fixed amounts of environmental 'susceptibility' per contributing gene were postulated, and found curves of regression plotted against mid-parent of the same general type as the one reported by Davenport for height—*i.e.*, pronounced regression at one extreme of the distribution, and no regression at the other extreme. It will be interesting, as a future problem, to find out whether data from mental traits also show such characteristic properties of mid-parental regression.

While the general agreement between Davenport's experimental results and the writer's theoretical calculations cannot be said to establish the theory, at least it suggests the expediency of further work in the same direction.

COMMENTS UPON "STATISTICAL HAZARDS IN NATURE-NURTURE INVESTIGATIONS"

By Truman L. Kelley

Miss Burks has rendered an important service in having very clearly pointed out a large number of treacherous pitfalls that lie in wait for the unwary pilgrim treading the narrow and obscure

[20] The writer is indebted to Professor L. L. Burlingame, Stanford University, for this latter suggestion.

path that winds its way between the eternal oaks of nature and the ephemeral shrubs of nurture. It is difficult to meet the request to comment briefly upon the many very pertinent issues raised in the foregoing chapter because most of the hazards of experimental and statistical research are matters of detail. Errors are very insidious in their entry into a study and in their effect upon conclusions drawn. For this reason comment concerned with the larger issues will generally fail to note the critical matters wherein danger lies, and such comments may be but an elaboration of the obvious.

An experimental study with its attendant statistical treatment may be thought of as constituting a logical argument wherein one false step vitiates the subsequent development. It has at times been considered analogous to a chain, the strength of which is only that of the weakest link. This analogy does not always hold, for in some experimental studies there are so many independent and semi-independent lines of evidence that there is not a single connection of premise with conclusion. The chain analogy should be replaced by that of a thread composed of many strands. When a study is of this type, each strand should be subjected to a thorough scrutiny and tested as rigorously as though upon it alone fell the entire weight of the argument. In the process of testing a line of argument one is almost certain to find assumptions that are only approximately true, such as, for example, "The correlation between height and weight is linear," or "the longer nurture acts, the greater is its influence," or "the distribution of 'leadership' in the case of random twelve-year-olds is normal," etc. Every such assumption is undoubtedly a source of error. Not uncommonly we may be sure that the error is not serious, but even so it is far better to continue to note that it has introduced some uncertainty than to dismiss it from mind. If there is a succession of such assumptions in a study, we can clearly see that their total effect will be such as is no longer pictured by the analogy to a chain or to a thread of many strands. An imperfect link in a chain, if stronger than the weakest link, is just as serviceable as though it were without fault, but a succession of weak steps in a scientific argument may terminate in a conclusion far weaker than any of these steps singly, for there may have been correlation between flaws. A chain is held at both ends, but an original psychological experiment should proceed from given data whithersoever the facts truly judged compel. Let us, therefore, adopt another picture. A tower built of cardboard cubes standing one upon another is weaker and less stable than any one of the blocks singly. If the heavy edge of each block is placed to the north, the tower will be much less stable than if the heavy edges are now north, now south, now east, now west. If, in a social science experiment, chance errors now work for and now against a certain conclusion, a certainty is attained which is clearly in excess of that built upon steps, the successive errors of which are correlated. The only safeguard is, first, to attempt to determine carefully the size or nature—chance or biased—of the error in each step, and second, to give careful consideration to the nature and net outcome of the summation of all errors.

Miss Burks' presentation, though admirably calling attention to many sources of error, seems to me not to lay sufficient stress upon chance errors, such as are commonly present in social science investigations. In the section dealing with partial correlation attention is called to the fact that the partial correlation coefficient $r_{12.3}$ is the result of partialing out too much when "we render constant, factors which may be in part or in whole caused by either of the two factors whose true relationship is to be measured, or by still other unmeasured remote causes which also affect either of the two isolated factors." This is true, but so far as presenting a picture between true variables, *i.e.*, between such as have no error factors in them, $r_{12.3}$ is also the result of partialing out too little, for we really should have partialed out the chance factors in the several variables.[1] Thus, if the variables are x_1, x_2, x_3, and if x_1 and x_2 contain, respectively, the chance factors x_4 and x_5, the partial correlation that people generally think they are getting when they get $r_{12.3}$ is $r_{12.345}$. With such measures as we must now frequently deal with, x_4 is ordinarily large with respect to x_1, x_5 large with respect to x_2, so that the numerical value of $r_{12.3}$ may be very different from $r_{12.345}$. I do not object to the point made in Section 3, with reference to partial correlation coefficients, when, as is seldom the case, it is known which variable or variables must

[1] Attention is called to the fact that Miss Burks, in the example she offers, has partialed out the chance factors, inasmuch as the correlations used by her were coefficients corrected for attenuation.—L. M. T.

be the cause and which the effect, but suggest that an even larger source of error in interpretation, that due to chance factors, has been overlooked.

There are two serious difficulties in interpretation of correlation coefficients, and it should be noted that partial correlation coefficients hold no different place in this respect from that held by ordinary total correlation coefficients. One speaks of the score of an individual on the XYZ Intelligence Test as the individual's general intelligence, whereas this is ordinarily in substantial error for two reasons. First, the score has a large chance factor in it, and second, the XYZ Test, in so far as it is not chance, may be a measure of something quite other than intelligence. In Section 3 there is given an illustration wherein one of the variables is "cultural status." I do not know what was the specific measure of cultural status employed, but knowing the difficulty of securing any measure of this type, I venture to suggest that probably about half of this measure is chance,[1] and probably about one-half of the half not chance is something other than cultural status as conceived by the modal American. One cannot partial out cultural status by partialing out a measure, one-fourth of which only, roughly, is entitled to the name. This probably sounds very elementary, as in fact it is, but why it should appear simple when dealing with partial correlation and not equally obvious when dealing with total correlation is a puzzle which must be left to the reader. Is it not clear that one cannot correlate social status with something else by correlating a measure, only one-fourth of which is social status? If the person labelling the social-status measure objects to the statement just made, and states that he explicitly wishes it understood that 'social status,' as he uses the term, means score on his social status scale, the situation is but slightly improved, for (a) it is extreme to incorporate a purely chance factor of large amount into a thing named in such a manner as not to imply chance, and (b) it is a severe, even if not impossible, requirement of a reader to ask him to discard an already established concept and build up another which is perhaps dependent upon the summation of scores or even of hundreds of items. "Social status" has been used merely as an illustration, for every mental measure is open to the same criticism.

[1] See footnote, p. 35.

In brief, though I agree that such interpretations of partial correlation coefficients and regression coefficients (and I wish total correlation coefficients had been included as well) as are criticised in Sections 3 and 4 are unsound, I doubt if such precautions as are suggested are sufficient. Rather—it seems to me—one should not attempt to reach such abstract conclusions. Why cannot experimentalists use measures in many connections; note the properties of each measure as shown by its reliability, as shown by age, race, sex, vocation, and other differences; calculate total, partial, and multiple correlations between these measures, now rich in meaning; and draw conclusions in terms of these specific measures and not in terms of unmeasured abstractions. Surely, we would be in an interminable morass if the relationships found between some hastily conceived "persistence of motives" or "honesty" test had to be accepted as giving us the true relationships of these traits. I believe that the agreement between the modal social concept corresponding to variously named character or mental traits and the traits measured by tests similarly designated will only run from 10 to 70 percent of the variance of the test employed. If I am correct in this, the point here raised is clearly of major importance.

In the discussion of the relative efficacy of different variables in estimating a criterion (Sections 4 and 5) is the statement: "We are justified in holding that a variable is effective in estimating a criterion in proportion to the amount of variance remaining in the criterion when all other variables are held constant, but the variable in question varies as much as before." The variance is equal to the standard deviation squared, so that the statement is equivalent to saying that the argument should be based upon the square of the standard deviation instead of upon its first power or any other power. Personally, I prefer this method to any other, but I would hesitate to say that this procedure is "justified" in any conclusive sense.

Section 13 deals with the cumulative effect of environment. Therein is found a criticism of the hypothesis: "It is fair to assume that the longer environment acts, the greater is its effect." This either is, or might well be, a quotation from the present writer. I have used the principle that the longer environment acts, the greater its effect, but only in connection with growing traits. In fact, I have not taken the environmental influence as proportional to the time through which it has acted, but proportional to the change, as judged by the growth of average individuals, that has taken place in the function during the time through which the environment has acted (see Kelley, *The Influence of Nurture Upon Native Differences,* page 16). This seems to me the most reasonable hypothesis to make with reference to growing traits. There are, of course, other situations wherein environment quickly accomplishes its maximal effect. For example, Jones is introduced to Brown and continues to know him for ten years. The sixty seconds in which the introduction took place may well be more potent than the entire remainder of the ten years in causing Jones to remember the name Brown. Here, however, after the first learning of the name, we no longer have a growing function.

Every hazard noted in the foregoing chapter is real, and unfortunately each has been overlooked time and again. A careful study of this chapter should make explicit, and should therefore lead to an avoidance of, many serious errors that now contaminate studies of heredity and environment.

PREFATORY NOTE TO CHAPTER III

This study has the three special merits of reliable and accurately scaled tests, large numbers of subjects, and data from two forms given a year apart (thus providing a reliability to use in correcting for attenuation). It has the drawback of limiting the subjects to a group of siblings in certain high-school grades, thus bringing in selective influences whose effect is difficult to measure, although the author has made a very reasonable estimate of some of these influences and has attempted to allow for them.

The conclusion is reached that the correlation, corrected for attenuation, between the intelligence of *unselected* siblings would not be far from .60. Comparing this value with the average sibling correlation, .52, found by Pearson for physical traits, which are free from ordinary environmental influence, the inference is drawn that home environment acts so as to raise the correlation between siblings from .52 to .60.

This inference must not of course be regarded as thoroughly established, since it depends upon the unproved hypothesis that mental abilities are transmitted by heredity in exactly the same manner as physical traits are. This may or may not be the case, and can only be ascertained through experimental means. Until definite data upon this point are available, the assumption that mental traits of siblings would correlate .52 if unaltered by environment must be considered as speculative.

CHAPTER III

THE RESEMBLANCE OF SIBLINGS IN INTELLIGENCE[1]

EDWARD L. THORNDIKE
and the Staff of the Institute of Educational Research,
Teachers College, Columbia University

This article will report measurements of the resemblance of siblings in the case of an intelligence examination made up of a selection of tests from standard instruments—the Institute of Educational Research Tests of Selective and Relational Thinking, Generalization, and Organization. The work has three special merits: The numbers which are the scores obtained in this examination form approximately a true scale of difficulty, as the successive differences between scores of 150, 160, 170, and so on, represent truly equal increments of altitude of intellect. In the second place, the populations are large so that satisfactory statistical precision can be obtained. In the third place, two alternative forms of the examination were given, approximately a year apart, so that correction for attenuation can be made and the resemblance can be computed which would be obtained if each of the individuals concerned had been subjected to an infinite number of such examinations. It has not been possible to attain the first and third of these advantages in previous studies of mental inheritance.

The work suffers from one notable demerit. The population concerned represents boys and girls attending the high school in Grades IX, X, and XI in 1922 and in Grades X, XI, and XII (plus a few pupils held back in Grade IX) in 1923, also pupils in Grades IX and X in September, 1924, and June, 1925. The results of this selection for our purpose are somewhat complicated.

First of all, since only the intellectually abler children progress to the high school, we have a tendency to select within a family two children both of whom are relatively intelligent. Obviously, for example, if one of the children tested had an idiot brother, that brother would not appear in our records. This selective action

[1] This investigation was made possible by a grant from the Commonwealth Fund.

tends to make all our measures of resemblance higher than they would be if we had a purely random selection within families.

It is the case further that any person in our lists will tend to be represented, in respect of the rest of his family, by brighter ones from his younger brothers and by duller ones from his older brothers. For example, suppose that we start with a boy of sixteen. If he has a bright brother of thirteen, that brother may have reached the high school and be in our lists. If he has a dull brother of thirteen, that brother will not appear in our lists. If he has a dull brother of eighteen (at the date of the first testing) that brother may appear in a later grade of the high school on our lists. If it is a very bright brother, however, he will have graduated from the high school.

This tendency to unfair selection is somewhat alleviated by the fact that we have records of all the pupils in the high school in Grades IX, X, and XI in June, 1922 and all those in Grades IX and X in September, 1924, but it still remains influential. It acts obviously to make resemblances lower than they would be if we had records within a family of all the children who ever went to the high school.

A further complication lies in the fact that, in order to correct for attenuation, we use only those individuals who had the two tests in June, 1922 and June, 1923 or in September, 1924 and June, 1925. Consequently, we are in every case measuring a group who remained in the high school for approximately one year or more. This is an accentuation of the general selective effect of going to the high school. We shall consider later the corrections probably necessary for the effect of these various selective factors.

Taking our group of siblings as it stands, we have to express the score of each individual as a divergence from the score which the average person of his age and sex would attain in this examination. Direct measurements of random samples of the male and female populations at ages 13, 14, 15, 16, 17, and 18 are not available for this examination or any similar examination; so we have to estimate them by somewhat roundabout methods, as described in the next few pages.

We have elsewhere[2] shown that the boys in the high schools of this city are in general slightly superior to the girls of the same

[2] *Pedagogical Seminary*, Vol. 33, pp. 167-181.

age in the ability measured by this examination. However, the superiority is very probably fully accounted for by the fact that one sixth more girls than boys attend high school. The selection of boys, being narrower, is presumably better. So we have not assumed any superiority of boys in general to girls in general at ages thirteen to twenty in this city.

We have elsewhere[3] shown that the gains in successive years from age 13 to age 19 in the ability measured by this examination are, for a group which enters the high school and stays two years, nearly the same. The gain drops about half a point for each year of age. The gain varies also with ability among individuals of like age. These facts are shown in Table I.

TABLE I.—THE PROBABLE AVERAGE ANNUAL GAIN IN SCORE IN THE I.E.R. TEST (SEL.-REL., GEN.-ORG.,) AFTER ALLOWANCE FOR DIFFERENCE IN DIFFICULTY OF FORMS AND FOR PRACTICE EFFECT, FOR SEVEN GROUPS OF VARYING ABILITY, MEASURED BY THE AGE IN GRADE IX (IN JUNE)

Age in Grade IX in June	12.5 to 13.5	13.5 to 14.5	14.5 to 15.5	15.5 to 16.5	16.5 to 17.5	17.5 to 18.5	18.5 to 19.5
12	15.0	13.5	13.0	12.5	12.0	11.5	11.0
13	14.5	13.0	12.5	12.0	11.5	11.0	10.5
14	14.0	12.5	12.0	11.5	11.0	10.5	10.0
15	13.5	12.0	11.5	11.0	10.5	10.0	9.5
16	13.0	11.5	11.0	10.5	10.0	9.5	9.0
17	12.5	11.0	10.5	10.0	9.5	9.0	8.5
18	12.0	10.5	10.0	9.5	9.0	8.5	8.0

The average pupil who stays at least two years in high school in this city is 14 years, 10 months, old at date of June first in his first year, and has then a median score of 174 in his first trial score with this examination. The probable first trial score for such a pupil at any age from 12 years, 0 months, to 20 years, 0 months, may then be computed from the gains shown in Table I, to be as shown in Table II.

The average of all children in this city would not, however, make scores as high as these. The high school selects the abler pupils. How much abler they are is not known with precision for this city or any similar city. We may attain a rough, approximate objective estimate by considering that an ordinary child of 14 years,

[3] *Jour. Educ. Psych.*, Vol. 14, pp. 513-516, and Vol. 17, pp. 73-76.

10 months, would be graduating in June, not from Grade IX but from the middle of Grade VII, and so lowering the figures of Table II by an amount which should on the average equal the amount of superiority shown in this test by a child whose grade is one and a half years ahead of his age. This amount may be com-

TABLE II.—AVERAGE FIRST-TRIAL SCORES IN I.E.R. TEST, FORM B, FOR PUPILS WHO CONTINUE IN THE HIGH SCHOOL FOR AT LEAST TWO YEARS

Year	Month											
	0	1	2	3	4	5	6	7	8	9	10	11
12	134	136	137	138	139	141	142	143	145	146	147	148
13	149	150	151	152	154	155	156	158	159	160	161	162
14	163	164	165	166	168	169	170	171	172	173	174	175
15	176	177	178	179	180	181	182	183	184	185	186	187
16	188	189	190	191	192	193	194	195	196	197	198	198
17	199	200	201	202	203	204	205	206	206	207	208	209
18	210	211	211	212	213	214	215	216	216	217	218	219
19	220	221	221	222	223	224	225	225	226	227	228	229
20	230											

puted from the facts of Table III. They show that pupils of the same age two years apart in grade are 51 points apart in score; hence, 39 points may be allowed on the average for a difference of a grade and a half. Since the abler individuals gain more rapidly than the less able, we may subtract increasing amounts as the age increases, beginning with 20 at 12.0 and ending with 50 at 20.0. The result is the set of values of Table IV. These may be

TABLE III.—DIFFERENCES IN MEDIAN SCORES IN I.E.R. TEST (SEL.-REL., GEN.-ORG.) OF PUPILS OF THE SAME AGE IN GRADES IX AND XI OR X AND XII

14-year boys in Grades XI and IX	30.4
14-year girls in Grades XI and IX	42.4
15-year boys in Grades XI and IX	62.3
15-year girls in Grades XI and IX	51.8
16-year boys in Grades XI and IX	42.1
16-year girls in Grades XI and IX	44.5
17-year boys in Grades XI and IX	52.3
17-year girls in Grades XI and IX	52.2
16-year boys in Grades XII and X	63.2
16-year girls in Grades XII and X	52.1
17-year boys in Grades XII and X	45.7
17-year girls in Grades XII and X	49.7
Median	50.7
Average	49.1

used as approximate central tendencies of the average child in this city at any age from 12.0 to 20.0 until more accurate determinations can be made.

TABLE IV.—ESTIMATED AVERAGE FIRST-TRIAL SCORES IN I.E.R. TEST, FORM B, FOR THE AVERAGE INDIVIDUAL OF CITY

Year	Month											
	0	1	2	3	4	5	6	7	8	9	10	11
12	106	107	108	109	110	111	112	113	114	115	116	117
13	118	119	120	121	122	123	124	125	126	127	128	129
14	129	130	131	132	132	133	134	135	136	137	138	139
15	139	140	141	142	142	143	144	145	146	146	147	148
16	148	149	150	151	151	152	153	154	154	155	155	156
17	156	157	158	158	159	160	160	161	162	162	163	164
18	164	165	166	166	167	168	168	169	170	170	171	172
19	172	173	174	174	175	176	176	177	178	178	179	180
20	180	181										

For the second-trial score we may make two similar tables, but adding 9 or 10 to each entry, since the practice effect of second trial over first trial is in general 11.9[4] and the A form (which was used in the second trial) was 2.4 points harder than the B form (used in the first trial). For obvious reasons of convenience we add 10.

We can now express the score of each pupil as a + or − deviation from the score which the average pupil who is found in Grade IX in this city in June, and remains for another year, would have at the age of the pupil in question, or as a + or − deviation from the score which the average boy in this city would have at the age of the pupil in question. The scores of C. S. and T. S., which were 15 yr. 5 mo., 166; 16 yr. 1 mo., 190; and 16 yr. 9 mo., 221; 17 yr. 9 mo., 210, respectively, become − 15, − 9, and + 24, − 7 as the deviations of the former sort, or + 23, + 31, and + 66, + 38 as deviations of the latter sort. These deviations in each case should next be expressed as multiples of the variability of the sex and age characteristic of the individual in question. We have elsewhere shown that the sex difference in variability is very slight; so this need not be considered.[5] The relation of variability in this examina-

[4] This practice gain may vary with age and ability, but how it varies therewith is not known; so we use the average for all.

[5] *Pedagogical Seminary*, Vol. 33, pp. 182-184.

tion score to age in the case of persons of 13, 14, 15, and so on, in this city is not known by direct measurement. We can, however, obtain accurate data concerning the variability in this test of 14-year-olds in the high school in June, 1922, and of the same individuals in a second form of the test a year later, and similarly for 15-year-olds, and 16-year-olds, and 17-year-olds. Except for the possible disturbing influence of the practice effect, these comparisons of variability are analogous to comparing the variability of all 15-year-olds with that of all 14-year-olds save those who died during the year, and similarly for the older ages.

The facts for the group tested are very clear. The variability is just the same in June, 1923, as it is for the same individuals in June, 1922.

For 220 boys of age 14 in June, 1922, the variabilities are: 29.6 in 1922; 28.8 in 1923 (median deviations).

For 378 boys of age 15 in June, 1922, the variabilities are: 34.8 in 1922; 35.3 in 1923.

For 375 boys of age 16 in June, 1922, the variabilities are: 32.4 in 1922; 32.1 in 1923.

For 156 boys of age 17 in June, 1922, the variabilities are: 30.2 in 1922; 31.3 in 1923.

The averages are: 31.8 for 1922, and 31.9 for the same groups a year later.

The differences in variability with age for the ages with which we are concerned are, then, so small that they may be neglected for our purpose.

We then measure sibling resemblance by the two correlations, using the two sets of deviations described above. The first takes as the zero of resemblance the resemblance of a pupil who stays in the high school in City A at least two years to any other pupil of the same sex and age who stays in the high school in City A at least two years. The second takes as the zero of resemblance the resemblance of any person in City A from 12 to 20 years of age to any other person of the same sex and age.

Our group of siblings is not a random selection of either of these two groups. The correlation obtained for our group of siblings in the first case will be lower than the correlations for all siblings, since we are working with a restricted range and are favoring, within that range, the inclusion of dull older sibs and bright younger sibs. The correlations obtained for our group of siblings in the second case will probably be higher than the correlation for all siblings, since the fact of attendance upon high school by both siblings means that specially dull sibs are not included. In the second case the correlation plot will be unlike those commonly shown, in that the zero points will not be near the means of the group of siblings, but far below them.

TABLE V.—CORRELATION OF SIBLINGS. EACH MEASURE IS EXPRESSED AS A DIVERGENCE FROM THE AVERAGE OF INDIVIDUALS OF THE AGE IN QUESTION AMONG THE HIGH SCHOOL POPULATION DESCRIBED IN THE TEXT

	−125 to −116	−115 to −106	−105	−95	−85	−75	−65	−55	−45	−35	−25	−15	−5 to +4	+5 to +14	+15	+25	+35	+45	+55	+65	+75	+85	+95	+105	+115	+125 to +134
−125 to −116							2			1																
−115 to −106			1		1				2				2	1												
−105		1		1		1			1	2		1	1		1	3										
− 95			1		1	1		2	2		3	3		2	1		1									
− 85				1			2		7	2	4		2	1	2	1										
− 75		1	1	1			1	2	3	1	2	2		3	1	2										
− 65	2				2	1		2	2	4	6	3	4	2	3	2										
− 55				2		2	2	2	3	1	3	4	7	3	4	2	4	1	1	1		1				
− 45		2	1	2	7	3	2	3		2	5	3	5	1	4	4	4	2	2					1		
− 35	1		2		2	1	4	1	2	8	8	7	2	2	6	6	1	3	2	2	4	1				
− 25				3	4	2	6	3	5	8	4	5	6	7	5	5	3	1	2	2			1			
− 15			1	3		2	3	4	3	7	5	8	5	7	7	3	1	4	3	2	1	1				
− 5 to + 4		2	1		2		4	7	5	2	6	5	6	9	8	5	9			2	2					
+ 5 to +14		1		2	1	3	2	3	1	2	7	7	9	2	4	11	5	6	6		4		2			
+ 15			1	1	2	1	3	4	4	6	5	7	8	4	2	6	4	3	3	3	3	1	1			1
+ 25			3		1	2	2	2	4	6	5	3	5	11	6	12	1	5	3	2	2	3	1			
+ 35				1				4	4	1	3	1	9	5	4	1	4	3	3	4	2	2	1	1		
+ 45								1	2	3	1	4		6	3	5	3	6	5	5	4	1	2	1		
+ 55								1	2	2	2	3		6	3	3	3	5		2	4	1	1		1	
+ 65			1					1		2	2	2	2		3	2	4	5	2	2	4	1				
+ 75										4		1	2	4	3	2	2	4	4	4	2		1		2	
+ 85								1		1		1			1	3	2	1	1	1		2	1		1	
+ 95											1			2	1	1	1	2	1		1	1				
+105									1								1	1							1	
+115																			1		2	1		1		
+125 to +134															1											2

There are also these pairs: −115, −155; 175, −5; and +165, +45.

TABLE VI.—CORRELATION OF SIBLINGS. EACH MEASURE IS EXPRESSED AS A DIVERGENCE FROM THE ESTIMATED AVERAGE OF ALL INDIVIDUALS OF THE AGE IN QUESTION IN THE CITY IN QUESTION

	−95 to −86	−85 to −76	−75	−65	−55	−45	−35	−25	−15	−5 to +4	+5 to +14	+15	+25	+35	+45	+55	+65	+75	+85	+95	+105	+115	+125	+135	+145	+155	+165	+175 to +184
−95 to −86														1														
−85 to −76							2				1		1	1														
−75						1		2						1						1								
−65				2	1				1		3		2	1	1	3	1				1							
−55				1		3			2	1	2	2	1		4			1										
−45			1		3				2	8	2	3	2	2	1		1											
−35		2						2	2	1	3	2	1		4	2	1											
−25			2				2		4	1	5	3	7	2	1	1	2						1					
−15				1	2	2	2	4	2	1	4	3	2	5	9	4	3	6	1		1							
− 5 to + 4					1	8	1	1	1		8	2	6	3	2	4	5	3	2	1	1	2			1			
+ 5 to +14		1		3	2	2	3	5	4	8	8	8	10	3	2	7	5	3		2	2	2						
+15					2	3	2	3	3	2	8	4	7	3	3	5	5	2	5	1	3	1						
+25		1		2	1	2	1	7	2	6	10	7	4	7	9	7	2	2	4				2					
+35	1	1	1	1		2		2	5	3	3	3	7	4	14	4	7	1	2	3	1	1	1	1	1			
+45				1	4	1	4	1	9	2	2	3	9	14	4	6	15	4	2	5	3	1		1				
+55				3			2	1	4	4	7	5	7	4	6	6	5	4	8	2	2	2				1		
+65				1		1	1	2	3	5	5	5	2	7	15	5	4	4	3	3	4	2	4					
+75					1				6	3	3	2	2	1	4	4	4	2	6	5		4	1	2				
+85									1	2		5	4	2	2	8	3	6		2	4	4	3					
+95			1							1	2	1		3	5	2	3	5	2		3	4	2	1	1			
+105				1					1	1	2	3		1	3	2	4		4	3		1	1					1
+115										2	2	1		1	1	2	2	4	4	4	1	2			1			
+125								1					2	1			4	1	3	2	1		2		2			
+135														1	1			2		1					1			
+145										1				1						1		1	2	1				
+155																1										2		
+165																												
+175 to +184																					1							

There are also these pairs: +35, −125; −85, −105; +195, +85.

The two correlation plots are shown as Tables V and VI. In each the average of the two scores is used for each individual. Each pair of sibs is entered twice, (− 4) (− 46) appearing as − 46 under − 4, and also as − 4 under − 46. The coefficients computed from these tables are .40 ± a P.E. of .026, and .66 ± a P.E. of .018. Correction for attenuation due to the fact that the average of the

TABLE VII—CORRELATION OF SIBLINGS IN CITY K. EACH MEASURE IS EXPRESSED AS A DIVERGENCE FROM THE AVERAGE OF ALL INDIVIDUALS OF THE AGE IN QUESTION AMONG THE HIGH SCHOOL POPULATION DESCRIBED IN THE TEXT

	−95 to −86	−85 to −76	−75	−65	−55	−45	−35	−25	−25	−5 to +4	+5	+15	+25	+35	+45	+55	+65	+75	+85	+95	+105	+115	+125	+135	+145	+155	+165	+175 to +184
−95 to −86																		1										
−85 to −76													1		1													
−75				1						2		1		1														
−65			1			1	1		1	1		2		3														
−55					4	2	2	2	5			1	2	2	1		1											
−45				1	2	2	2	1	2	3	2	2	1	2	1	2	2		1	1	1							
−35				1	2	2	2	3	2	5	3	4	7	1	1	3	2	1										
−25					2	1	3	4	7	6	7	4	6	3	5	1		4		2	1	1						
−15				1	5	2	2	7	2	5	7	6	8	5	4	5	6		1									
−5 to +4			2	1		3	5	6	5	8	5	8	6	8	7	5	7	7		1	2	2						
+5 to +14						2	3	7	7	5	6	5	8	6	13	6	6	4	5		1		1			1		
+15			1	2	1	2	4	4	6	8	5	8	13	11	11	14	7	5	2	4		2	2					
+25		1			2	1	7	6	8	6	8	13	6	12	13	17	15	10	5	5	5	2						
35			1	3	2	2	1	3	5	8	6	11	12	12	6	13	9	16	3	5	5	4		1				
45		1			1	1	1	5	4	7	13	11	13	6	8	25	7	7	12	3	3	3	2					
55						2	3	1	5	5	6	14	17	13	25	16	12	6	6	4	12	5					1	1
65					1	2	2		6	7	6	7	15	9	7	12	10	15	8	7	4	3	4	2		1		
75	1						1	4		7	4	5	10	16	7	6	15	14	10	12	4	2	2	2	3			
85						1			1		5	2	5	4	12	6	8	10	14	3	8	1	4	3	1		1	
95						1		2		1		4	5	5	3	4	7	12	3	16	6	4	1	1			1	
105						1		1		2	1		5	5	3	12	4	4	8	6	10		3				1	
115								1		2		2	2	4	3	5	3	2	1	4				2	1		1	1
125											1	2			2		4	2	4	1	3							
135														1			2	2	3	1		2		4				
																		3	1			1			2			
											1						1									2		
																1			1	1	1	1						
175 to 184																1						1						

There is also one pair at +205, +115.

two scores is still not a perfect measure of the ability in question raises the .40 to .435 and the .66 to .70.[6]

[6] The correlations for first trial score with second trial score are .852 and .902, respectively; hence, the correlations of the average of the two scores with the average of the scores for two other forms of the test, had such been given, would be, by the Spearman formula (commonly attributed Brown formula), .920 and .948.

We have checked these results by computing the resemblance of 823 pairs of siblings in another city. Each was tested with Forms A and B of the I. E. R. Tests of Selective and Relational Thinking, Generalization, and Organization at dates a year apart. Each individual's score was his deviation from the estimated score for the average person of his age in the city. The general intellectual status of youth in the city was assumed to be the same as in the city first studied.

The resulting correlation plot is shown in Table VII. The coefficient is .73, compared with .66 obtained for the city first studied. Even if we allow generously for a possible superiority of this city in general intellectual status, the coefficient will be over .68 raw, and over .72 when corrected for attenuation.

Fraternal resemblance in the ability measured by a large number of alternative forms of this examination, then, almost surely lies between .41 (.43½ − 1 P.E.) and .72 (.70 + 1 P.E.).

Consider now the estimate of .70. This is surely too high because of the selection of children both of whom enter the high school and remain at least two years (or one year in the group entering in 1924). It would also be unduly high if our estimated averages for all individuals in the city are too low. Our estimates were such as to make the average or median of all individuals coincide with the 21st percentile of the group of individuals who remained through two years of high school (in about two-thirds of the cases) or one year of high school (in about one-third of the cases).

From the Army data on Alpha, it can be computed (*Memoirs*, p. 748) that the median Alpha score for the native born white draft is the 12th precentile score for those of the native born white draft who reported themselves as having completed at least the second year of the high school.

From the data of Cobb[7] it can be computed that the 20th percentile score of high-school sophomores in Illinois, Wisconsin, Iowa, and Michigan about 1920 is 90, which is 25 above the median adult score of the draft in 1918, 25 above the median adult score of the native whites of the draft who reported themselves as having completed Grade VIII, but no further grade, and 11 above the median for the draft from the highest state.

---

[7] *Jour. Educ. Psych.*, Vol. 13, pp. 449-464 and 546-555.

It appears, therefore, that there is fully as much likelihood that our estimates of average scores for the general population of the city at ages 13 to 20 are too high as that they are too low, and that there is no reason to reduce the correlation from .70 so far as concerns the zero points from which ability is measured. We are unable, from data in our possession or easily obtainable, to discover empirically what reduction would be made from .70 if a random sampling of sibs were used instead of those measured by us. Some light may, however, be gained by comparing the resemblance in the case of pairs, one sib of which was very bright, with the resemblance in the case of pairs one sib of which was below or only moderately above the average ability for his age. In the first case the individuals in our list represent nearly a fair sampling of all pairs of sibs one of whom is very bright, since with a correlation as low as .50 these very bright children will very rarely have sibs eliminated for reason of dullness before Grade X. In the second case the individuals in our list may represent an undue proportion of the bright sibs as compared with the dull sibs of these moderately bright children. The comparison is complicated by the chance error of the measurements and distorted by the other effect of selection whereby the duller from older sibs and the brighter of younger sibs are unduly selected by reason of the time conditions of the testing; but it will prevent our estimate from being a mere intuition.

To make this comparison, we compute r by $\sigma_{array}=\sigma_{dis}\sqrt{1-r^2}$ for each array, and obtain a weighted average of the r's from (1) arrays where one sib scored + 95 or over, (2) arrays where one sib scored + 55 to + 94, from arrays where one sib scored + 15 to + 54, and for arrays where one sib scored − 25 to + 14. For our purpose it seems best to attach weights in proportion to the number of cases. It is also best to measure the deviations, not from the actual means of the arrays, but from the most probable central tendency for each array in view of the relation line as a whole. The actual means are + 1.7, + 3.6, + 3.3, + 2.2, + 3.4, + 2.5, + 3.6, + 3.9, + 4.3, + 4.9, + 5.7, + 7.0, + 7.25, + 6.0, + 7.3, + 8.3, + 9.0, 11.6 for the thirteen cases of the next two arrays combined, and 11.6 for the five extreme pairs of the five extreme arrays; so we use as probable true means for the arrays under − 25, − 15,

− 5, + 5, etc., 1.5, 2 −, 2 +, 2.5, 3 −, 3 +, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, and 12. The weighted average σ's for the arrays are as follows:

+ 95 and over	44.3
+ 55 to + 94	44.8 −
+ 15 to + 54	45.6
− 25 to + 14	44.8

By these determinations the resemblance is then a little closer for the pairs containing one very bright sib than for those containing one relatively dull sib. It therefore seems that if all pairs of sibs in this city had been tested at or near the age of 15, the correlation would not have been very much lowered below what it was for our group. The exclusion of pairs where both were dull or both were of moderate ability, and the excess inclusion of bright young and dull old sibs of ordinary or slightly superior sibs, in part counter-balances the general tendency of the high school to select superior children, and so to select unduly pairs where both are superior as compared with pairs where one is superior and the other inferior. So an allowance of .10 seems reasonable. If we had all the sibs of the ages in question in the city, the correlation (corrected for attenuation) would probably be nearly .60.

We may check this estimate in the following manner:

If we estimate that the individuals of our lists represent one fourth of the population of comparable ages and that our lists contain 90 percent of the top one percent (in intellect) of the population of those ages, the σ of the whole population for any one month age in average score in two trials of the test will be 53.5. Using this estimate of σ for an unselected age population, r for our cases of sibs where one sib scores + 95 or over is .56 (before correction for attenuation) or approximately .60 after correction for attenuation.

On the whole, then, we conclude that the resemblance of siblings in the ability measured by the I. E. R. Sel., Rel., Gen., Org., test is near .60.

If we may accept Pearson's results for the resemblance of siblings in eye color, hair color, and cephalic index (.52, .55, and .49), and regard .52 $\pm$ .016 as the resemblance in traits entirely free from environmental influence, we may infer that *the influence upon intelligence of such similarity in environment as is caused by being siblings two to four years apart in age in an American family today is to raise the correlation from .52 to .60.*

PREFATORY NOTE TO CHAPTER IV

The material summarized in this chapter represents the first mental-test data upon parents and their children gathered in sufficiently large quantities to establish parental correlations with fair precision. The importance of supplementing and checking, by studies employing reliable mental tests, what has been found in the earlier biometric *trait rating* studies of family similarities has long been recognized. Up to the time of Willoughby's study (which was started in 1923) a number of investigations of *sibling* resemblance had been made (see section on family resemblance in the chapter entitled "A Summary of Literature"); but no comprehensive mental-test study of *parent-child* resemblance had been reported. Willoughby presents both parent-child and sibling correlations. The latter can be compared with results in previous published literature as well as with the results of several Yearbook studies; the former can be compared with the early non-mental-test studies, as well as with the parent-child correlations based upon tests reported in this Yearbook by Jones (Chapter V) and by Burks (Chapter X).

CHAPTER IV

FAMILY SIMILARITIES IN MENTAL-TEST ABILITIES[1]

RAYMOND R. WILLOUGHBY
Clark University, Worcester, Massachusetts

A battery of verbal and non-verbal group tests was administered to parents and children of 141 families living in the vicinity of Palo Alto. Through the courtesy of the schools, all families were listed (total number, 214) having a child who would be between 12½ and 13½ at the estimated middle of the testing period for the vicinity in question. Tests were obtained upon about 141 such children, 90 fathers, 100 mothers, and 280 siblings (who apparently ranged in age from 7 up).

The tests used were taken from standard group intelligence and achievement tests, and comprised the following:

1.	Opposites	(N. I. T.)
2.	Number-series completion	(Army Alpha)
3.	Arithmetic reasoning	(N. I. T.)
4.	Symbol-series completion	(Army Beta)
5.	Sentence meaning	(Stanford Achievement)
6.	Geometric forms	(Army Beta)
7.	Analogies	(N. I. T.)
8.	Symbol-digit	(Army Beta)
9.	Science-nature information	(Stanford Achievement)
10.	History-literature information	(Stanford Achievement)
11.	Checking similarities	(N. I. T.)

The tests taken from the National Intelligence Tests were made more difficult by the addition of other items, chiefly from the Terman Group Test of Mental Ability.

Standard (z) scores were derived for subjects of all ages—from 7 to 60—by constructing (a) growth curves for each of the 11 tests, and (b) curves showing the change in group variability (S.D.) with

[1] This study is a doctorate dissertation completed at Stanford University in 1926 under the direction of Professor Kelley and myself, and not published at the time this summary was made for the Yearbook by Barbara Stoddard Burks.—L. M. Terman.

age upon each of the 11 tests. According to the author, the groups upon which the curves were constructed "varied in number from about 30 individuals through the fast-growing years of childhood to about 60 where the curves were not so steep."

Results were reported as (a) raw correlations of family resemblance upon the separate tests, and (b) correlations corrected for attenuation upon the separate tests. To obtain reliabilities for use in treatment (b), the tests were divided into odd-numbered and even-numbered halves; z scores on the two halves were correlated; and the Spearman-Brown correction formula was applied. The tables here are copied or adapted from tables appearing in the original study.

The computation of probable errors for the correlations of Table II was rendered somewhat difficult by the fact that certain individuals entered the correlation tables more than once (as parents of more than one child or as children having more than one sibling). Owing to a correlation between errors of sampling under such circumstances, the N to use in the P.E. formula becomes less than the total number of pairings (though it remains more than the *smallest* number of individuals whose scores constitute either one of the correlated variables). By empirical methods the geometric mean between the possible upper and lower limits of N is found to yield the most reasonable value of the P.E., which is almost surely correct in every case within one point in the second decimal.

In the discussion of results the following observations made or suggested by the author are of especial significance or interest:

1. The average corrected correlations, fraternal and parental, are .42 and .35, respectively.

2. On the verbal tests the average corrected correlations, fraternal and parental, are .43 and .39; and on the non-verbal tests, .41 and .30.

3. On the separate tests the fraternal correlations run from .48 (history-literature and also symbol-digit) down to .32 (geometric forms). The paternal correlations run from .48 (history-literature) down to .17 (checking similarities).

4. "In the individual tests, for the most part, the sex differences are negligible, but there are a few striking exceptions, such as the superiority of the male fraternal to the female in the verbals,

TABLE I.—THE UNCORRECTED FAMILIAL COEFFICIENTS AND THE RELIABILITY COEFFICIENTS USED IN THEIR CORRECTION

(MS indicates the uncorrected (mother-son) coefficient; M will here indicate the reliability of the mother group entering into this particular correlation, and S, similarly. YB stands for younger brother, OB for older brother, etc.)

Test	1	2	3	4	5	6	7	8[1]	9	10	11
Mother-Son											
MS	.24	.33	.23	.41	.41	.29	.19	.25	.43	.48	.17
M	.86	.85	.80	.91	.89	.76	.87	.87	.95	.93	.92
S	.82	.72	.65	.83	.83	.71	.74	.80	.93	.90	.83
Mother-Daughter											
MD	.23	.37	.34	.23	.28	.18	.43	.20	.43	.40	.20
M	.90	.89	.87	.84	.74	.74	.91	.87	.95	.93	.89
D	.77	.78	.79	.75	.82	.63	.69	.77	.91	.78	.85
Father-Son											
FS	.36	.32	.16	.31	.33	.14	.38	.32	.40	.49	.13
F	.93	.85	.78	.92	.79	.77	.91	.88	.92	.97	.91
S	.74	.74	.49	.86	.81	.68	.73	.77	.92	.87	.81
Father-Daughter											
FD	.38	.38	.41	.23	.11	.20	.33	.25	.19	.32	.02
F	.94	.84	.79	.92	.84	.78	.84	.88	.95	.97	.89
D	.81	.82	.79	.83	.81	.72	.79	.79	.91	.73	.77
Brother-Brother											
BB	.45	.45	.37	.31	.47	.29	.31	.30	.53	.47	.16
YB	.83	.73	.80	.75	.86	.85	.82	.84	.94	.93	.89
OB	.88	.88	.67	.79	.78	.80	.90	.85	.96	.92	.89
Sister-Sister											
SS	.29	.46	.26	.31	.21	.28	.30	.48	.38	.44	.42
YS	.66	.78	.89	.74	.70	.65	.57	.73	.86	.73	.70
OS	.80	.82	.72	.82	.89	.81	.76	.80	.90	.83	.70
Brother-Sister											
BS	.30	.31	.37	.31	.28	.16	.28	.36	.23	.31	.27
B	.81	.86	.58	.82	.78	.70	.79	.80	.94	.84	.88
S	.75	.85	.80	.83	.83	.61	.80	.80	.92	.79	.83
Husband-Wife											
HW	.34	.17	.34	.50	.41	.25	.26	.56	.43	.55	.41
H	.92	.75	.68	.92	.84	.73	.88	.85	.95	.96	.90
W	.90	.94	.98	.91	.88	.79	.86	.88	.92	.94	.92

[1] As Test 8 is not susceptible of division into equivalent halves, the reliabilities used therein are averages of those for the remaining ten tests. The essential soundness of this procedure would seem to be indicated by the fact that the average of the reliabilities of the child-groups above differs by less than .005 from that (.788) determined for a large population, in three school grades, on a closely related symbol-digit test of the same number of elements

TABLE II.—FAMILIAL COEFFICIENTS CORRECTED FOR ATTENUATION

Test	1	2	3	4	5	6	7	8	9	10	11	Verbal Aver.	Non-Verbal Aver.	Total Aver.
Parentals														
Maternals														
Male, r	.28	.42	.32	.47	.48	.39	.23	.29	.46	.52	.19	.38	.35	.37
$P.E._r$	.07	.07	.07	.06	.06	.07	.08	.06	.06	.05	.07			
Female, r	.27	.44	.41	.29	.31	.27	.52	.24	.47	.47	.33	.41	.31	.37
$P.E._r$	.07	.06	.07	.07	.06	.08	.06	.07	.06	.06	.06			
Paternals														
Male, r	.41	.37	.25	.35	.40	.20	.47	.39	.43	.53	.15	.41	.29	.36
$P.E._r$	.07	.07	.09	.07	.07	.08	.07	.07	.06	.06	.08			
Female, r	.44	.46	.51	.27	.13	.27	.40	.29	.20	.38	.02	.34	.26	.31
$P.E._r$	.07	.07	.07	.07	.08	.07	.08	.07	.07	.07	.08			
Aver. Parentals	.35	.42	.37	.35	.33	.28	.41	.30	.39	.48	.17	.39	.30	.35
Fraternals														
Male, r	.52	.56	.50	.39	.58	.35	.36	.36	.55	.51	.18	.50	.37	.44
$P.E._r$	.07	.08	.09	.09	.07	.09	.09	.09	.07	.07	.09			
Female, r	.40	.58	.32	.39	.26	.38	.45	.63	.43	.56	.60	.40	.52	.45
$P.E._r$	.10	.07	.09	.09	.09	.09	.11	.07	.08	.08	.07			
Mixed, r	.38	.36	.54	.37	.35	.24	.35	.45	.25	.38	.32	.38	.35	.36
$P.E._r$	.07	.06	.07	.06	.07	.08	.07	.06	.07	.07	.07			
Aver. Fraternals	.43	.50	.45	.38	.40	.32	.39	.48	.41	.48	.37	.43	.41	.42
Marital														
r	.37	.20	.47	.55	.48	.33	.30	.65	.46	.58	.45	.44	.44	.44
$P.E._r$	.07	.08	.08	.05	.07	.05	.08	.04	.05	.06	.06			

vice versa in the non-verbals, and the superiority of both to the mixed (brother-sister)."

In an attempt to bring together the parental and fraternal correlations in a single, unified interpretation, the author has included a brief note upon Fisher's "coefficient of environment."[2] Quoting Willoughby: "This is defined as 'a constant equal to the ratio of the variance (σ^2) with environment absolutely uniform to that when difference of environment also makes its contribution. Its

[2] Fisher, R. A., "The correlation between relatives on the supposition of Mendelian inheritance." *Trans. Royal Society, Edinburgh,* 52: 1918, 399-430.

derivation depends upon certain reasonable hypotheses and much intricate logic and cannot be reproduced here. Its value (which depends, apart from the derivation, only on the obtained correlations) is [using our average correlations of marital, .44; fraternal, .42; and parental, .35] .54. That is, granted the soundness of Dr. Fisher's argument, in the case of mental-test ability, environment operates not quite to double the variability as measured by the square of the standard deviation. Whether this is equivalent to saying that inheritance and environment are approximately equal determinants of mental-test ability is a question we only suggest here.''[3]

A note on the growth and decline of the abilities measured on the battery of tests introduces other considerations of importance. In the author's words:

''We observe first a type of mean curve in which the senescent decline carries the curve back to pubertal or pre-pubertal level, as opposed to one in which a large proportion of the early ability is retained. Pronounced examples of the former are the curves for opposites ability, symbol-series completion, analogies, symbol-digit, history-literature information, and checking similarities; in some of these cases (symbol-series completion, and symbol-digit substitution) the mean at 60 has regressed to the 10-year level. It is interesting that none of these except the history-literature information is a specifically school-trained ability. The marked example of 'retained ability'—arithmetic reasoning—is obviously school-trained. Intermediate are number-series completion, sentence meaning, geometric forms, and science-nature information. Differential selection or survival ought to be mentioned as a possible contributing cause in the long decline of both mean and standard deviation curves. It is possible, too, that both the remoteness of the older individuals' education and its more meager absolute amount may have some influence on the former.''

[3] Further comments upon the Fisher technique may be found in Chapter II (Statistical Hazards in Nature-Nurture Investigations). There would appear to be considerable doubt as to whether the conditions postulated by Fisher in deriving his methods are met sufficiently well in any type of mental test data to warrant the computation of the Fisher coefficient of environment. Fisher himself applied the technique only to data upon physical traits.—L. M. T.

PREFATORY NOTE TO CHAPTER V

This study is the most recent available contribution to the province of mental test resemblance in parents and offspring. Results which demand attention, and which are not easy to explain, are the consistently higher values found for parent-child resemblance upon raw intelligence scores in this investigation than in studies similar (though carried on with different mental tests) by Willoughby and by Burks (Chapters IV and X of this Yearbook). The raw correlations reported by Jones are about as high as the parent-child correlations corrected for attenuation in the Burks study, and higher than the corrected coefficients in the Willoughby study.

One fact mentioned in this chapter, but passed over with little comment, would seem well worth further investigation. This is a drop in median I.Q. from 94, at age 10, to 83 at age 14. The author attributes this drop to the "nature of the test, and not to differences in the sample." This is appreciably more than the age drop in I.Q. of the original subjects upon whom the Stanford Binet was standardized. It is to be hoped that the possibility of cumulative environmental influence upon the subjects of this study may sometime be investigated.

CHAPTER V

A FIRST STUDY OF PARENT-CHILD RESEMBLANCE IN INTELLIGENCE[1]

HAROLD ELLIS JONES
Director of Research
Institute of Child Welfare, Berkeley, Cal.

The resemblance of parents and children in certain physical traits (stature, span, forearm, eye-color) has been reported by Pearson[2] as averaging approximately .50 in terms of coefficients of correlation. Nearly identical coefficients have been obtained for parent-progeny resemblance in the coat color of animals[3] and in the body characteristics of insects and other lower forms.[4] The present report deals with the resemblance of parents and children in mental test traits. The statistical analysis is limited to correlations; other modes of treatment of the data will be undertaken in a later article.

SELECTION OF A TEST SAMPLE

In planning the locus of the study, weight was given to the following considerations:

1. The population examined should be relatively homogeneous in economic status and in educational opportunity. Environmental variables should approach a minimum, to the extent that this may occur under non-experimental conditions.

2. The population should be entirely native-born of native-born stock, in order to reduce factors of language handicap and of differential group traditions.

3. The population should be relatively stationary within a limited district, in order that later additional surveys can be made if desired. A stable group has the further advantage that its

---

[1] Acknowledgment is due to the Social Science Research Council of Columbia University in providing an allotment for the research; to Professor A. T. Poffenberger as representative of the Council; and to Dr. Mary Cover Jones and Mr. Herbert Conrad as assistants in the collection of data.

[2] *Biometrika*, Vol. 2, p. 358. *Philos. Trans.*, Vol. 195, p. 93.

[3] *Proc. Roy. Soc.*, Vol. 66, p. 154. *Jour. of the Anthrop. Instit.*, Vol. 33, p. 204.

[4] *Biometrika*, Vol. 1, p. 139. *Proc. Roy. Soc.*, Vol. 65.

earlier genealogy and the present family relationships can be more readily traced than in a group comprising many transients.

The district finally chosen for the survey included a block of 9 counties in central and north central New England. These were selected as offering representative rural conditions for the states of Massachusetts, New Hampshire, and Vermont. In the 193 townships (under 2500 population) of these counties, 19 villages were utilized for the purposes of the survey; they were considered, with their adjacent farming areas, to provide a fair sample of the total rural population of the nine counties.[5] The representativeness of these villages appears to be attested by all of the social and economic criteria which have been applied.

Approximately 2500 test records were obtained. The subjects, who ranged between the ages of 3 and 65, included two or more members of over 300 families.

TABLE I.—DISTRIBUTION OF 317 CHILDREN IN 105 FAMILIES

Number of Children	Number of Families	Number of Children	Number of Families
2	51	6	4
3	26	7	1
4	13	9	1
5	9		

TABLE II.—DISTRIBUTION OF 213 I.Q.'S BY STANFORD REVISION

I.Q. Intervals	Number of Children	I.Q. Intervals	Number of Children
50– 59	1	100–109	36
60– 69	8	110–119	22
70– 79	28	120–129	5
80– 99	60	130–139	3
90– 99	49	140–149	1

Median I.Q.	92.1 ± .87
Mean I.Q.	93.5 ± .70
S.D.	15.1

[5] The 19 centers had the following populations in 1920: 34, 110, 115, 118, 242, 289, 294, 307, 343, 399, 500, 565, 609, 811, 907, 1087, 1175, 1577, 1667. From 1910 to 1920 these communities suffered a population decrease of 5%, and from 1900 to 1910, of 4.1%. For rural Vermont as a whole, the corresponding figures are 5.7% and 4%, or practically identical. The 14th census does not give comparable data for the total rural population of Massachusetts and New Hampshire.

The present report deals with the results from the total number of families (105) in which both parents and two or more children were tested. The Army Alpha (Form 5 or 7) was used with the parents and with children above ten years of age. The Stanford Revision of the Binet-Simon Scale was used with the 213 children in the age interval from three and a half to fourteen years.

Characteristics of the Sample

The sample includes 527 individuals (210 parents and 317 children). Except in two or three families, in which there was an absence of a child from town, tests were obtained on all sibs above the age of three and a half years.

The figures for central tendency and variability of the 213 children are the same, within two decimal places, as the constants obtained for the larger sample of 375 children tested by the Stanford Revision in these towns. We have considered our group representative of New England rural children.

Table III.—Distribution of Army Alpha Scores of the 210 Parents

Score	Fathers	Mothers
160–179	3	4
140–159	11	5
120–139	5	12
100–119	9	9
80– 99	13	22
60– 79	15	21
40– 59	20	17
20– 39	20	12
0– 19	9	3
Median	65.2 ± 3.6	79.5 ± 3.2
Mean	73.4 ± 2.9	82.6 ± 2.6
S.D.	44.8	29.1

Evidence on the normality of this sample, with reference to the total group tested, will be given on a later page. The superiority of the wives, which is reliable, is probably due in some part to the relatively stronger selective influence of migration on the adult males.

No illiteracy was encountered, although two parents reported that they had had no formal schooling.[6]

[6] By the 14th Census, the percent illiteracy among rural native whites in Vermont is 1.5 for males and .7 for females. In Massachusetts and New Hampshire the figures are still lower.

TABLE IV.—AGE OF THE PARENTS

Age	Fathers	Mothers
60–64	2	..
55–59	6	1
50–54	17	10
45–49	14	12
40–44	23	17
35–39	16	23
30–34	23	24
25–29	4	15
20–24	..	3

TABLE V.—EDUCATION OF THE PARENTS

	Family Sample	Main Sample of 375 Adults
	Percent	*Percent*
Completed college	1.4	2.4
Entered college	4.9	7.5
Completed high school	20.3	27.2
Entered high school	47.6	53.6
Completed Grade VIII	85.3	86.4
Completed Grade VI	91.6	94.4
Completed Grade IV	97.2	98.1

Racial Origin of the Parents

All of the parents were native-born whites; the great majority were derived from old New England stock. No language other than English was spoken in any of the homes. The only considerable admixture of a non-English strain was in the case of about 12 percent of the parents, who showed French-Canadian derivation.

Occupations of the Fathers

Table VI shows the range of occupations in the family sample. The percentage falling in the various groups is closely similar to that of the total sample tested, and of the local male population as a whole.

When placed in the occupational classification used in the United States Census, it is found that our 105 fathers are distributed similarly to the adult males in the census records of Vermont (1920).

The remaining cases (18.1 percent in the family sample and 14.9 percent in the census population) are classified in "Transporta-

TABLE VI.—OCCUPATIONS OF THE FATHERS

	Family Sample 105 Males	Main Sample 277 Males
	Percent	*Percent*
Professional Group	5.7	4.6
Minister (2)		
Dentist (2)		
Physician, lawyer		
"Technicians"	2.9	4.6
Telephone engineer		
Scene painter		
Small Business Executives	9.5	10.8
Retail dealers (4)		
Postmaster (2)		
Mill owner		
Restaurant owner		
Section foreman		
Skilled Laborers	15.2	10.8
Mechanic (6)		
Carpenter (4)		
Painter (2)		
Stone cutter		
Leather worker		
Plumber		
Farmers	38.1	41.5
Semi-skilled laborers	11.4	15.1
Truckdriver (7)		
Mail carrier (3)		
Chauffeur, trainman		
Unskilled laborers	17.1	12.2
General laborer (8)		
Wood cutter (4)		
Teamster (3)		
Section hand (3)		

TABLE VII.—OCCUPATIONS CLASSIFIED ACCORDING TO U. S. CENSUS

	Family Sample 105 Males	U. S. Census Adult Males of Vermont
	Percent	*Percent*
Agriculture	40.0	44.1
Manufacturing and mechanics	29.5	29.0
Trade	6.7	8.8
Professions	5.7	3.2

tion'' and ''Public Service.'' It is evident that, with respect to the occupational cross-section, our group of fathers is representative of the rural communities studied. The widest range of economic status is of course found in the group of farmers, for these include both small farmers, cultivating no more than five or ten acres of unproductive hillside, and also land owners operating fertile farms of several hundred acres. The actual range in annual income and in standards of living is not so great as this would suggest, however. None of the families studied is near the borderline of dependence, and none would be considered moderately wealthy.

Community Status of the Fathers

In five of the larger communities ratings on ''social status'' and ''mental test ability'' were obtained on all of the adult males tested. In each town the rating committee consisted of three judges, usually a minister, town clerk, or some other official well acquainted with the town and its inhabitants. They were instructed to rate on a scale of 10, with 5 as ''average for your town.'' The significance of each point on the scale was stated as explicitly and concretely as possible.

TABLE VIII.—RATINGS OF 66 FATHERS

Ratings	"Social Status"	"Mental Test Ability"
	Percent	*Percent*
9	1	1.5
8	7.5	4.5
7	6	10
6	15.5	12
5 ("Average")	39	37
4	21	13.5
3	6	13
2	2.5	6
1	1.5	2.5
Median rating	4.90	4.97

Subject to the limitations of this rating method, we may conclude that the group of 66 fathers constitutes a fair sample of the adult males in the towns studied. Careful examination of their Alpha distribution and of the data on educational and occupational status fails to show any significant differences in this group as compared with the larger sample of 105 fathers. The technique employed in our test survey would be expected to yield an unselected (rural) population. From the various lines of evidence presented above, it would appear that this result has in a fair measure been achieved.

The Army Alpha scores and mental ages presented in Table IX cannot be used directly in determining coefficients of resemblance because of the influence of the age factor in increasing scores up to 16 and in depressing scores beyond the age of 45. In the case of the Stanford Revision material, the intelligence quotients are unfitted for our purposes, owing to the fact that in this rural population the median I.Q. drops steadily from 93.6 at 10 years to 82.7 at 14 years. This is due to the nature of the test, and not to imperfections in the sample.[7] In order to overcome these difficulties, and to make the Stanford and the Alpha records more directly comparable, each score was transmuted into a sigma score by the following method:

For our 375 Stanford Revision records, a median mental age was computed at each age interval, and a smoothed curve constructed by the formula $\frac{a + 2b + c}{4}$. A smoothed curve for the standard deviations was similarly constructed. It was then possible to make an interpolation table giving the mental age median and sigma for each month of chronological age from 3 years to 15 years. A mental age score was converted into a sigma score by the formula $x = \frac{X - Md_1}{\sigma_1}$ where x = sigma score, X = M.A., and Md_1 and σ_1 represent the median and standard deviation M.A. for the individual's chronological age. A similar age curve, from 10 to 65 years, was constructed from 1215 Army Alpha records obtained in the same communities. This use of sigma deviation scores involves certain theoretical difficulties, but so far as correlations are concerned it is not open to criticism for placing the Stanford Revision reference point in the same scale position as the Alpha reference point, since a discrepancy between these, unless

[7] A similar result has been reported by N. D. M. Hirsch, in a recent study of rural children in Kentucky. A decrease in the average I.Q. was found, from 81.0 at 8 years to 73.1 at 13 years.
Proc. Nat. Acad. of Sci. Vol. 13, 1927, 18-21.

inconstant, would neither attenuate nor inflate the correlation coefficients. The reliability of the Alpha for the adult group[8] is .977.

The reliability of the Stanford Revision for these children has not been determined at each chronological year. Probably it does not extend below .90 at any age level within our range.

With the computation of sigma scores, we have a further means of checking the representativeness of our group. For if the average sigma scores do not differ sensibly from 0, and if the standard deviations are close to 1, we may assume that the distribution of our sample is practically equivalent to that of the larger group of which it is a part.

TABLE X.—MEAN AND MEDIAN SIGMA SCORES

	N	Mean	Median	S.D.
Fathers	105	−.18 ± .07	−.37 ± .09	1.04
Mothers	105	.00 ± .06	−.12 ± .08	.93
Midparents	105	−.10 ± .06	−.23 ± .07	.87
Sons	158	.06 ± .06	−.06 ± .08	1.11
Midsons[1]	88	.09 ± .07	.02 ± .09	1.03
Daughters	159	.26 ± .06	.16 ± .07	1.07
Middaughters[2]	89	.20 ± .07	.11 ± .09	.97
Midchildren	105	.14 ± .06	.11 ± .08	.94
Stanford Children	213	.11 ± .05	−.06 ± .06	1.09
Alpha Children[3]	136	.21 ± .09	.15 ± .08	1.05

[1] The midson group includes 39 cases with only 1 son, 34 cases of 2 sons, and 15 cases of 3 or more.

[2] The middaughter group includes 44 cases of 1 daughter, 28 cases of 2 daughters, 17 cases of 3 or more.

[3] In the correlations, if a child was tested on both the Alpha and the Stanford Revision, the Alpha was used above the age of 12-6 and the Stanford Revision record below that age.

The only measures differing significantly from the norm are the median of the fathers and the mean of the daughters. These differences are not quite reliable, when the P.E. of the zero point is taken into consideration.

INTERPRETATION OF THE DATA

The data presented in our tables may be considered in relation to a number of current hypotheses of mental inheritance.

[8] Reliability computed by correlating odd and even items, on 102 cases of adults taken at random, and correcting for length of the test (Brown-Spearman formula).

TABLE XI.—PARENT-CHILD COEFFICIENTS OF CORRELATION (r) IN MENTAL-TEST ABILITY
(All correlations rectilinear)

	Son	Daughter	Son and Daughter	Midson*	Mid-daughter*	Midchild*	Average
Father	.524 ± .039	.505 ± .040	.509 ± .028	.580 ± .048	.492 ± .054	.592 ± .043	.534
Mother	.544 ± .038	.557 ± .037	.548 ± .026	.608 ± .045	.591 ± .044	.653 ± .038	.584
Midparent	.586 ± .035	.586 ± .035	.589 ± .025	.648 ± .042	.610 ± .045	.693 ± .034	.619
Superior Parent	.562 ± .037	.574 ± .036	.566 ± .026	.624 ± .044	.578 ± .048	.662 ± .037	.594
Inferior Parent	.572 ± .036	.553 ± .037	.546 ± .026	.656 ± .040	.579 ± .048	.660 ± .037	.594
More Deviate Parent	.572 ± .036	.546 ± .037	.560 ± .026	.602 ± .046	.580 ± .047	.689 ± .035	.592
Less Deviate Parent	.500 ± .040	.540 ± .038	.515 ± .028	.550 ± .050	.510 ± .053	.610 ± .041	.538
Average	.551	.552	.548	.610	.563	.651	

*It is of course to be expected that the correlations of parental scores with those of midson, middaughter and midchild should be higher than those in which single offspring are paired with parents because the tendency to variation within a sibship is partly ironed out by using group averages.

For coefficients of the order of magnitude reported here, the author is in some doubt as to the desirability of applying the formula for correction by attenuation. In any event, such a correction would increase the coefficients by not more than from 1 to 1.5 times the P.E. of the correlation.

1. The Dominance of Superior Intelligence

No confirmation is found for the doctrine that the superior grades of intelligence are dominant to the inferior grades. To the extent that correlations may be taken as evidence, the inferior parent exerts the same degree of influence as the superior parent, in determining the mental test score of the progeny. While not in agreement with a naive unit-inheritance theory of intelligence, our results are undoubtedly amenable to statement in terms of multiple Mendelian factors, additive complexes of genes, such as have been assumed to function in the inheritance of quantitative traits in plants and animals.[9]

2. The Influence of the Mother

From a number of considerations involving the pre-natal and post-natal environment, the mother's influence on the offspring would be expected to be slightly greater than the father's. The hereditist, however, basing resemblance on the germplasm contribution, would predict little or no difference. In our table, the mother's coefficient averages five points higher than the father's —again an unreliable, but a fairly consistent, difference. No conclusions can be drawn from it, particularly when one considers the possibility that in a small percentage of the children, paternity may extend outside of the family.

3. The Influence of the Same Sex

In rural more than in urban families, the son's activities are from an early age closely associated with those of the father, while the daughter remains at home to share her mother's duties in the household. From the standpoint of certain environmentalists, we would then expect the like-sex coefficients to be higher than those of unlike-sex. The four coefficients in each of these classes average .563 and .537 respectively, with the difference neither reliable nor consistent.

---

[9] Pearson, K. "On a mathematical theory of determinantal inheritance," *Biometrika*, Vol. 6, pp. 80-93.

Nilsson-Ehle, H. Einige Ergebnisse von Kreuzungen bei Hafer und Weizen. Bot Notiser, 1908.

Warren, D. C. "Inheritance of egg size in Drosophila Melanogaster." *Genetics*, Vol. 9.

4. The Issue of Unlike Parents

Galton observed that the stature of children depends in a prime degree upon the stature of the midparent, and that the progeny of unlike parents are in general no more diverse than the progeny of like parents. Our data substantiate this conclusion, for mental test ability. For when we correlate the individual pair differences of the parents with the average deviation of the (Stanford Revision) children, we obtain $r = -.145 \pm .075$. In 33 cases the parents differ from each other by as much as .75 sigma; in this extreme group, the correlation between the diversity of the parents and the diversity of the children is $-.127 \pm .12$. In these cases the mating of "weakness with strength," following Davenport's recommendation,[10] has produced no increase in the variability of the offspring; it has, however, resulted in intermediate progeny more often than in strong progeny, an outcome which may have an obvious application in eugenic theory.

5. "The Maximum Influence of Heredity"

Pearson has considered that the multiple correlation of A and an infinite number of brothers should be the same as the multiple correlation of A and two parents (or of A and all of the forebears at any given generation level.)[11] This may be expected to be true, however, only in cases of non-assortative mating. Comparing our mental test correlations with those reported by Pearson for physical traits, we find a remarkable agreement in some respects; the chief difference is in the greater degree of assortative mating in our study, and the consequent lowering of the multiple R with parents.

While the multiple R with two parents is less than that found or predicted by Pearson, it may be noted that our midparent-midchild correlation ($.693 \pm .034$) is practically identical with his result. Doubtless, the increase in reliability with mental test data, when children are combined into midchild, and parents into midparent, is one source of the increase in the coefficients obtained between these composites.

---

[10] Davenport, C. B. *Heredity and Eugenics*, p. 281.

[11] Eugenics Laboratory Lecture Series III, 1915. *Biometrika*, Vol. 17, p. 129.

TABLE XII

	Pearson's Data	Jones' Data	
	(Approx. Averages)	N	Correlation
r siblings.............	.50	828*	.490 ± .180
r parent-child.........	.50	317	.508 ± .028 (father)
		317	.548 ± .026 (Mother)
r husband-wife........	.00†	105	.598 ± .042
Multiple R of sibs (theoretical)........	.707		.700
Multiple R of children and 2 parents.......	.707		.590

*The sibling correlation was determined by the method of the maximum number of pairs (Elderton), plotting each individual on both the x and the y axes. Furfey, P.H., "A formula for correlating interchangeable variables." *Jour. Ed. Psych.*, vol. 18, Feb. 1927. There were no pairs of twins in the group.

†.15 in the middle classes.

Pearson's thesis has been: "We are forced, I think literally forced, to the conclusion that the physical and psychical characters in man are inherited within broad lines in the same manner and with the same intensity."[12] The present material is at every point consonant with such an interpretation, although it can scarcely be regarded as furnishing conclusive proof.

It will occur to the reader that much analysis is possible with various groupings of this material, and with the addition of our data on first cousins and avuncular pairs. We considered it advisable, however, to deal with this first representative sample of 105 families as a unit, and to devote further inquiries not to this group alone but to the larger sample of which this is a part.[13]

[12] *Jour. Anthrop. Institute*, Vol. 33, p. 204.

[13] The remaining families are those containing only one child of test age, or only one parent who could be reached during the period of the survey.

PREFATORY NOTE TO CHAPTER VI

These data upon a pair of Siamese twins are unique. If it were only possible to obtain measurements upon a large number of conjoined twins to compare with (1) measurements of ordinary identical twins, (2) measurements of identical twins reared apart, and (3) measurements of ordinary fraternal twins, an important part of the nature-nurture problem might be brought nearer solution. This concerns the relative contributions to intellectual differences of *random* environmental influences, both pre-natal and post-natal[1] (which of course would act to reduce correlations between twins), and *consistent* environmental influences (which would act to increase correlations).

It is to be hoped that more data of the kind presented in this study may eventually be gathered. Even with the results from this single pair, effects which are probably those of *random* environment can be discerned, since the differences between the two members of the pair upon two standard group intelligence tests are as large as or even larger than those usually found between ordinary identical twins, despite the utmost opportunity that environment has had to act with equal effect upon each twin.

A question fully as fascinating and profitable as that which concerns the intelligence of these Siamese twins is that of their personality likenesses or dissimilarities. Data are reported upon personality tests administered to the pair, but the tests used are not reliable enough to make their interpretation more than problematic.

---

[1] The pre-natal and post-natal random influences can themselves be divided into those having cumulative effects from the instant of conception throughout life, and those working upon the organism over a brief period of time. Those working over a brief period of time account for what is usually called the 'probably error of a score' and their effects are observable in the fluctuations of scores of single individuals.

CHAPTER VI

A STUDY OF A PAIR OF SIAMESE TWINS

HELEN LOIS KOCH
University of Texas

Many possibilities of dispute lie in the allegation that, since the heredity and the environment of the members of a pair of Siamese twins are practically identical, a study of the behavior likenesses and differences of these unique individuals should aid in the solution of the problem of trait determination. The antithesis of heredity vs. environment as factors in trait determination some authors choose to abandon, preferring to face the issue in terms of the more objective considerations of the ease and of the specific techniques with which trait moulding can be effected or controlled. Even if, however, the value of the antithesis be granted, the application of the term "identical" to either the hereditary or environmental forces which play upon Siamese twins may still legitimately be questioned. Their environments, broadly considered, are, to be sure, highly similar, but it must not be overlooked that each twin is a part of the environs of the other and that a constant interaction between the two members of the pair is inevitable. We are not sufficiently informed, moreover, in regard to the *modus operandi* of most of the so-called environmental forces to weigh with any justice the multifarious significant details in which the environments of conjoined twins have from earliest embryonic life been different as well as alike.

Whether Siamese twins have an identical heredity or not is an issue which stands or falls with the validity of the chromosome theory of inheritance and of the doctrine of the monozygotic origin of a certain portion of the twin population. Neither of these theories is it our intent to review here. A somewhat detailed presentation of arguments as well as a much more extensive series of measurements is given elsewhere.[1] We shall indicate merely those traits in our subjects which, if the arguments for the uniovular

[1] Koch, H. L. "Some measurements of a pair of Siamese twins." *Jour. Comp. Psych.*, 7:1927, 313-333.

origin of some twins have cogency, would tend to establish these twins as of that type. Our subjects are of the same sex—girls. Being conjoined in symmetrical parts as well as in ventral structures, they probably represent an incomplete organic separation rather than a fusion of disparate individuals. Their joining, let it be said, involves primarily the lower division of the intestinal tract. In respect to hair and eye color, hair type, tooth and eye pattern, as well as general body contours the two sisters are highly similar. No data, unfortunately, were available concerning the

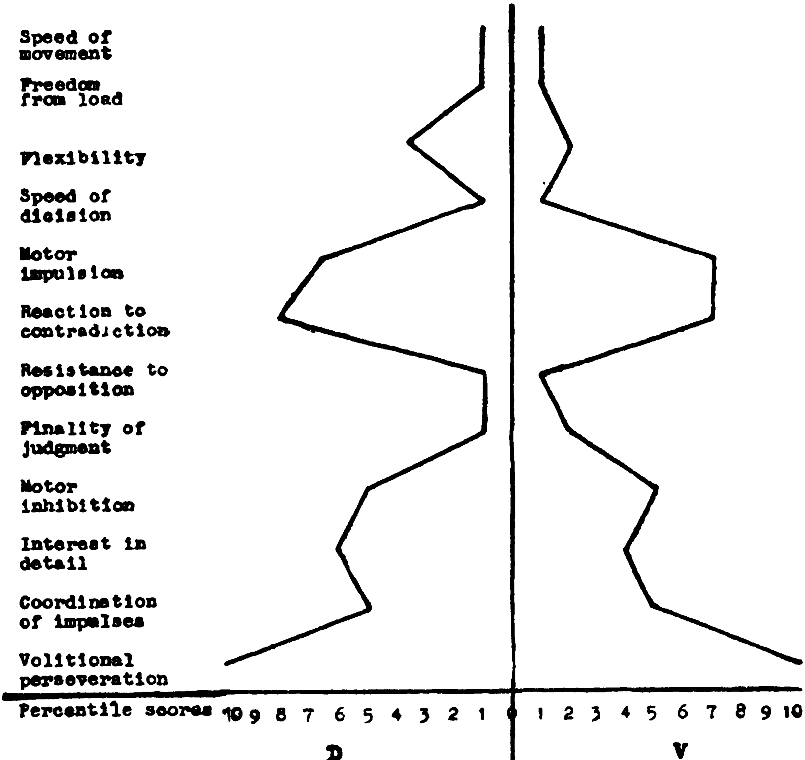


Figure I

singleness or duplicity of the embryonic membranes nor concerning the presence or absence of any degree of *situs inversus* in the viscera.

In spite of many marked similarities, differences are readily apparent in the girls. D, the right-hand member, for instance, is

1½ inches taller than is V, a fact which might be explained in terms of the more favorable placement of the former. The finger prints also show significant dissimilarities. Analyzed according to Bonnevie's system,[2] the prints of the two individuals present differences slightly greater, on the one hand, than the average of those exhibited by a group of alleged identical twins, and, on the other hand, of about half the size of those noted in a group of alleged fraternal twins. It is worthy of comment in this connection, too, that a comparison of the right with the left hand of each twin reveals about as marked dissimilarities as does a comparison of the hands of one twin with those of the other.

All of our observations considered, then, the chances favor the point of view that our subjects are identical twins.

The subsequent cursory account of their personal history may further aid the reader in the interpretation of our measurements. The girls were 14 years and 10 months old at the time of examination, Christmas, 1922. They were born in Brighton, England, of English parents. Orphans from birth, they have been cared for by an uncle and aunt. During most of their lives they have been associated with travelling shows or vaudeville circuits, as their livelihood was dependent in part upon exhibiting themselves. In their travels they have covered three continents. An informal sort of schooling has necessarily been theirs, with those aspects of their training stressed which fitted them most effectively for the task of interesting the public. For seven years, however, they received instruction in scholastic fundamentals at the hands of one tutor or another. The girls enjoy music and play with considerable skill the saxophone, clarinet, and piano. Their conjoinment has not interfered seriously with their locomotion, for they dance as well as run. They have been able to walk since they were 2½ years old. While their speed of reading is somewhat retarded, they both claim to enjoy reading, especially in the field of detective stories. At the time of examination they were much absorbed in their dolls and pets. Emotionally, they seem well poised. Their health has been almost perfect. D, who is the right-hand member of the pair, both cut her teeth and mensturated first, anticipating her sister in the

[2] Bonnevie, K., "Studies on papillary patterns of human fingers." *Jour. of Hered.*, 15:1924, 1-111.

former case by three weeks and in the latter by exactly one month. Although at present D is the more talkative of the two, it was V who as an infant uttered the first word, performing this feat, however, only a few days in advance of her partner.

In order to prevent the girls from aiding one another in any way during our examinations we subjected them to written tests whenever that was possible. A cardboard partition was also placed between them. The physical measurements were made in accordance with the techniques outlined by Baldwin in his monograph on growth.[3]

In our results the following relationships are observable:

1. D's height, head length, and head width are greater than are V's (see Table I). The differences in the head measurements of the two are significantly less than the P.E.dist. of such measures in a fourteen-year-old female population described by Porter.[4] The difference in height, however, falls not far short of the P.E.dist. of the heights of fourteen-year girls (Porter).

2. The left hand of V is, according to dynamometer records, stronger than is her right hand and also slightly stronger than is D's right hand. D's left hand, on the contrary, is not so strong as is her right, being equal in power to V's right (see Table I). It is a question for speculation whether these relationships are a product of the adaptation necessitated by the conjoinment of the girls, are a chance difference, or are on a par with the mirror-imaging phenomenon alleged to be common in identical twins.

3. Consistent superiority is exhibited by D in both the Terman and Army Beta tests (see Table II). The differences between the scores of the two on the Beta is 77 percent of the P.E.dist. of the 307 scores obtained by Darsie[5] from a mixed male and female American population of twelve-year-olds. The median score of Darsie's subjects was 68.3.

Since suitable age norms for the Terman test were not available, no appropriate comparisons can be made. The difference in

---

[3] Baldwin, B. T. *The Physical Growth of Children from Birth to Maturity.* Univ. of Iowa Studies in Child Welfare, I: 1921, 411 pp.

[4] Porter, W. T. "The growth of St. Louis children." *Trans. of the Acad. of Sci. of St. Louis*, 6:1895, 263-369.

[5] Darsie, M. The mental capacity of American-born Japanese children. *Comp. Psych. Mon.*, 3: 1926. 89 pp.

TABLE I.—PHYSICAL MEASUREMENTS OF SIAMESE TWINS

Trait Measured	D	V
Height (cm.)	110.7	106.7
Head length (cm.)	16.9	16.7
Head width (cm.)	13.5	13.4
Cephalic index	0.7988	0.8024
Grip (kg.) right hand	20.0	19.5
Grip (kg.) left hand	19.5	21.0

TABLE II.—SCORES OF SIAMESE TWINS IN TWO INTELLIGENCE TESTS

Test	Score of D	Score of V
Terman Group Test of Mental Ability, Form A		
Information	2	3
Best answer	10	8
Word meaning	4	3
Logical selection	8	7
Arithmetic	2	2
Sentence meaning	4	0
Analogies	5	2
Mixed sentences	3	1
Classification	13	10
Number series	8	6
Total score	59	42
Group Examination Beta (Army)		
Maze	3	2.50
Cube analysis	7	8
X-O series completion	10	7
Digit-symbol substitution	10.66	10.66
Number checking	16	13
Pictorial completion	7	7
Geometrical construction	5	5
Total score	58.66	53.17

TABLE III.—PERFORMANCE OF SIAMESE TWINS IN KENT-ROSANOFF ASSOCIATION TEST

	D	V
No. of common specific responses	79	53
No. of common non-specific responses	8	0
No. of individual responses	13	46
No. of doubtful responses	0	1

the scores of the sisters, relative to the size of their total scores, however, is greater in the case of the Terman test than it is in the case of the Beta.

4. The Kent-Rosanoff Association Test called forth a very different sort of reaction from our subjects; V gave 46 individual or unusual responses out of 100; D, only 13 (see Table III). Only 8 responses are common to their lists, whereas a comparison of the response series of 32 adults (16 twin pairs) revealed, on the average, 17.8 common responses.

5. The Downey Will-Temperament profiles of the two sisters are strikingly similar. In 7 of the 12 tests the girls fall within the same percentile, and in only 2 of the tests do they differ by as much as 2 percentiles (see Figure I).

6. While differences between the two individuals do stand in relief in our results, observations made elsewhere emphasize the fact that the deviation of the scores of the pair from the appropriate age norms offered for comparison are almost invariably in the same direction. This agreement in an extensive series of measurements, coupled with frequent divergences of great size from comparable norms, throws into prominence the essential similarity of our twins.

PREFATORY NOTE TO CHAPTER VII

One of the obvious methods of studying the relative influence of nature and nurture is to classify a large number of twins into three groups: (1) twins known to be identical, (2) twins of the same sex, among whom must be all of the identical twins (because identical twins are always of the same sex) and a certain number of non-identical twins (since non-identical twins are as likely to be both boys or both girls as to be of opposite sex), and (3) twins known to be non-identical; then to test the intelligence of each pair of twins carefully in order to determine the amount of difference in the brightness of the pairs in these three groups. These differences may also be compared with the differences found between (4) a group of ordinary siblings, not twins.

This is the nature of the study reported by Miss Tallman, whose results confirm the general conclusions of other students of this problem. In order of likeness in intelligence the order stands: identical twins, like-sexed twins, non-identical twins, siblings. Of course, it remains possible to argue that such a result might still be thought of as due to similar degrees of closeness of environment, rather than as due to inherited constitutional dispositions, but such an argument certainly seems too far-fetched to be taken very seriously—it would have to be assumed, for instance, that the environment is distinctly more nearly the same for identical than for non-identical twins.

Just what the figures obtained by Miss Tallman and others mean in terms of any precise quantitative determination of the relative effect of nature and nurture remains a matter challenging statistical research in the future.

CHAPTER VII

A COMPARATIVE STUDY OF IDENTICAL AND NON-IDENTICAL TWINS WITH RESPECT TO INTELLIGENCE RESEMBLANCES

GLADYS G. TALLMAN, A. M.
Psychological Laboratory, Columbia University,
Psychological Laboratory, Neurological Institute,
New York City

The purpose of this study was two-fold: first, to compare the differences in intelligence quotient ratings of siblings, from one to four years apart in chronological age, with those of twins, and secondly, to study the differences in intellegence quotient ratings in twins appearing very much alike and those appearing very different.

The term 'sibling' was first used by Karl Pearson to denote brothers and sisters in the same family. From a biological viewpoint a sibling is a single birth unit. Before birth the embryo is enveloped in an amniotic sack and chorionic coat and has its umbilical cord attached to one placenta. Twins are known to be of two types, those developing from one egg, called uniovular, duplicate or identical, and those developing from the fertilization of two eggs, called biovular, fraternal or non-identical. Before birth a pair of identical twins would be enveloped each in his own amniotic sack but both would be in the same chorionic coat. Each would have an umbilical cord but both cords would be attached to a single placenta. A pair of non-identical twins in the embryonic stage would be enveloped as are single units.

The chief assertions with respect to twins on which embryologists, obstetricians, and biologists agree seem to be these: (1) Twins of opposite sex are always non-identical; (2) Non-identical twins of the same sex need look no more alike than ordinary siblings; and (3) Identical twins are always of the same sex, are strikingly similar in form and feature, and much alike in 'mental characteristics.' But what these 'mental characteristics' are these authors do not say.

Some Previous Studies

However, mental characteristics of twins have been the subject of study. Francis Galton in 1875 reported on eighty pairs of twins. He makes note of insanities in twins, similarities in association of ideas, illnesses occurring simultaneously in both twins although apart, etc. This work was based on the questionnaire method and had obvious limitations on that account.

In 1905 Thorndike published *Measurements of Twins.* His data were concerned with fifty pairs of twins from nine to fourteen years in age who were tested in eight physical and six mental traits. He concludes that "the resemblance of twins in mental traits is roughly twice that of ordinary siblings."

Arnold Gesell's work on the *Mental and Physical Correspondence in Twins,* published in 1922, is devoted to the comparisons of physical and mental traits of one pair of identical twins. These children were born by Caesarian section, so that their embryonic condition should have been known. They presented a remarkable degree of correspondence in physical and mental constitution.

In 1924 Merriman published *The Intellectual Resemblances of Twins.* He applied both group and individual tests to two hundred pairs of twin children. As one hundred and five pairs were given the Stanford Binet test, it has been possible to make use of the results of this group. Dr. Merriman's results show (1) that environment appears to make no significant difference in the amount of twin resemblance, (2) that twins suffer no intellectual handicap, and (3) that there are two distinct types of twins.

"Mental Traits and Heredity—The Extent to Which Mental Traits are Independent of Heredity, as Tested in the Case of Identical Twins Reared Apart" is the title of an article by H. J. Muller in a recent publication. Dr. Muller reports a detailed study of one pair of adult twins and shows that, although brought up apart, they tested very much alike on tests for intelligence, but differed decisively on tests "for emotions and social attitudes." In this article is also presented a mathematical method of determining the identity or non-identity of twins. This method was not usable in the present study, as it is necessary to have data on other children in the families having the twins in order to determine the type of twins. Such data were not collected at the time of examining the children to be reported upon and could not be obtained later, except in isolated instances, because of graduation from school, family removals, etc.

The Present Study

The present study was started four years ago. In order to locate twin children a circular letter was sent to all New York City's private schools and permission was obtained from the Department of Education of the City of New York to test twins in public schools. The coöperation of both groups of schools was excellent. The data on siblings were assembled in the same way; the same private and public schools coöperated by providing families for testing.

The test used in all cases was the Stanford Binet. In the case of the twins note was made of the coloring of hair, eyes, and skin, general appearance as to facial symmetry, height, and weight. The children were also asked if their teachers and friends knew them apart. One examiner made all the tests. Both twins were tested on the same day, one immediately following the other. All the twins were living in their own homes. In the case of the siblings, complete families were tested the same day. The ages of both twins and siblings ranged from three to twenty years. The average differences of the intelligence quotient ratings, together with the probable error of the average difference, were estimated. In the case of the siblings the differences of children adjacent in age were used.

In all, 72 families of 199 children and 158 pairs of twins were tested. The average difference in the case of the siblings was found to be 13.14 with a probable error of .709. The average difference of all the twins, that is those of this study and those tested by Merriman, is 7.07, with a probable error of .265. Here it would seem that the difference in the siblings is about twice as great as that in the case of the twins. However, these siblings were from one to four years apart in age; so it might be argued that changes in home environment over the four-year period might have influenced these ratings. Out of this group were taken 128 children who were less than two years apart in chronological age. Their average difference in intelligence quotient ratings was found to be 11.96, with a probable error of .963. This figure is much closer to the original sibling difference than it is to that of the twins.

In the complete twin group the like and unlike sexes were treated separately. There were 84 boy-girl pairs. The average difference here is 8.48, with a probable error of .517. The like sex group was composed of 178 pairs and had an average difference of 6.42, with a probable error of .299. One would expect this measure to be lower than that of the boy-girl group, because of the presence of identical twins in the former group.

Of the 158 pairs personally examined, 63 pairs were noted as looking so much alike that it was all but impossible to tell them apart, while 39 pairs were obviously different in appearance. The 63 pairs are here called the identical group. The average difference is 5.08, probable error .469. The non-identical group of 39 pairs had an average difference of 7.37, probable error .624. In the identical group there were 29 pairs of girls and 34 pairs of boys. The average difference in the girl group is 4.22, probable error .434, and that of the boy group 5.82, probable error .766. Within the non-identical group there were 22 pairs of girls and 17 pairs of boys. The average difference of the girls is 7.14, probable error .710; that of the boys is 7.56, probable error 1.06.

The results of the work are, then, that by intelligence-quotient ratings twins are about twice as much alike as are siblings; that boy-girl twins, known to be non-identical, resemble the siblings, results more than does any other twin combination, whereas, of the like-sex twins, those which look much alike test more nearly alike than those which look distinctly different.

PREFATORY NOTE TO CHAPTER VIII

The work represented by the following chapter is notable as the first study of sibling resemblance in mental traits, based upon reliable measures of *laboratory animals.* No ambiguous factors of home training, schooling, or social *milieu*—in short, none of the influences proposed by environmentalists as sufficient to account for mental resemblances found among human families—could possibly confuse the issue when the subjects of the experiment are groups of albino rat littermates reared under uniform conditions. If the coefficients of sibling resemblance reported in this study can be interpreted as establishing the existence of hereditary mental traits among animals, an enormous weight of evidence immediately falls in the direction of heredity in the human domain.

The corrected correlation found for sibling pairs matched from 224 individuals, comprising 40 different litters, upon errors on runs 6 to 15 of a reliable maze was .31 ± .04. It may occur to the reader at once that while .31 is a significant correlation in the light of its probable error, it is not as high as the correlations for human siblings which are ordinarily found upon mental tests. (Most sibling correlations are close to .50.) If the rat populations were comparable to unselected human populations with respect to heterogeneity, freedom from inbreeding, assortative mating, etc., then the tentative inference might be drawn that sibling resemblance was greater for human beings than for rats because environment enhanced the resemblance of human siblings. But the authors point out that the degree to which the factors enumerated above are present in the rat population in question is not accurately known. What *is* known regarding the presence of these factors leads to the conclusion that the correlation of .31 would have been higher if subjects could have been drawn from an unselected population of rats.

CHAPTER VIII

FAMILY RESEMBLANCE IN MAZE-LEARNING ABILITY IN WHITE RATS[1]

MILDRED BURLINGAME AND CALVIN P. STONE
Department of Psychology, Stanford University,
Palo Alto, California

The inheritance of learning ability in animals is a subject of great interest and importance to students of psychology as well as to geneticists. Nevertheless, contributions to this subject[2] have been relatively few in number and of a preliminary character, owing, no doubt, to the relatively great span of an experimenter's time and to the outlay of research capital required for successful prosecution of programs in this field.

Our paper deals with sibling resemblance in maze-learning ability of rats tested during the past year in the Stanford Psychological Laboratory. Although the data were gathered primarily for the purpose of contrasting learning ability in rats of different ages, they will serve to indicate the trend of sibling resemblance in maze-learning ability, despite the fact that they are unsuitable for an ideal study of this subject because of certain shortcomings to which attention will be directed.

THE ANIMALS

The subjects are descendants of the Slonaker colony, Stanford University, which has been maintained from a small original stock of albinos during the past 25 years. Undoubtedly, there has been a great amount of in-breeding, although accurate information on this point is lacking. The parental stock was mated without knowledge of their learning abilities. Furthermore, the details of paternity for some groups cannot be established, because at certain times more than one male was kept in the breeding cages. This condi-

[1] This report is an abridgment of a Master's thesis by Miss Burlingame, now on deposit in the Stanford Library. The cost of training the animals was defrayed by a Carnegie grant.

[2] See the article by Mildred Burlingame: "Literature on the heredity of behavior traits in animals." *Psychol. Bull.* 24: 1927, 62-68.

tion has probably produced more homogeneity in the different litters than would have been the case if a different male had sired each litter.

For convenience of the experimenter, the animals were run on the maze in groups[3] of 16 to 35 animals. The individuals of any group are of the same age, although they represent a total of from 7 to 9 litters. Table I lists the groups and gives their respective ages, the numbers of litters represented, and the number of cases in each group distributed as to sex. Groups comprised of littermates are indicated by naming the groups of which they are littermate controls. The ages at the beginning of the experiment are stated.

TABLE I.—ANIMAL GROUPS WITH THEIR RESPECTIVE AGES, NUMBERS OF LITTERS, AND NUMBERS OF INDIVIDUALS, DISTRIBUTED AS TO SEX

Age Group	Number of Litters	Number of Individuals		
		Male	Female	Total
75 days	7	12	12	24
100 days	8	14	6	20
5 months (Con. of 100-day gr.)	7	7	9	16
6 months	9	18	11	29
8 months (Con. of 75-day gr.)	9	11	14	25
9 months	7	13	12	25
12 months (Con. of 6-mo. gr.)	9	11	14	25
12 months	9	16	19	35
18 months (Con. of 12-mo. gr.)	9	12	13	25
Total No. Dif. Litters	40	114	110	224

THE APPARATUS

Fig. 1 gives the floor plan of the maze[4] from which our records were derived. As will be seen by examining this drawing, each

[3] In order to study learning ability in rats of different ages each litter was split into two groups. One of the groups was run on the maze at one age and the other at a later age. This technique tended to control the genetic factor and certain environmental influences pertaining to early nutrition. Within the age limits herein considered, however, and with the technique of maze operation followed, learning ability seems to be approximately equal at the different ages. Hence, there is no apparent objection to throwing together the data for the separate groups in the manner we shall presently describe.

[4] A more detailed description and more elaborate drawing of this Multiple —T maze will be found in a forthcoming paper by Calvin P. Stone and Dorothy Nyswander.

true pathway leads to a crossroad running at right angles to it. At each crossroad the animal must make a choice of direction. One turn will enable him to continue in the direction of the goal; the other involves a blind alley from which he must return. Short pathways at the ends of blind alleys make them appear essentially like the true pathways and thus enhance the difficulty of making

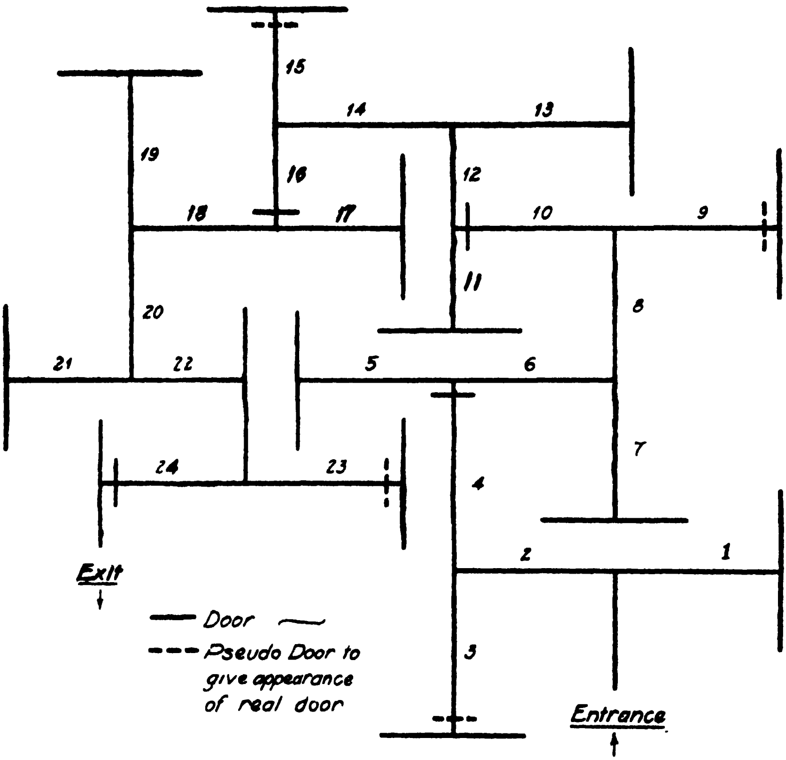


FIG. I.—FLOOR PLAN OF MULTIPLE-T MAZE

correct choices. As the drawing indicates, doors are put at the ends of pathways 4, 10, 16, and 24. These prevent retracing, once the animal has passed them in his run. Pseudo-doors were put in the opposite ends of each of the aforementioned pathways to make both ends appear alike, lest the animals' choices at these points be made on the basis of presence or absence of doors at one end of the pathways. Only in the early trials, when the animals are forming the habit of forward running in the direction of the goal, is it necessary to make use of the doors to prevent retracing.

Salient Points of Technique

Before starting the animals on the maze they were given five trials, one trial per day, on a simple platform escape box. The details of this apparatus need not concern us here, since acts performed in it are not specific for maze treading and, although very important forerunners[5] of the latter, only in a general way contribute to the maze situation. They permit the animal to become adjusted to routine handling in the experimental situation, the novel requirement of performing a specific series of acts to obtain food, a restricted food ration suitable for stable hunger motivation, and a 24-hour feeding schedule associated with maze running. Thus, they increase the reliability of the early trials.

As a general rule, the animals were deprived of food for a period of 24 to 36 hours before the initial trial on the platform box. For the next 10 to 12 days thereafter they were permitted to eat 20 minutes at the end of the learning trial. Adult animals usually lose weight at the rate of 2 to 5 grams per day during this period, but relatively young, growing animals maintain their weight as a result of continual growth. After the 10th or 12th day, the body weights of all animals were held relatively constant, except those of the 75-day group, which were allowed to gain at the rate of approximately 0.5 to 1.0 gram per day. Under this condition of weight control, hunger motivation appeared to be relatively constant from day to day and sufficiently strong to insure close attention to the business of maze running. Each day the learner is weighed and the weight used to guide the experimenter if it is necessary to increase or decrease the ration.

Measures of Learning

Although records of time as well as errors are recorded for each animal, we are not yet prepared to evaluate time scores as measures of sibling resemblance. The errors herein considered consist exclusively of entrances into blind alleys while the rat is moving for-

[5] See the article by C. J. Warden: "The preliminary period of feeding in the problem box." *Jour. of Compar. Psychol.*, 5:1925, 365-372.

ward through the maze. Retracing on the true path or into blinds is relatively inconstant, even in the beginning, and unimportant after the first four or five trials on our maze; hence, it does not afford a suitable basis for comparing learning ability within families. The maze test consists of one trial per day for 30 consecutive days.

Fig. II graphically illustrates the mean error scores for each group of animals from trials 6 to 15 with which we are chiefly concerned in this paper. It will be noticed that the curves for groups comprised of littermates show a slight tendency to coincide in their general levels from trials 6 to 15. Under the conditions of this experiment and with the ages herein represented there does not seem to be a consistent age difference in learning ability. For that reason we have not hesitated to pool the data for groups according to the method mentioned further on.

THE RELIABILITY OF ERROR SCORES

The reliability of error scores[6] for the Multiple-T maze has been calculated by four different methods, each of which yields a fairly high coefficient of reliability. The methods involve:

1. Correlating the sum of the errors on the odd trials with the sum of the errors on the even trials.
2. Correlating the sums of the errors on the odd-numbered blinds with the sum of the errors on the even-numbered blinds. (Sum of errors on blinds 1, 5, 9, 13, 17, 21, with the errors on blinds 3, 7, 11, 15, 19, 23; Fig. I.)
3. Correlating the sum of all errors for the first half of the maze with the sum of all errors for the second half of the maze. (Sum of errors on blinds 1 – 11 vs. sum of errors on blinds 13 – 23; Fig. I.) Roughly stated, we assume that the animal has run two adjacent mazes without interruption of progress from one to the other. The two mazes in this case are identical as to mode of construction and number of blinds.

[6] For a more extensive discussion of the reliability of maze records see the forthcoming paper of Professor E. C. Tolman and Dorothy Nyswander, *Jour. of Comparative Psychology,* and that of Calvin P. Stone and Dorothy Nyswander, *Ped. Seminary and Jour. of Genetic Psychol.,* now in press.

4. Correlating the sum of errors for any segment of the trial series with the sum of the errors for any succeeding trial series. Example: Sum of the errors for trials 1-10 vs. 11-20; or trials 6-10 vs. 11-15. This coefficient throws some light on the degree to which individuals of a group hold their relative positions with respect to rate of error elimination. When this coefficient is relatively high, the individual curves hold their relative trends to a marked degree.

Table II gives the coefficients of reliability for various trial series for the animal groups herein considered. It will be noted that our reliability coefficients for all but the first portions of the

TABLE II.—COEFFICIENTS OF RELIABILITY FOR ERROR SCORES BASED ON VARIOUS PARTS OF THE TRIAL SERIES (Trials 1–30)

Methods of Determining Reliability	Portion of Trial Series						No. Cases
	1–10	11–20	21–30	1–20	10–30	1–30	
Odd vs. Even Trials	0.781 ± .018	0.933 ± .007	0.937 ± .005	0.893 ± .011	0.960 ± .004	0.947 ± .005	205
Odd vs. Even Blinds	0.859 ± .016	0.845 ± .018	0.839 ± .018	0.909 ± .012	0.878 ± .014	0.923 ± .009	120
1st Half vs. 2nd Half Maze	0.805 ± .020	0.831 ± .018	0.852 ± .016	0.865 ± .014	0.887 ± .012	0.898 ± .010	145
Segment vs. Segment of Trial Series	1–10 vs. 11–20	1–15 vs. 16–30	21–25 vs. 26–30	6–15 vs. 16–25	1–5 vs. 16–20	11–15 vs. 16–20	
	0.750 ± .025	0.652 ± .033	0.748 ± .025	0.777 ± .022	0.452 ± .044	0.807 ± .020	145

learning series are very high, and in this connection it is important to recall that the reliability of maze data for animals heretofore reported in the literature has without exception been very low. This no doubt explains, in part, the failure of others to find sibling resemblance in animal learning scores.

SIBLING RELATIONSHIP

For our study, sib relationship was calculated by correlating the error score of each animal with that of each of its sibs in the group with which it was run. The original division of animals to form age groups (for the purpose of an experiment on learning ability of animals of various ages) consisted of splitting litters so that each litter would have representatives in a group to be run at an early age and in a group to be run at a later age. Since the means and variability of error scores for these littermate groups are not respectively equal, we have considered it inadvisable to cross over from one group to the other to secure all possible pairings of littermates irrespective of ages at which the animals were run. Hence, our method of pairing animals for correlating error scores of sib-

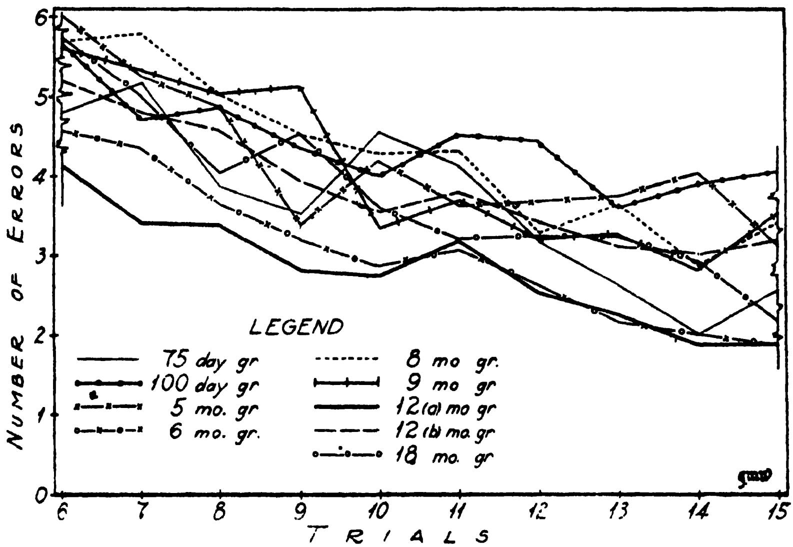


FIG. II.—MEAN ERROR CURVES FOR THE ANIMAL GROUPS LISTED IN TABLE II, TRIALS 6 TO 15

lings may be illustrated as follows: Suppose the animals of any litter within a group to be labeled A, B, C, and D. Then, all the possible pairings of error scores are obtained by pairing the score of:

A with B	B with C	C with D
A with C	B with D	
A with D		

That is the method we have used.

By way of preliminary prospecting for the best part of the trial series for the study of sib resemblance, several different segments of the trial series between 1 and 30 trials have been studied.

In our opinion, that segment which embraces the most rapid elimination of errors is the most valid index of learning ability. This falls between trials 1 and 15 for this experiment. It is usually advisable to drop the first few trials from consideration since they reduce the reliability of the error data. This leaves as our most valid and reliable measure some segment, such as that 6 to 15 which we have used in this study.

One would hardly expect to find statistically significant correlations in every instance between the error scores of littermates from only 9 or 10 litters unless these litters represented widely divergent levels of learning ability. As a rule, the family sampling would be inadequate. Our expectations in this respect are borne out by the coefficients obtained. About half of the small groups of 25 to 30 animals yield coefficients for sib resemblance that are statistically significant. As a rule, however, when a significant coefficient is obtained for one segment of the trial series, it appears again when other segments are considered. If an insignificant coefficient is obtained, the reverse of the foregoing generalization usually holds. Typical of the groups yielding significant coefficients for sib re-

TABLE III.—COEFFICIENTS OF RELIABILITY AND SIB RESEMBLANCE BASED ON ERROR SCORES FOR VARIOUS TRIAL SERIES FOR THE SIX-MONTH GROUP

(29 Cases)

Trial Series	Reliability (1st half vs. 2nd half of maze)	Sib Relationship*	Trial Series	Reliability (1st half vs. 2nd half of maze)	Sib Relationship*
1–30	0.87 ± .03	0.39 ± .10	6–25	0.91 ± .02	0.46 ± .10
1–25	0.91 ± .02	0.55 ± .09	6–15	0.98 ± .004	0.49 ± .09
1–20	0.87 ± .03	0.31 ± .10	11–20	0.88 ± .02	0.49 ± .09

* The P.E.'s of correlation coefficients used in this study have been computed by the standard formula, $P.E._r = .6745 \frac{(1 - r^2)}{\sqrt{N}}$. When maze scores of littermates have been correlated, the N used in this formula has been the total number of pairs, taking all combinations of pairs from each set. Some of the P.E.'s, therefore, have actual values which are slightly higher than those reported, since the use of the same individuals twice or more in a single correlation array results in a correlation between errors of sampling. There are no available methods by which these P.E.'s can readily be corrected to their true value, when, as is the case in this study, the number of offspring in the different litters varies.

semblance is the six-month group. (Table 1). Table III gives the coefficients for sibling, resemblance in error scores for various trial groupings. Each of these is statistically significant in terms of its P.E. For this group the reliability as calculated by each of the four methods is relatively high (see Table II, six-month group). However, the same may be said for other small groups in which the coefficients for sib resemblance are insignificant.

Since small groups of 21 to 30 animals do not give an adequate sampling of rat families as to learning ability, we have considered it advisable to pool the data for separate groups in order to enlarge the scope of genetic sources. The pooled data enable us to base our correlations on scores from 224 cases and from 40 different litters. This even yet, no doubt, falls considerably short of the requirement for an ideal sampling of rat learning ability. *For this grouping individuals were compared only with littermates within the same age group.* This method has been followed in order to permit us to check against the introduction of a spurious factor that might account for such correlation as we have obtained. Concerning this latter point, we shall speak in a later paragraph dealing with the coefficient for non-sibling resemblance.

The reliability of the error scores on trials 6 to 15 as measured by correlating errors on the first half with the errors on the second half of the maze is .87 ± .012. The coefficient expressing the sibling relationship, corrected for attenuation, is 0.31 ± .04.[7]

At the present time we are unable to determine how nearly this coefficient approaches the true sibling resemblance in maze-learning ability, since we are still uncertain as to the absolute validity of the scores under consideration. Our maze technique is adequate to insure quite satisfactory reliability coefficients. Furthermore, our daily records and observations indicate that we have eliminated or reduced to a practical minimum such factors as inactivity in the maze, lethargic running, sporadic tendencies to explore rather than run the maze, etc., which might alter the validity of the scores without necessarily reducing their reliability, provided their influences in successive trials and various portions of the maze on given trials are distributed at random, as is usually the case. Nevertheless, we still lack an adequate standard by which to determine the

[7] See note following Table III.

discrepancy between perfectly valid learning scores and those now at hand. Probably the data from experiments now in progress in which certain animals are run successively on mazes of different patterns will throw some light on this relatively obscure point.

Since the curves representing the mean number of errors made by various groups of animals herein considered do not exactly coincide, we have raised the question whether this condition was due to systematic rather than chance factors, and also whether special conditions of the experiment might have caused homogeneity within groups but divergence between groups and thus have given rise to a spurious coefficient for sibling resemblance. The answer to the first question has not yet been found. To test the second possibility, we have correlated error scores for non-littermates. For the latter correlations pseudo-litters were made up from the non-littermates of each group, with care to insure that the number and size of litters corresponded to the real litters of the group. Thereafter, pairings of the non-littermates were made according to the plan previously described for littermates.

The coefficient of correlation obtained from these data was $0.036 \pm .04$. If a significant coefficient had been obtained, we should have been uncertain as to whether it was due wholly or in part to a spurious factor causing homogeneity in groups but divergence between groups; its absence is taken to indicate that this factor is not responsible for the obtained coefficient for sibling resemblance. Hence, for the present, we have no grounds for regarding the coefficient $0.31 \pm .04$ as other than an indication of actual sib resemblance in the error scores.

In this connection it is important to note that the resemblance expressed by the coefficient of 0.31 is probably less than would be obtained if all litters had been sired by different males. As we have already pointed out, some of our litters ranked as 'independent' may really have been half-siblings, thus producing greater homogeneity in our group as a whole than would have been the case had each litter been sired by a different male.

As an additional point indicating that the coefficient found is not exaggerated, but is, on the contrary, actually reduced, one may pertinently recall that rat litters are sufficiently large to give considerable chance for the appearance of genetic extremes. In as much as there has not been any selective breeding for learning ability, the range in a litter may conceivably be approximately as large as that of the group as a whole. This being the case, one might reasonably expect that the coefficients for sibling resemblance would rise and stabilize in amount for family populations no larger than that of our composite group with even a small amount of selective breeding of parental stock. The latter, if properly done, will increase the range of learning scores and thus make less prevalent the situation in which unreliable individual scores totally conceal real but slight resemblance in learning ability.

Our program for improved approach to studies on the inheritance of learning includes: (1) selective mating of parental stock through several generations; (2) increasing the reliability of individual scores by giving more attention to the motivation of individual, recalcitrant animals; (3) increasing the number of families under consideration; and (4) obtaining several litters from the same selected parental stock to secure more adequate representation of genetic classes.

PREFATORY NOTE TO CHAPTER IX

The investigation described in the following chapter should be considered in relation to that set forth in Chapter X. Both deal with the influence of nature and nurture upon the development of foster children. Because of the opportunity which this method of approach offers to evaluate the nature and nurture factors separately, the evidence offered in these two chapters is more nearly crucial than that coming from any of the other Yearbook investigations.

The two foster children studies involve such a complex chain of data and treatment that no attempt will be made to indicate their results here. Each contains a summary at the end which condenses some of the more important conclusions. The reader is also referred to Chapter XI which compares many of the results of the two investigations, item for item, and which attempts to account for several seeming discrepancies between the contributions. It is sufficient to state here that the two following chapters represent a specific type of approach to the nature-nurture problem which has been needed and recognized for many years.

GUIDE TO CHAPTER IX

CHAPTER IX

THE INFLUENCE OF ENVIRONMENT ON THE INTELLIGENCE, SCHOOL ACHIEVEMENT, AND CONDUCT OF FOSTER CHILDREN[1]

By Frank N. Freeman, Karl J. Holzinger, and Blythe Clayton Mitchell
Assisted by Helen R. Bobo and Clara H. Lorenzen
University of Chicago, Chicago, Illinois

I. Problems and Methods of the Investigation

1. General Problem

The chief problem of this investigation was to determine whether the intelligence of the child is affected by the character of his environment. This is a question upon which there is a wide variety of opinion, both among trained psychologists and among non-technical observers. There are some who contend that the degree of intelligence possessed by the individual is fixed by inheritance and determined at birth. There are others who maintain that one's intelligence is a product of both factors—one's native capacity and the sum of home influence, educational training, and all other environmental conditions.

This divergence in view may be based in part upon a difference in the definition of the term 'intelligence.' One investigator may define the word in terms of the measurable reactions of the individual to the situations which confront him. Intelligence under this conception is simply the quality of a person's behavior in dealing with intellectual problems. It is measured by the score which the individual makes upon adequate tests of intelligence. Intelligence may be thought of, on the other hand, as something behind

[1] This study was made with the aid of a grant from the Commonwealth Fund. Indispensable assistance in prosecuting the study was given by the Illinois Children's Home and Aid Society. This Society also furnished a large amount of collateral information. The authors make grateful acknowledgment to the directors of the Commonwealth Fund, to Superintendent C. V. Williams and the staff of the Illinois Children's Home and Aid Society, and to children and parents who generously coöperated in the investigation.

one's performances—some characteristic or quality of the brain or the nervous system which makes the individual capable of acting as he does and of solving intellectual problems. One who thinks of intelligence in this way might maintain that it is unaffected by environment or training. Even though the individual's performance exhibits such effect, he would interpret this as meaning only that performance is the joint product of intelligence and training.

It was assumed in this study that intelligence is defined in terms of the individual's performance, namely, as his intelligence-test score. Our intelligence tests may not be perfect measures of intelligence, but they are the best we have. If they are shown to be affected by environment, we shall have to conclude, so far as our present scientific evidence goes, that intelligence is affected by environment. In case tests are later devised which can be shown to measure the inner capacity for intellectual performance and which are more independent of environment than our present tests, their evidence will have to be accepted, but we are not warranted, from the scientific point of view, in assuming an intelligence, or native capacity, which cannot be measured.

The measures of intelligence used in the present study were the Stanford Revision of the Binet-Simon Scale for Measuring Intelligence and the International Group Mental Test, a scale which was particularly designed to be independent of environmental influences. The problem may, then, be restated as follows: How far may a child's performance on the Stanford Revision of the Binet Scale and on the International Test be affected by the character of his environment?

A subordinate problem was to study the effect of the environment of foster children upon their achievement in school and upon their conduct. In order to secure evidence on this problem, as full information as possible was gathered concerning the education and conduct of the natural parents of the children and of the children themselves. Information was also gathered concerning the education of the foster parents. From a comparison of the educational achievement of these three groups it is possible to draw inferences concerning the effect of the environment of the foster homes upon the attainments of the foster children. From a comparison of the conduct of the foster children with that of their natural parents some inference may be drawn concerning the effect of the foster homes upon the conduct of the children.

2. Plan of the Study

The chief difficulty which has been encountered in investigations of the effect of environment upon intelligence has been that of isolating environment from inherited capacity. When differences in environment and differences in heredity are associated, it is impossible to determine which is the cause and which is the effect. An example of this association of the two factors is found in the correlation between the amount of schooling which an individual possesses and his intelligence as measured by intelligence tests. It has been demonstrated beyond question that the individual with the greater amount of schooling has the higher intelligence. It is apparent that the mere fact of this correspondence can be interpreted to mean either that high intelligence is necessary to enable the individual to progress to the higher levels of training or, on the other hand, that his intelligence score is determined by the amount of schooling which he has had. The mere fact of correspondence gives no means of determining which factor causes the other. The same difficulty attaches to the interpretation of other correlations, as, for example, that between the intelligence of the inhabitants of the various states of the Union and the degree of development of their school systems or that between the intelligence of immigrants and the length of their residence in the United States.

The uncertainty of the interpretation in these cases is typical of the general difficulty in attacking this problem. The cause of the uncertainty is that both the grade of the individual's heredity and the character of his environment vary together, or at least it is not possible to demonstrate that one varies independently of the other. In order to overcome this difficulty, it is necessary to compare persons who are alike in heredity, but who have been subjected to different environment or, on the other hand, those who are of diverse heredity, but who have been under the influence of the same or similar environment. By the application of intelligence tests the extent to which one or the other of these factors affect the score may then be determined.

The most completely satisfactory method of bringing about such variations of one factor independently of the other is the experimental method. The experimental method might be used if a group of individuals with similar heredity could be divided into two groups and then subjected to widely different environmental influences. Such a procedure is obviously difficult to follow in dealing with human beings. It occurred to the writers, however, that the adoption of children might give conditions which resemble the main features of such an experiment. It may be assumed that foster children placed by a society for the care of dependent children will have a heredity which is somewhat below the average. The evidence in our study goes to show that this assumption is justified. There is also reason to believe that when such children are adopted they are introduced into an environment which is superior to that of their own family. This assumption also has evidence to support it. In the light of these assumptions, the original plan of the study was to make three comparisons.

The first was to be made between two children of a family, one of whom had been adopted into a foster home, while the other had remained in his own home. Sufficient data for this type of comparison could not be obtained, because, when one child of a family is adopted, the whole family is usually broken up.

The second comparison was to be made between children of a family (siblings) who had been separated and adopted into homes of different type. Such conditions would make possible, as in the first comparison, the separation of the factors of heredity and environment, since the former would be fairly similar, but the latter quite different, in the case of each pair of siblings. This, of course, assumes that children of the same parents have, on the average, the same inheritance. Besides making this assumption, it would be necessary to show that there was no tendency for the brighter child of a family to be placed in the better home. It was found possible to obtain some data of this type, a consideration of which forms an important part of the study.

It was hoped that the investigation might furnish another crucial comparison, namely, that between a foster child and an own child in the same family. If the foster child were adopted when quite young, the environment of the two would be similar but the heredity different. Under such conditions it would be possible to observe the effects of a difference in heredity. Data of this sort were found to be very limited because of the fact that foster parents do not usually have children of their own.

Because of the small number of cases involved in the foregoing comparisons, it seemed desirable to broaden the scope of the study by including a larger body of cases in which the factors of heredity and environment were less clearly differentiated. In spite of the fact that these cases could not be used in the crucial comparisons desired, nevertheless their greater numbers added reliability to certain of the more general and less direct comparisons which were made.

3. Method of Securing Cases

Information concerning the children who were studied in this investigation was secured through the kind coöperation of the Illinois Children's Home and Aid Society. This Society is the chief organization in the state for the care and placement of dependent Protestant children.

It was hoped that there would be a considerable number of instances in which a pair of siblings had been placed in different foster homes before either was three years of age. It was further desired that these children should have been in these foster homes at least five years and that they should be not over fourteen years of age at the time of the study. The search for such cases revealed the fact that a very large proportion of the children who had been placed by the Society were illegitimate and that, of the legitimate children placed, only one half were sibling pairs. A further limitation was brought about by the Society's general policy of placing brothers and sisters in the same home whenever possible. All these limitations greatly reduced the number of cases meeting the desired requirements of early entrance into different homes. In fact, only five instances were found in which a pair of siblings had been placed in different foster homes before either was three years of age and were now not over fourteen years of age.

By raising the age of entrance into homes to four years and by shortening the period of residence to four years it was found that approximately two hundred children would meet the requirements. The number of these actually included in the study was only 159, because often one member of a pair could not be located or had died, or because foster parents were unwilling to coöperate.

Examination of the files showed that a considerable number of children had had previous intelligence tests. It was decided to include this group in the study because of the direct comparisons possible between the results of these early tests and of those given in the study. In 74 of these cases the child had been in one foster home during the whole period between the tests.

In order to secure a greater number of cases for the general comparisons of foster children with various environmental factors, the study was broadened to include both illegitimate and legitimate children, regardless of whether they had a brother or sister in a foster home. By including such cases it was possible to secure many more children entering foster homes at the earlier ages and living under widely differentiated home environments.

A number of other children were tested by the members of the staff in order to assist the Society in the placement of these children. This accounts for the large number of tests given.

4. Tests and Ratings

The data of this study consisted in the results of tests given to children and parents, and in information obtained for the making of certain ratings of the children, the foster parents, and the homes. A description of the various tests used will be given next, after which the information blanks will be discussed, with an explanation of the methods of rating.

The tests given to the children were the Stanford Revision of the Binet Scale and the International Group Mental Test. The former is the well-known revision of the Binet Scale made by Terman and does not require description. The Binet Scale was chosen because it was thought that it would give the best individual measurement of the mental ability of children of various ages. During the investigation, however, it was discovered that the intelligence quotient appeared to drop with age and in the search for explanations a careful examination was made of the standardization of the Stanford Revision. It was discovered that, even for the group of

905 unselected children used in the final revision, the mean I. Q.'s for ages 5 to 14 showed a drop from 111 to 98, an average of over one point per year. Since this characteristic of the test was not discovered until the study was nearly complete, it was impossible to do more than make certain approximate corrections for age by group means.[3]

The International Test is still in the experimental stage and will, therefore, be described briefly. It is the outgrowth of intelligence testing in the army and was sponsored by the Migrations Committee of the National Research Council. The construction and standardization of the test was carried out by Dr. Stuart C. Dodd under the direction of Professor Carl C. Brigham of Princeton University.

The aim in the construction of the test was to devise a method by which intelligence could be measured independently of differences in culture or environment. In contrast with the Binet Scale, this test is entirely non-verbal in character, the material consisting only of pictures and diagrams drawn upon cardboard rotators. The rotators are arranged in rows on large sheets of cardboard. The principle which determines the setting of the rotators is discovered by the subject, without verbal directions, from a key at the top of each sheet. The scale is made up of a number of short tests which vary in difficulty. The large number of items in the separate tests and their range of difficulty is such that the ability of even superior adults may be adequately measured. It was hoped that the non-verbal character of this test and its wide range would make it especially valuable in the present study.

The International Test was originally designed to be presented in pantomime. Since all of the children tested in this investigation understood English, very brief verbal directions were substituted. The entire test is made up of two parts of about equal length and difficulty, one of which was considered sufficient for the present study. The fact that no norms were available for the test did not limit its value, in as much as only the point scores were used and these only in comparisons and correlations within a given group.

[3] The critical reader may wish to turn at once to VI, 8, "Relationship Between I.Q. and Chronological Age," in order to interpret the results of the study more satisfactorily.

The correlation between the International Test score and chronological age for a group of 297 cases was .56. For the same group the correlation between Stanford-Binet mental age and chronological age was .71. The higher correlation of the Stanford-Binet scores with chronological age may be interpreted in various ways. It may be that the Stanford-Binet Scale depends more upon experience than does the International Test, so that part of the correlation with chronological age is due to the accumulation of experience. Again it may be that, since the Binet Scale was standardized by the criterion of age differences, a closer correlation with age would exist in the case of this test.

The correlation between Stanford-Binet mental age and the score on the International Test for the 297 cases was .80, but this was reduced to .70 when chronological age was held constant. This is a very substantial correlation and indicates that the two tests are measuring much the same ability. It is, however, lower than that usually found between two verbal mental tests. This lower correlation would seem to indicate that the International Test and the Stanford-Binet Test are measuring somewhat different aspects of mental ability.

It may be assumed that the score on a mental test is determined by the three factors of native capacity, maturity (in the case of a child), and training. The correlation between the test score and age is due to the combined effect on the score of maturity and of training. If one test is less affected by training than another, we should expect it to show a lower correlation with age. Conversely, if one test has a lower correlation with age than another, it may be due to the fact that it depends less on training. The score on such a test might depend in greater proportion on the other two factors, native capacity and maturity. The conclusion from this reasoning is that, since the International Test and the Stanford Revision are correlated somewhat less closely than is common with language scales, and since the International Test is correlated less closely with age than is the Stanford Revision, the International Test may depend somewhat less on training and be a closer measure of native capacity.

The tests given to the foster parents were the Otis Self-Administering Test of Mental Ability, Higher Examination, and a Vocabulary Test especially designed for the present study. Owing to the large amount of time required in securing the data from a home, it was thought advisable to select an intelligence test for the parents which was short and easily administered. The Otis test was chosen because it met these conditions and is generally recognized as a reliable test suitable for adults. The twenty-minute time limit was used.

The Vocabulary Test was constructed by selecting words from various fields of knowledge. It was thought that material of this sort would give a rough measure of the intellectual and cultural level of the parents. The scale used is not standardized, but was tried with several groups of students before it was adopted in its present form. The Thorndike Test of Word Knowledge was given to the same groups and gave a correlation of .84 with the test used. A copy of the vocabulary test is shown in the Appendix.

In addition to the scores on the tests mentioned, other facts were obtained by the use of a blank which was filled out by the field worker through interviews with the foster parents and teachers. A copy of the blanks used will be found in the Appendix.

The first section deals with the material environment, including the neighborhood, building and grounds, furnishings, care and upkeep, and evidences of culture. The three field workers recorded the necessary information and indicated their qualitative estimates of the various factors shown on the blank. The same general procedure was followed on the blanks for rating the parents and the child, supplemented by specific items of information, such as the occupation and education of the parents.

In reducing the ratings to a numerical scale so that they could be used for statistical analysis the following scheme was adopted: The field workers agreed upon a standard for each factor by means of which their estimates were converted into quantitative scores. A general measure of home rating was then made up by considering the following factors:

1. Material environment.
2. Evidences of culture.
3. Occupation of foster father.
4. Education of foster father.
5. Education of foster mother.
6. Social activity of foster parents.

Each of these six items was given a value from one to five. The total made up a home rating index ranging from six to thirty. The procedure followed in assigning values to these six factors will next be described.

1. The blank headed "Material Environment" contains five items. The first four of these were rated 1-2-3, the fifth 1-2-3-4-5. By totaling the points on these five items, values ranging from five to seventeen were obtained, which were then converted into a scale of from one to five.

2. The item "Evidences of Culture" occurs on the blank headed "Material Environment" and its rating is included in the total just described. Because of the great importance of this factor, it was thought advisable to give it additional weight by including it a second time—as a separate item.

3. The rating of the occupation of the foster father was made by the use of Taussig's five-fold classification, which was used as a scale for indicating roughly the occupational level.

4 and 5. In making the educational ratings of the foster parents the number of years of formal schooling was tabulated. The distributions of these ratings proved to be quite similar in form for both parents and indicated a natural grouping into five levels of schooling. The ratings were assigned thus:

(1). Less than 4 years.
(2). From 4 to 7 years.
(3). From 8 to 11 years—at least grammar-school education.
(4). From 12 to 15 years—at least a high-school graduate.
(5). From 16 to 21 years—at least a college graduate.

6. The rating of "Social Activity" was made on the basis of information given concerning social activity and use of leisure time on the blank entitled "The Parents." It is obvious that information of this sort does not lend itself readily to appraisal or to rating on a numerical scale.

Of the six ratings just described, the two concerning the education of the foster parents are the most objective and reliable. The others involve the usual difficulties in applying rating scales. While the Taussig classification of occupations is fairly objective, the use of the five categories as a scale furnishes only an approximate measure of occupational standing. Still further difficulties were encountered in making the other ratings, which required subjective estimates on the basis of rather indefinite standards.

In spite of the fact that not all of the ratings are as objective and valid as would be desired, it is felt that the sum total represents the best index of home environment that could be obtained from the information available.

5. The Groups Studied

It has already been noted that the present study was not confined to siblings, but that the original plan was extended to include cases to be used in more general comparisons.

The Binet test was given to a total of 671 children, 374 of whom were also given the International test. Of the total number of children tested, 401 were in foster homes. These cases, which furnish the chief data of the study, will be referred to hereafter as the "Home Group."

Of the remaining number tested, 146 were dependent children who had recently been committed to the care of the Illinois Children's Home and Aid Society, and were awaiting placement in foster homes. This group will be known as the "Newly Committed Group."

In order to make the comparison between a foster child and an own child raised in the same home, 36 own children were tested. These children, together with their foster-siblings, make up the "Foster-Own Group."

A miscellaneous group of 88 consisted chiefly of siblings to members of the Home Group and other wards of the Society who had not found permanent homes.

This total of 671 children furnished several groups, each of which was used as the basis for a particular study. These groups will next be described in the order in which they will be discussed in the following sections.

Pre-Test Group. There were 74 children of the Home Group who had been given a Binet test before they were placed in their present homes. These children were considered separately in order to make a direct comparison between their intelligence before placement and after a period of residence in a foster home.

Sibling Group. A group of 185 children had siblings (brothers or sisters) from whom they had been separated. Since several of the families contained more than two children, the group furnished 159 sibling pairs. When the period of separation was restricted to four years or more, the number of sibling pairs was reduced to 130. These are used for the chief comparisons between siblings separated.

Foster-Own Group. In 30 of the families included in the study there was at least one own and one foster child. Since more than one foster or more than one own child was found in 9 of the 30 families, it was possible to make 40 pairings of an own with a foster child.

Home Group. This group of 401 children in foster homes furnished data for various general comparisons between the child's intelligence and such other factors as grade of foster home, intelligence of foster parents, age of placement, and character of heredity.

Table I shows the character of the Home Group. It will be noted that about two-thirds of this group are legitimate children, who furnished practically all of the sibling cases. While a few colored children are included in the study, their number is not sufficient to alter any of the comparisons within the study, although they tend to lower the average of the entire group. It was found that by including this small percentage of colored children a greater range in grade of homes was secured. The inclusion of illegitimate children also served to extend the range, because these were more often found in the superior homes.

Tajle I.—Composition of the Home Group (401)

Legitimate (260)				Illegitimate (141)			
White 234		Colored 26		White 133		Colored 8	
Siblings	Not Siblings	Siblings	Not Siblings	Siblings	Not Siblings	Siblings	Not Siblings
164 (143 separated)	70	17 (14 separated)	9	10 (2 separated)	123		8

The location of the Home Group cases is shown in the distribution given below. It will be noted that approximately 40 percent of the cases were in Chicago homes. Chicago and its suburbs furnished over 60 percent. The remainder were widely distributed over the state of Illinois, since siblings were often found in homes quite distant from one another.

LOCATION OF CASES IN THE HOME GROUP

Chicago	166
Suburbs of Chicago	87
Cities of 20,000 or more	34
Cities from 5000 to 20,000	17
Cities below 5000	34
Rural districts	63
Total	401

II. PRE-TEST GROUPS

1. Introduction

If the intelligence of a group of children were measured before and after they had been under the influence of a certain environment, it would be possible to measure the effects of that environment independent of heredity.

Such conditions are met by a group of 134 children who were given mental tests several years previous to the present study, in the majority of cases just before their placement in foster homes. These tests were made by the Institute for Juvenile Research by means of the Stanford Revision of the Binet Scale. Since they were administered by the trained psychologists of the Institute, they may be regarded as reliable. The results of the tests may, therefore, be directly compared with those found in the present study.

The first tests on these children were made for various reasons, the chief being a suspicion of low mentality. In case a child showed signs of dullness, tests were usually given to the other children in the family also. Other children were tested because their parents were known to be defective. In some cases a prospective foster parent asked that the child be given a test in order to be assured that he had normal mentality. Occasionally, a child exhibiting abnormal behavior was given a mental test.

In order to trace separately the effect of different types of environment, the 134 cases were divided into three groups.

1. In 74 cases the child was placed in a foster home shortly after his first mental test and spent the entire intervening time in that home.

2. The second group is made up of 38 children who spent the interval between the tests in more than one foster home or institution.

3. The remaining 22 were children who were given their first test some time after their placement in a foster home.

It will be noted that these groups represent three different environmental situations. In Group I the environment changed, probably for the better, just after the first test and remained the same after that time. Since the children in Group II spent the interval between the tests in several environments, no complete measure of the influence could be obtained. In Group III the children lived in the same environment before and after the first test, hence it was impossible to determine the entire influence of the foster home by the change in intelligence during the interval.

Since Group I furnishes the most direct and useful comparisons, it will be discussed in more detail than the other two groups.

2. The First Pre-Test Group

This group of 74 contained 64 legitimate and 10 illegitimate children. Three of the group were colored children. The average age at the time of the first test was 8 years, considerably above the

TABLE II.—DISTRIBUTIONS OF INTELLIGENCE QUOTIENTS FOR THE PRE-TEST GROUP I

Intelligence Quotient	First Test Made by Institute for Juvenile Research	Later Test Made in Present Study
130–134	1	—
125–129	—	—
120–124	—	2
115–119	—	3
110–114	5	4
105–109	7	8
100–104	8	8
95– 99	8	13
90– 94	6	5
85– 89	13	8
80– 84	10	9
75– 79	9	7
70– 74	5	4
65– 69	—	2
60– 64	2	1
Total	74	74
Mean	91.2±1.05	93.7±1.1
Standard deviation	13.4	14.0

average placement age of the entire group of children studied. The mean age at the second test was 12 years, 2 months, indicating an average residence of 4 years in the foster home.

Table II shows the distribution of intelligence quotients for the two tests. It will be noted that the mean on the first test was 91.2, which is within the accepted normal range. The mean I. Q. on the second test was 93.7, indicating a mean gain of 2.5. By using the probable errors of the individual means as shown in the table and the correlation of .68 found between the two tests, the probable error of this difference was determined by the usual formula involving correlated means.[3] The observed difference may therefore be written 2.5 ± .8, which is about three times its probable error. While this difference is barely significant according to the usual test, it probably indicates some gain in intelligence due to residence in the foster home. Such a result could hardly be accounted for by practice effect, because of the long interval of time elapsing between the tests.

The increase appears much more significant than the actual figures indicate when considered in the light of certain other facts. A loss would not be surprising, since in the entire group of children tested there exists a negative correlation between I. Q. and age, which is due in large part to the nature of the Binet Scale. This drop with age will be discussed more fully in VI, 8, further on, where it will be shown that for the Terman unselected group, the probable value of the mean I. Q. for 8-year-olds was 104, while that for the 12-year-olds was only 99, as shown in Table XXXVIII. The observed gain of 2.5 points, therefore, might actually represent a gain of 7.5 points.

It should also be mentioned that the homes in which these children were placed were below the average of the entire Home Group —the mean indices were 16.7 ± .4 and 18.8 ± .2, respectively. If the gain in intelligence were due chiefly to an improvement in environment, a still greater gain would be expected for the entire Home Group. Unfortunately, pre-test data were available for only a small part of that group.

Another factor to be considered in interpreting this apparent change in intelligence is the period of the child's life spent in the

[3] $P.E._{M_1-M_2} = \sqrt{P.E.^2_{M_1} - 2r\,P.E._{M_1}\,P.E._{M_2} + P.E.^2_{M_2}}$.

foster home. The typical child of this group spent the years from eight to twelve in his new home, a period of life surely not the most formative in his childhood. If he had been under the influence of such an environment for four years at an earlier age, he would probably have made a considerably greater improvement in intelligence.

In order to throw further light on the effect of these factors, type of home and age of placement, the data were divided in two ways. As noted before, the mean home rating index of this group was 16.7. The children were put into two groups for which the home ratings fell above and below this value. The results for the two groups are shown in Table III.

These figures indicate that the children in the better homes made a gain of 5.3 points in intelligence. Since this value is four times its P. E. it may be regarded as significant. In the case of the poorer homes an insignificant gain was found (.1). Owing to the faulty standardization of the test, however, the gain of 5.3 would probably mean an actual gain of about 10.4 points, while the slight gain in the case of the poorer homes group would mean a gain of about 5.0 points. (See Table XXXVIII). These facts show that improvement in intelligence is directly related to the type of foster home.

The figures for the second division of the group, made on the basis of age, are shown in Table IV. The two groups consisted of children who were above and below the mean age of the entire group. The children of the older group spent the years from 10 to 14 in the foster home, while the interval for the younger children was from 6 to 10.

The observed gain for the younger group was 5.2 points, while for the older group there was a slight loss in I. Q. It might also be noted that both groups actually made significant gains, since about 5 points would need to be added in both cases to correct for the test. The difference of 5.6 points between the two gains is significant, which would indicate that improvement in intelligence is appreciably greater in the case of younger children.

Additional evidence of the influence of environment on intelligence is furnished by the correlations between the home rating indices and the intelligence quotients. The correlation between

TABLE III.—MEAN INTELLIGENCE QUOTIENTS OF THE PRE-TEST GROUP DIVIDED INTO BETTER AND POORER HOMES

Range in Home Rating	Cases	Mean Ages in Years and Months		Mean Intelligence Quotients		Change Between Tests	Standard Deviations	
		First Test	Second Test	First Test	Second Test		First Test	Second Test
17–30 (Better homes)	33	7–8	11–11	95.2±1.7	100.5±1.5	+5.3±1.3	14.6	12.5
7–16 (Poorer homes)	41	8–3	12–4	88.0±1.2	88.1±1.3	+0.1±1.0	11.3	12.6

TABLE IV.—CHANGE IN INTELLIGENCE OF PRE-TEST GROUP DIVIDED ACCORDING TO AGE

Age on Second Test	Cases	Interval Between Tests	Mean Intelligence Quotients		Change in Intelligence	Standard Deviations	
			First Test	Second Test		First Test	Second Test
12 yrs., 4 mos. or older	37	3 yrs., 9 mo.	89.7±1.6	89.3±1.7	−0.4±1.3	14.3	15.1
Less than 12 yrs., 4 mos	37	4 yrs., 6 mo.	92.8±1.4	98.0±1.3	+5.2±1.1	12.2	11.3

the home rating and the intelligence of the child at the time of his placement was $.34 \pm .07$ for this group. This indicates a certain degree of selection, that is, that there was some tendency for the superior children to be placed in the better homes. After an interval of about four years in the foster homes, the correlation was $.52 \pm .06$. This would indicate that the intelligence of the children more nearly corresponds to the type of home at the end of the period than at the beginning. This is further evidence that the change in intelligence is directly related to the type of environment under which the child has been living.

Taken as a whole, the results for this Pre-Test Group show, in our opinion, that a significant improvement in intelligence is associated with the type of foster home in which the children are placed. It is also shown that the younger children made the greater improvement. These gains appear even more significant in view of the observed tendency for the I. Q. to drop with age. The differences found would probably have been even more striking had the children spent an equal period at an earlier age in higher grade homes.

3. Other Pre-Test Groups

As was noted in the first section, there were 134 children who had been given mental tests several years before the present study was made. The same types of comparison as those just discussed were worked out for Groups II and III, in order to note the changes that occurred during the interval between the tests. Table V gives a summary of the results for the three groups.

It will be noted that, while the observed gain for the entire group of 134 children was only .8, this result is raised to 5.6 when adjustment is made for the standardization of the test. When the results for the three sub-groups were considered, it was found that Group I had made the greatest gain (7.5 points). The gains of the other two groups were both about three points and are hardly significant.

To interpret these results, it will be necessary to recall the character of the three groups. Group II was composed of children who had been shifted about during the interval between the tests. Although each of the homes in which a child lived might have been

TABLE V.—CHANGE IN INTELLIGENCE FOR ALL PRE-TEST GROUPS

Group	N	Mean Ages		Time Interval	Mean I. Q.'s		Change in I. Q.		S. D.'s	
		First Test	Second Test		First Test	Second Test	Observed	Estimated Actual	First Test	Second Test
I	74	8–0	12–2	4–2	91.2±1.0	93.7±1.1	+2.5± .8	+7.5± .8	13.4	14.0
II	38	10–1	13–9	3–8	86.3±1.5	85.0±1.6	−1.3±1.2	+3.1±1.2	14.2	15.1
III	22	9–5	12–8	3–3	87.2±2.5	86.2±1.9	−1.0±1.8	+2.9±1.8	17.3	12.9
Combined Group...	134	8–9	12–9	4–0	89.2± .8	90.0± .9	+0.8± .7	+5.6± .7	14.4	14.7

good in comparison with his own home, it was impossible to measure their total influence or to estimate the effect of the various shiftings and unsettled nature of his life during the period. When one considers the influence under which these children lived during the interval, as well as the fact that they averaged 10 years of age at the time of the first test, it does not seem surprising that they failed to gain as much as Group I.

Group III contained 22 cases in which the first test was given after the child had been in the foster home for some time (mean time, 4 years, 9 months). The group shows no significant gain in intelligence between that test and the later one. The inference might be that improved environment exerts its influence relatively quickly, that is, within a few years. If the child's intelligence increases to its final level soon after placement and then remains fairly constant, this group would not be expected to show a significant gain in intelligence.

The large gain made by the children of Group I appears even more significant in the light of the smaller gains made by the other groups. The children in this group were those who were placed in foster homes just after their first test and who have remained in those homes since that time. This constitutes the chief difference between this group and the other two. It would seem, therefore, that their greater gain was due, at least in part, to their more favorable environment. It should be noted that the children of Group III may have at some time made a gain as great as that of Group I, but this could not be determined in the present comparisons, because there was no record of their intelligence before they entered the foster home.

The results on these Pre-Test groups have shown that children placed at eight years of age in relatively poor homes gain significantly in intelligence. A greater gain might be expected had they been placed at an earlier age in higher grade homes.

III. Sibling Groups

1. Introduction

In the study described in the preceding section the factor of heredity was controlled, since the comparisons made were of the same group of children, to whom successive intelligence tests had been given. The present section describes a second attempt to hold the heredity factor constant so as to obtain independent measures of environmental influence. Comparisons will be made between children of the same family who had been placed in different homes. If, as seems probable, differences in heredity are negligible when many such pairs of siblings are considered, the observed differences in intelligence may then be ascribed to environment.

The entire Sibling Group consisted of 185 children who had been separated for periods ranging from 2 to 13 years. Not all the group had been continuously in one foster home up to the time the study was made. Such conditions were not satisfactory for the direct comparisons mentioned above, but this "Main Group," as it will be called, was used for studying certain general relationships by means of correlations.

By limiting the comparisons to pairs separated at least four years, with continuous residence in a foster home, a group of 159 children remained. These children, referred to hereafter as the "Limited Group," furnished 130 sibling pairs. The relatively large number of pairings obtained was due to the fact that in some cases several children came from the same family. Thus, out of a total of 64 families, there were 4 instances of four children in a family, while in each of 23 families three children were found. The remaining 37 families furnished one pair each.

The distributions of intelligence quotients at the time of the study are shown in Table VI. It will be noted that the means for both groups were practically the same and that both were somewhat below normal.

As in the case of the Pre-Test groups, some illegitimate and some colored children are included. In both the Main and Limited groups there is one pair of illegitimate siblings (same father) and one pair of twins. There were also 15 colored children from six families, furnishing 12 pairs of siblings. The relatively low average of intelligence found is probably due to several factors, one of which is the small proportion of illegitimate children in the present groups. These children, as shown later, were the most intelligent of those tested in the study. Other factors are the relatively late placement age and low home ratings of these groups as compared with the Home Group treated in the sixth section of this chapter.

TABLE VI.—DISTRIBUTIONS OF INTELLIGENCE QUOTIENTS OF SIBLING GROUPS

Intelligence Quotient	Main Group	Limited Group
125–129	2	2
120–124	1	—
115–119	5	5
110–114	14	11
105–109	11	8
100–104	24	20
95– 99	17	17
90– 94	23	19
85– 89	21	19
80– 84	18	14
75– 79	18	15
70– 74	18	15
65– 69	5	7
60– 64	6	5
55– 59	2	2
Total	185	159
Mean I. Q.	90.9 ± .7	90.4 ± .8
S. D.	15.0	15.0
Mean Home Rating	16.7	16.7

The more general comparisons made with the Main Sibling Group will next be discussed, chiefly by the use of correlations. These results will assist in interpreting those found in the subsequent analysis of the Limited Group.

2. Main Sibling Group

Correlations between the six factors studied in this section are presented in Table VII. It should be noted that pairs of siblings are not considered in any of these correlations, but that all of the children are taken as a group, regardless of their relationship to one another. Several of the results found here will be substantiated by similar correlations worked out for the larger Home Group in the sixth section.

The whole group of correlations indicates the very complex interrelationship of the factors. Most of the coefficients in the table must be interpreted in the light of the influence of other variables upon the results found. The method of partial correlation was therefore used for the purpose of holding constant as many of the conditioning factors as possible.

Some evidence of selection according to age is furnished by the negative correlation of −.31 between age of placement and home rating. This shows a tendency for parents in better homes to adopt younger children. The correlation becomes −.42 if the time in the home is held constant.

The correlation between home rating and intelligence is .32, indicating a tendency for the children raised in better homes to be brighter. This correlation is somewhat lowered when age of placement is held constant ($r_{12.4} = .24 \pm .05$), showing that the relationship is due in part to the fact that parents in superior homes adopt younger children. If age at the time of the test is held constant, the correlation of .32 is reduced to .26 ± .05. This reduction is probably due to the fact that the older children tend to be in poorer homes ($r = -.20$). When both age of placement and age at test are held constant, the correlation between intelligence and home rating is reduced to .24 ± .05.

Table VII.—Correlations for the Main Sibling Group of 185 Children

Factors	1	2	3	4
1. Intelligence (I. Q.)				
2. Home Rating......	.32 ± .04			
3. Time in Home	−.11 ± .05	.11 ± .05		
4. Age of Placement..	−.30 ± .05	−.31 ± .04	−.51 ± .04	
5. Age When Tested..	−.36 ± .04	−.20 ± .05	.44 ± .04	.56 ± .03
6. Percent of Life in Home.........	.18 ± .05			

The correlation of −.11 between intelligence and time in the foster home is probably negligible in view of its P. E. It is, however, raised to .06 ± .05 if age tested is held constant. While this value is very small, it is at least in the direction to be expected. Evidently, the child's intelligence bears no significant relation to the number of years he has spent in his foster home. A more significant correlation results ($r = .18$) when the percent of his life spent in the foster home is used. This would seem to indicate that a certain number of years in the foster home may mean more in the early period of his life than later, since the percent would be greater if the child had been placed at an early age.

In the comparisons of this section the siblings have been considered as though they were unrelated individuals. The next section will also deal with the Limited Group as unrelated cases and discuss the possible influence of selection upon correlations with home rating.

3. The Selective Factor in Correlations with Home Rating

Before considering the siblings as pairs, it will be well to note an important correlation found for the Limited Group as a whole. The correlation between the intelligence of these 159 children and their foster home rating was $.40 \pm .04$, indicating a tendency for the children raised in better homes to be brighter. A point to be considered is the extent to which selection might affect this relationship. If the better parents had adopted brighter children, that is, made their selection on the basis of intelligence, the observed correlation might have been due to this fact alone and could not be attributed to subsequent environmental influence. While it is impossible to determine the amount of actual selection for this group, a consideration of the matter from several angles has led to the conclusion that selection probably exists, but not in sufficient amount to account entirely for the strong correspondence found between the child's intelligence and his foster home.

In the discussion of the Pre-Test Group it was stated that there was a correlation of .34 between the child's I. Q. at the time of adoption and the type of home in which he was placed. This correlation represents the amount of selection for this particular group, but it is important to note that this group had the most favorable conditions for selection of any included in the study. The child's standing on the Binet test was known to the Society placement workers and, in the majority of cases, to the prospective foster parents. The fact that this initial correlation of .34 was raised to .52 after a residence of only about four years in the foster home would indicate that the home had some effect upon the child's intelligence. The difference in the correlations is evidently significant, since it is almost three times the probable error of either coefficient. It would seem that the increase in the correlation was due to the influence of environment.

There are other reasons why the Pre-Test Group affords more opportunity for selection than other groups studied. Besides the fact that the children of this group had had a mental test, it should be noted that they averaged 8 years of age at the time of their placement. If foster parents were selecting on the basis of their judgment of the child's intelligence, such selection would be more easily made with children of this age than with those much younger. The children of these sibling groups averaged about five years of age at the time of their placement, while those of the Home Group were even younger—over one third of them were placed before they were two years of age.

Besides being relatively old at the time of their placement, the children of this Pre-Test Group were placed in the more recent years when more attention is being paid to "fitting the child to the home." For this reason, the degree of selection in this group would probably be greater than that in the Sibling Group, the members of which were placed in foster homes much less recently.

The very fact that some of the pre-tests were given at the request of the foster parents indicates the likelihood of a greater amount of selection with these children than with those for whom such requests were not made.

A study of the family history of the child would probably furnish the chief data for selection. In the years in which the most of the siblings were adopted such records were relatively meager. In the case of the children of the Pre-Test Group, most of whom were placed within the last five years, prospective foster parents and Society workers can refer to more complete family records in order to judge the probable mental ability of the child in question.

These facts all go to show that, although selection probably exists to some extent with all groups, it can hardly be so great in other groups as in the Pre-Test Group, and must therefore be considerably less than would be indicated by a correlation of .34. Selection, therefore, could hardly account entirely for the observed correlation of .40 between intelligence and foster home rating for this Limited Sibling Group.

4. Comparisons of Siblings Separated

Correlations between siblings are of a different nature from those discussed so far. In an ordinary correlation, measures of two different traits are obtained for each individual, while in a sibling correlation each of a pair of siblings is measured with respect to the same trait. The proper method of pairing the entries in the correlation table thus becomes a matter of importance. In the present discussion the siblings were paired in two ways. According to the first method, the score for the older of a pair was always tabulated in the same direction on the table, while in the second scheme every pair of scores was entered twice. The double entry method produces a symmetrical table, since each pair of scores, *e. g.*, 105, 97, is also entered in the reverse order, as 97, 105.

It will be recalled that the Limited Sibling Group consisted of 125 pairs of siblings who had been separated in different foster homes for from 4 to 13 years. These pairs were formed by taking all possible combinations in a given family. Thus, a family with four children furnished six pairs, while a family of two produced only one such pair. In the double entry table there were of course twice as many entries as pairings.

The average age at separation of these pairs was 5 years, 4 months. The mean age at the time of the test was 12 years, 8 months, indicating an average separation period of over 7 years.

One of the most important correlations found in the study was that between the present intelligence of these siblings. This coefficient was .34 ± .05 by the method based on age and .25 ± .06 by the double entry scheme. Both of these results are considerably lower than the correlations usually found for siblings reared together. The "coefficient of heredity," as it is called, is approximately .50. It thus appears that residence in different homes tends to make siblings differ from one another in intelligence.

A factor which affects this relationship is the age at which the siblings were separated. For the siblings of this group who were separated after both were five years of age, the correlation between their intelligence was found to be .49 ± .08 on the age basis (N = 38) and .43 ± .09 by double entry. These results are similar to the usual coefficients of fraternal resemblance. The corresponding correlations found in the case of siblings separated before either was six years old (46 cases) were .32 ± .09 and .25 ± .09 respectively. Separation in foster homes at the early formative years of life thus appears to account in a large measure for the lowered heredity coefficient.

These correlations show that siblings separated quite early in life develop greater differences in intelligence than do those who remain together until they are at least five years old. The data were further examined to see if those placed in contrasting homes became more dissimilar than those raised in homes of the same type. There were very few cases in which the foster homes of a pair differed in any marked degree. In fact, the mean difference was only 4.6 points when the possible difference was 24 points. The correlation between the home ratings of a pair was .22 ± .06.

In order to study the differential effect of the homes upon the resemblance of the siblings, the 125 pairs were divided into two groups according to whether the difference in the home ratings of the pair was more or less than 4 points. The correlation between the intelligence of the siblings raised in the more similar homes (62 pairs) was .39 ± .07 by the age and .30 ± .08 by the double entry method. For the pairs in widely different homes (N = 63) the correlations were .28 ± .08 and .19 ± .08, respectively. These results furnish some evidence that siblings raised in widely different environments are less alike in their mental ability than are those raised in the same or in similar environments.

It will be remembered that the 125 pairs of siblings had been separated for periods of from 4 to 13 years. The correlation between the intelligence of the 59 pairs who were separated less than 7 years was .41 ± .07 by age entry and .23 ± .08 by double entry. For those separated more than 7 years the corresponding correlations were found to be .27 ± .08 and .21 ± .08, respectively (66 pairs). While there is little difference in the coefficients by the double entry method, there is a slight indication that the length of the separation period affects the correlation between the intelligence of siblings.

A summary of the correlations just mentioned is given in Table VIII. On the whole, the coefficients are smaller than those usually found for siblings reared together. This difference is evidence that intelligence is affected in some degree by the environment in which

TABLE VIII.—CORRELATIONS BETWEEN THE INTELLIGENCE OF SIBLINGS WHO HAVE BEEN SEPARATED AT LEAST FOUR YEARS

Method of Pairing	Entire Group	Before Six Years of Age	After Five Years of Age	More Than Seven Years	Less Than Seven Years	Unlike Homes	Like Homes
Age entry..........	.34	.32	.49	.27	.41	.28	.39
Double entry.......	.25	.25	.43	.21	.23	.19	.30
N................	125	46	38	66	59	63	62

a child lives. As has been pointed out, the lowest correlations were found when the siblings had been separated at an early age and had been for a considerable number of years in widely different homes.

Our correlations have shown that siblings raised apart resemble each other less than do siblings who are raised together. The next question is whether or not the differences of a pair are consistent with the type of environment under which each has lived—that is, whether the child in the better home tends to be the brighter of the pair. Another point to be considered is the effect of being the younger of a pair at the time of placement. It was found that in 85 of the 125 pairs the brighter sibling was the one in the better home, while in 80 cases the younger child of the pair had the greater intelligence. A study of mean intelligence quotients for the group divided on these two bases will best show the influence of the two factors.

The average intelligence quotient of the siblings in the poorer homes (poorer as regards the two homes of the pair) was found to be 85.7 $\pm$ .8 as compared with an average of 95.0 $\pm$ .9 for those in the better homes. The observed difference of 9.3 points is clearly significant, since its probable error is only 1.0. When this difference is corrected for the ages of the two groups (from Table XXXVIII), the difference in intelligence is reduced to about 6 points, which is still significant. This gives additional evidence that the intelligence of children is affected by the environment in which they are raised. Siblings tend to differ from one another as the environment differs and the child in the better home is usually found to be the brighter.

As has already been noted, the two factors, home rating and age of placement are quite closely related. A correlation of $-.34 \pm .05$ between the two for this Limited Sibling Group indicates that there was a strong tendency for the better homes to take the younger children. In this correlation the siblings are considered individually, that is, no account is taken of their paired relationship. That the tendency also exists if only the two children of a pair are considered is shown by the fact that in 74 of the 124 pairs the younger child was placed in the better home. The mean home rating of the younger siblings was $16.9 \pm .3$, while that of the older was $15.9 \pm .3$. It is evident, then, that a part of the observed difference in intelligence may be due to the fact of younger placement. In order to study this factor separately, the siblings were divided into two groups, one group consisting of the 129 who were the younger members of the pairs, the other of the 129 older members.

The mean intelligence quotient of those who were placed younger was $94.3 \pm .9$, of those placed older, $86.4 \pm .9$. The standard deviations were 15.2 and 14.5, respectively. The observed difference in the means (7.9 ± 1.0) is clearly significant. A part of this difference, however, is due to the standards of the Binet Test. When correction is made for the difference of 3 years, 5 months, in the ages of the two groups, the observed difference in the means is reduced to 3.8 ± 1.0, which is still significant.

It is evident that being placed young and being placed in a good home are both influential factors in the development of intelligence. The two factors are very closely related. In an attempt to discover which was more influential a study was made of the 49 pairs in which the advantage of being placed in the better home was counteracted by the disadvantage of being older at the time of placement and *vice versa*. The mean I. Q. of the 49 children placed older but in the better homes was 89.5 ± 1.5, while that of their younger siblings placed in poorer homes was 87.7 ± 1.4. The difference in the means is only 1.8, but when corrected for age (by Table XXXVIII) it becomes 5.2 ± 1.5. It would thus appear that improvement in intelligence is more dependent upon type of home than upon age of placement.

Additional supporting evidence of the influence of a good home and early placement is furnished by the scores on the International

Test. This non-verbal test was given to 136 of the siblings, and thus furnished material for the comparison of 96 sibling pairs. The first comparison will be made between the siblings according to whether they were in the better or the poorer home of the pair.

The mean International Test score of the siblings in the better homes was 140, while that of those in the poorer homes was 138. These point scores are all that the International Test gives in its present stage of development. They are, of course, not at all comparable unless the ages of the two groups are the same. In this case the mean age of the siblings in the better homes was 12 years, 6 months, while that of the siblings in the poorer homes was 13 years, 8 months. It appears, therefore, that the higher average was made by the younger group—the children in the better homes, who were over a year younger than their siblings in the poorer homes. This result points to the influence of a good home upon intelligence when intelligence is measured by a non-verbal scale.

In order to make a more exact comparison of these groups the cases were matched for age so that the mean scores would be comparable. When this was attempted, it was found that only 76 of the pairs could be included. This reduction was due to the fact that there was an excess of younger children in the better home group. Many of these could not be matched for age with children in the poorer home group because it contained an excess of older children. The results found for these 76 pairs are given in Table IX.

TABLE IX.—COMPARISON OF SIBLINGS IN BETTER AND POORER HOMES ON THE INTERNATIONAL TEST (76 PAIRS)

Group	Mean Age in Months	Mean Score	Standard Deviation
Sibling in better home of pair.......	158.3	146.2 ± 2.9	37.6
Sibling in poorer home of pair.......	158.5	135.4 ± 2.9	36.9

The difference in the mean scores is 10.8 ± 3.4 in favor of the siblings in the better homes. This difference is over three times its probable error and is therefore quite probably significant. It is, however, not so significant a difference as that found in the case of the Binet Test. This lack of agreement in degree might be taken to mean that the type of intelligence measured by the Binet Scale is more susceptible to environment, chiefly because of its verbal character. It is probably partly due to the fact that the number of cases used is much smaller and that the method of pairing necessary forced the exclusion of cases which by their very nature would have made a still greater difference between the means of the two groups.

The next comparison was made between the International Test scores of the siblings according to their relative age at placement. The mean scores for the two groups were found to be 133 for the younger of the pairs and 147 for the older. The scores are not at all comparable in this form, for the mean chronological ages of the two groups were 11 years, 4 months, and 14 years, 9 months, respectively, indicating a difference of nearly 3.5 years in chronological age. The scheme of pairing was again resorted to, though much greater difficulties were encountered than before, since it was necessary to match for age the members of an older group with those of a younger group. It was possible to use only 51 of the 96 pairs. The results shown in Table X are very similar to those found when type of home is considered.

TABLE X.—INTERNATIONAL TEST RESULTS FOR SIBLINGS DIVIDED ACCORDING TO THEIR RELATIVE AGE AT PLACEMENT

Group	Mean Age in Months	Mean Score	Standard Deviation
Sibling placed younger	154.6	148.7 ± 3.1	33.1
Sibling placed older	155.4	135.8 ± 3.2	33.9

The table shows that the children who were the younger members of the pairs made scores averaging 12.9 points higher than the scores of their older siblings. This difference is over three times its probable error and would be regarded as significant. Here, also, the cases necessarily omitted were such as to have made the result even more significant, could they have been included.

The results found for the International Test substantiate those for the Binet Scale. Such a check is of great importance in view of the distinctly different nature of the two tests. As already mentioned, the International Test differs from the Binet Scale in several respects. Its material is non-verbal in character, relatively freer from environmental influences, and less susceptible to coaching. The International Test probably gives a better measure of intelligence in the upper age levels than the Binet Scale, since it has a greater range in that direction. Moreover, the comparisons made with the International Test results were based on a point score, while the corresponding comparisons with the Binet Test were made by the use of a quotient.

The two tests seem to agree in showing the influence of good home environment upon intelligence. It would be well, however, to again consider what possible influence selection might have had in the production of the results. As has been pointed out, selection may affect comparisons in which home ratings are involved because the parents in better homes may have chosen brighter children for adoption. In the comparisons made in this section such selection could not affect the results to any appreciable extent, unless there was selection of the brighter sibling of a pair. Since the differences between siblings are much less than between children in general, the effect of such a choice would be relatively slight, even were such selection possible. In these cases the family history could not give a basis for selection, since it would be the same for the two children.

It has been noted before that in two-thirds of the sibling pairs the younger child was placed in the better home. This indicates some selection according to age, a basis of choice which was much in evidence throughout the study. If there were a relationship between intelligence and late birth in the family, then selection of the younger child would also mean selection of the brighter. In order to see if such a relationship existed, comparisons of older and younger children in a family were worked out for two groups.

The data for the first comparison consisted of Binet intelligence quotients for 77 pairs of siblings in the University of Chicago Elementary School. Each pair of siblings had been reared together and had attended the same school all their lives. The pairs used were those for which there was a difference of less than two years in test age. The mean I. Q. of the younger of the pairs was 120.4 ± 1.1, while that of the older was 118.2 ± 1.1. The difference between these two averages is 2.2 ± 1.1, which may be regarded as insignificant. It would, of course, be even smaller if correction were made for the slight difference in age at the time of the test.

Additional evidence is furnished by data from the University of Iowa Elementary School. It was possible to use 29 pairs for which the test ages differed less than a year. The mean I. Q. of the older siblings was 116.0, while that of the younger was 115.6, showing no significant difference.

These results indicate that order of birth in a family is unrelated to intelligence. It would seem, therefore, that selection of the younger child of a pair would not imply selection of the brighter. It is probable that the better families adopted younger children because they realized the importance of raising the child during the early formative years of his life.

Even if there were some slight selection according to intelligence, only the comparisons of means would be appreciably affected by it. The low correlations found between siblings who had been reared apart could hardly be influenced by such selection.

IV. Foster-Sibling Groups

1. Introduction

In the previous sections an attempt was made to control the factor of heredity so that the effect of environment might be observed. In the present section comparisons will be made between children whose environment is practically the same but whose heredity is different. Such conditions are met when two unrelated children have been raised in the same home.

It was possible to find thirty homes in which there was at least one own and one foster child. These families furnished 36 own children and 34 foster children distributed as shown below:

21 homes—one foster and one own
4 homes—two foster and one own
4 homes—one foster and two own
1 home—one foster and three own

These cases furnished 40 pairings of an own with a foster child.

Other cases which met the conditions were furnished by homes in which two or more unrelated foster children had been raised.

The number of pairings of this type was 72. Although both types of cases furnish pairs of unrelated children raised together, they are sufficiently different to warrant separate treatment.

2. Foster-Own Group

The distributions of intelligence quotients for the Foster-Own Group are shown in Table XI. It will be noted that the own children are superior, not only to their foster siblings, but also to children in general. Their mean intelligence quotient is 112.4 ± 1.6 as compared with that of 95.1 ± 1.7 for the foster children. The latter mean is increased to 99.9 if only those who entered the home before three years of age are considered. The two groups have about the same variability, as the standard deviations indicate.

TABLE XI.—DISTRIBUTIONS OF INTELLIGENCE QUOTIENTS FOR THE FOSTER-OWN GROUP

Intelligence Quotient	Own Children	Foster Children
130–134	3	—
125–129	4	—
120–124	4	2
115–119	7	2
110–114	4	1
105–109	5	5
100–104	3	5
95– 99	1	2
90– 94	3	2
85– 89	—	5
80– 84	—	3
75– 79	2	4
70– 74	—	3
Total	36	34
Mean	112.4±1.6	95.1±1.7
S. D	13.9	14.8

Although the environment for these two groups was substantially the same, the period of residence for the own and the foster children was quite different. The mean age of the own children at the time of the test was 10 years, 3 months, while that of the foster children was 11 years, 5 months, as shown in Table XII. Since the mean age at which the foster child entered the home was 4 years, 8 months; only 6 years, 9 months, was spent in the foster home. While the own children had been in the home since birth, the foster children had been under the home influence less than two thirds of their lives. Moreover, the foster children were not in the home during the early formative years. It is evident, therefore, that the environmental influence was by no means equal for the two groups.

TABLE XII.—COMPARISON OF PERIODS OF HOME INFLUENCE

Child	Mean Age Entered Home	Mean Age at Test	Mean Time in Home
Own........	Birth	10 years, 3 months	10 years, 3 months
Foster.......	4 years, 8 months	11 years, 5 months	6 years, 9 months

There is a possibility that, even though a foster and an own child had been in a home for practically the same period, the environmental influence might still have been different. The attitude of the parents toward their own child might not have been the same as that toward the adopted child. As far as the field workers were able to observe, however, there were no evidences of any such distinctions.

The thirty homes in this group were distinctly superior to those of the groups previously discussed. The mean home rating index was 20.7, as compared with that of 18.7 for the entire Home Group. This was accompanied by a high level of intelligence for the own children, whose mean intelligence quotient was 112.4.

The correlation between the intelligence of a foster child with an own child in the same family was $.34 \pm .09$ (40 pairs). This indicates considerable resemblance, which is probably due, for the most part, to the fact that the children have been raised in the same home, since it was found that, when the home rating was held constant, the observed correlation was reduced to $.15 \pm .10$.

The correlation between the intelligence of the own child and his home rating was .47 ($N = 36$), while the corresponding relation between the foster child and the home was also .47 ($N = 34$). It might at first seem impossible that the foster children could bear as close a relationship to the home as do the own children, especially since the periods of the home influence were so unequal. It is probable, however, that a part of the observed correlation in the case of the foster children is due to selection. On the other hand, if the children had been adopted earlier, their resemblance to their environment might have been even greater. The observed correlation might not have been raised, however, because of the lesser possibility of selection at the earlier ages.

The correlation of .47 in the case of the own children is a measure of the combined effect of heredity and environment upon intelligence. Since the corresponding correlation of .47 for the foster children involves selection and environment, the comparison between the two coefficients is complicated.

It was also found that the correlation of the intelligence of the own children with that of their parents (mid-parent Otis Test score) was $.35 \pm .11$ ($N = 28$). The corresponding correlation for the foster children was $.18 \pm .13$ ($N = 26$). The difference, $.17 \pm .14$, is evidently insignificant.

From the data for this group it is apparent that adopted children are considerably lower in intelligence than own children in the same homes. It is probable that the difference is due partly to heredity and partly to early environment. The intelligence of the foster children would quite probably have been greater had they been adopted earlier in life. Comparisons made later show that the earlier children are adopted, the higher is their intelligence.

Significant correlations were found when the intelligence of the foster children was compared with that of own children or with the rating of the foster home.

3. Foster-Foster Group

This group consists of 72 pairs of unrelated foster children raised in the same family. Two children of a pair in this group had different heredity, but substantially the same environment for a considerable part of their lives. Since both members of the pair were adopted children, the hereditary relationship to the family was not involved, as was the case in the Foster-Own Group.

A correlation of $.37 \pm .07$ (by both age and double entry) shows a distinct resemblance between the intelligence of two foster children raised in the same home. A part of this correlation may be due to a possible tendency for a family to adopt two children of about the same level of intelligence. On the other hand, the observed correlation might have been greater had the two children been in the home the same length of time and at equally influential periods.

4. Entire Foster-Sibling Group

By throwing the two preceding groups together it was possible to obtain 112 pairs of unrelated children raised in the same home. The correlation for this group on the Binet Test was $.31 \pm .06$ by age entry and $.25 \pm .06$ by double entry. Substantially the same result ($r = .27 \pm .08$) was found when 56 of the pairs were compared on the International Test. For this last correlation it was, of course, necessary to eliminate the age factor by means of partial correlation.

It may be further noted that a correlation of $.18 \pm .07$ ($N = 76$) was found between the school achievement ratings of two unrelated children raised in the same home. This rating was based on age-grade placement and character of school work done. Such an index is obviously not satisfactory in the case of children in the earliest school grades. This, as well as certain accidental factors in age-grade placement, might account for the relatively low correlation found.

Through the courtesy of Professor Terman, a number of additional cases meeting the stated conditions were furnished. When these are included, the following correlations are obtained:

Intelligence of foster with own..............	.38	(N = 47)
Intelligence of two foster in the same home....	.40	(N = 93)
Intelligence of two unrelated in the same home.	.34	(N = 140)

The results on this larger number of cases substantiate those already discussed.

In conclusion, it should be noted that there was probably some tendency for two children of similar mental ability to be placed in the same foster home. If such selection did not exist, the expected correlation between the intelligence of the two children at the time of their adoption would be zero. It would still be zero after several years of residence in the same home unless the intelligence of the children was modified by the new environment. Since all of the observed correlation of .40 could hardly be due to selection, it is highly probable that such modification actually took place. The influence of environment, therefore, is shown by the fact that when two unrelated children are raised in the same home, differences in their intelligence tend to decrease.

V. Characteristics of Own and Foster Parents

1. Introduction

In the previous sections attempts were made to control either the factor heredity or the factor environment. This control was possible only in relatively small groups which met the necessary requirements. The Home Group to be considered next contains a much larger number of cases, but does not permit of such a sharp distinction between the two factors.

Before considering the results found for this group of children, it will be well to consider the characteristics of both the own and the foster parents. Such information will furnish a basis for comparing the heredity and early environment of the children with their later environment in the foster homes.

2. The Foster Parents

The occupational status of the foster fathers on the basis of the Taussig classification is given below in Table XIII. In order to

Table XIII.—Occupational Status of Foster Fathers as Compared With the Male Population of Illinois[a]

Class	Foster Fathers	Males in Illinois
	Percent	*Percent*
Professional	15	3
Semi-professional and business	41	35
Skilled labor	38	39
Semi-skilled to slightly skilled labor	5	15
Common labor	1	8
Total	100	100
Number of cases	394	1,419,950

[a] In this and in other tables in this section the entries are made on the basis of the child. Thus, in Table XIII a foster father is included as many times as there were foster children in the home. It was thought that this method would give the fairest picture of the parental influence of the entire group of children.

show how this group compares with the general population, a similar distribution for the entire state is included in the table. This latter distribution is based upon a report from the 1920 census giving the number of persons engaged in various occupations. These occupations were grouped according to the Taussig scheme and percentages figured.

It will be noted that 15 percent of the foster fathers are in the professional class, whereas for the entire state only 3 percent are thus rated. This difference of 12 has a probable error of only 1 and is therefore clearly significant. On the other hand, the percentages in the semi-skilled and common labor groups are much smaller in the case of the foster fathers. These facts clearly indicate that the foster fathers have a higher occupational status than have the men of the entire state in which they live.

TABLE XIV.—SCHOOLING OF THE FOSTER PARENTS AS COMPARED WITH THAT OF THE UNITED STATES ARMY

Schooling	Foster Fathers	Foster Mothers	Army
	Percent	*Percent*	*Percent*
16 years or more (College graduation)	16	8	7
12 to 15 years (High-school graduation)	21	28	11
8 to 11 years (Grade-school graduation)	31	31	31
4 to 7 years	27	27	37
Less than 4 years	5	6	14
Total	100	100	100
Number of cases	394	401	91,487

The amount of scholastic training of both foster fathers and foster mothers is given in Table XIV. An inspection of these distributions shows that 16 percent of the fathers had at least a college education, while only 8 percent of the mothers had reached that level. Combining the two upper groups shows that practically the same percentage of fathers as mothers had a high-school education—somewhat over one third of the group. Since no comparable

TABLE XV.—SCORES MADE BY FOSTER PARENTS ON THE OTIS SELF-ADMINISTERING TEST, HIGHER EXAMINATION, FORM A

Score	Foster Father	Foster Mother
65–69	1	—
60–64	2	1
55–59	7	—
50–54	13	5
45–49	23	9
40–44	19	20
35–39	16	37
30–34	36	45
25–29	27	33
20–24	14	39
15–19	11	35
10–14	8	19
5– 9	3	9
0– 4	—	3
Total	180	255
Mean score	35.1±.6	28.1±.5
S. D.	12.6	11.1

data were available on the education of the general adult population, data from the report on the United States Army were used.[4] The army group was composed of officers and drafted men, the former constituting one seventh of the entire group. Although the comparison of such dissimilar groups may not be of great value, it appears that the education of the foster parents is distinctly better than that of the men in the army.

The average amount of schooling of the foster parents was slightly above 9 years. While no figures are available on the amount of schooling for the population in general, the usual estimate for adults of the same generation as these foster parents is about 6 years. There can be little doubt, therefore, that the foster parents as a group have had considerably more than the average amount of formal education.

Table XV gives the distributions of the two parents on the Otis Self-Administering Intelligence Test (20-minute time-limit).

[4] R. M. Yerkes. *Intelligence Examining in the United States Army.* National Academy of Sciences. Volume XV. Washington, 1921. Page 758.

The average scores of the father and mother were 35.1 and 28.1, respectively. These were converted into intelligence quotients according to Otis' interpretation chart, using 16 years as a basis. The mean intelligence quotient of the fathers is thus found to be 106, while that of the mothers is 97. While these are both fairly close to normal, it was felt that many persons taking the test were not adequately measured. The long time elapsing since many of the group had been in school and the unusual nature of such a performance led to considerable embarrassment. Although every effort was made to put the parents at their ease in taking the test, a good many seemed to feel that they were on trial and because of their nervousness probably failed to do themselves justice. Another difficulty arose over the language handicap of some of the parents. In view of all these disturbing factors, it was felt that the measure of intelligence found was too low, especially in the case of the mothers.

In addition to the Otis Test a Vocabulary Test was given. Since no norms are available, no comparisons with the general population can be made. It may be noted, however, that the mean score of the fathers was 35, that of the mothers 31.

In view of these facts it is apparent that the group of foster parents were above the average of the general population in occupational level, education, and intelligence. The close inter-relationship of these factors is shown by the correlations given below.

Father's intelligence with home rating	.62
Mother's intelligence with home rating	.56
Father's intelligence with his occupation	.57
Intelligence of husband and wife	.49
Intelligence of husband and wife if type of home is held constant	.22
Father's education with his intelligence	.72
Mother's education with her intelligence	.63
Father's vocabulary with his education	.69
Mother's vocabulary with her education	.69
Mid-parent vocabulary with mid-parent intelligence	.78

(Note: The P.E.'s for these correlations range from .02 to .05.)

3. The Own Parents

The information concerning the own parents was rather meager, especially in the case of those whose children who had been committed for the longer periods. The case histories were based upon information obtained from various sources, such as the family itself, neighbors, county officials, local probation officers, and ministers. The family records were most complete in the case of the legitimate children. Full information on the illegitimate cases was obviously difficult to obtain, especially that concerning the father. In spite of these limitations, it was found possible to obtain fairly complete data on such points as occupational status, education, and mental and moral characteristics.

The occupations of the own fathers were grouped according to the Taussig classification. The results of this grouping for the fathers of the legitimate and illegitimate children, together with the distribution for the foster fathers, are given in Table XVI.

TABLE XVI.—OCCUPATIONAL STATUS OF OWN FATHERS

Occupational Class	Own Fathers			Foster Fathers
	Legitimate	Illegitimate	Total	
	Percent	*Percent*	*Percent*	*Percent*
Professional	1	10	4	15
Semi-professional and business	2	16	6	41
Skilled labor	26	41	31	38
Semi-skilled to slightly skilled labor	34	28	32	5
Common labor	37	5	27	1
Total	100	100	100	100
Number of cases	177	78	255	394

The percentage of the own fathers engaged in the higher occupations is noticeably smaller than that for the foster fathers. Only 10 percent of the former were in the professional or business class, whereas the corresponding percentage for the latter is 54. This contrast is even more striking if the fathers of only the legitimate children are considered, of whom only 3 percent were in the two highest groups. Three fourths of the legitimate group were semi-skilled or common laborers. It is evident that the occupational level of the own fathers was decidedly below that of the foster fathers. Reference to Table XIII shows that it was also below that of the male population of Illinois.

Table XVI also shows that the fathers of the illegitimate children had a higher occupational status than those of the legitimate group. This superiority is again evident in the comparisons which follow and will be discussed more fully in a later section.

In 29 cases the mothers of the legitimate children were listed as having some occupation. Seventeen were domestics, cooks or waitresses; five were washerwomen; two did farm work; one was a rural-school teacher, one a factory worker, one a musician, one a cabaret singer, and one a store clerk.

There were 82 cases in which information was given concerning the occupations of the mothers of the illegitimate children. The occupations of this group seem superior to those of the legitimate group. The distribution is given herewith.

At home—do not need to work	12
Students or teachers	4
Office girls or stenographers	12
Nurses	3
Telephone operators	5
Home or hotel maids, waitresses	27
Factory workers	14
Farm workers	3
Actress	1
Prostitute	1
Total mentioned	82

The data on the education of the own parents are very meager. The material given in Table XVII includes only those cases for which definite information on the amount of schooling was available. Thus, in some instances the occupation, such as lawyer or physician, was mentioned, but no statement was made as to the years of training. Other cases which could not be included in the summary of schooling were those in which individuals were reported only as "illiterate" or with "very little education." While there was some tendency for the frequencies of these two types of cases to offset each other, it is still not certain that the group included in the table is a representative sample. The table as a whole shows the superiority of the illegitimate parents.

TABLE XVII.—SCHOOLING OF THE OWN PARENTS

Years of Schooling	Legitimate		Illegitimate	
	Fathers	Mothers	Fathers	Mothers
16 years or more	1	—	—	—
12 to 15 years	1	2	4	7
8 to 11 years	10	15	8	22
4 to 7 years	20	23	2	15
Less than 4 years	18	18	2	5
Total reporting	50	58	16	49
Approximate mean in years	5	5½	9	8

The most complete data on the own parents is that concerning their moral and mental characteristics. The extent of this information is shown in Table XVIII. Data on more than two thirds of the legitimate parents could be obtained from the family records. The information is, of course, much less complete in the case of the illegitimate parents. One reason for the incompleteness of the records was that, when a parent was reported as dead, it was seemingly thought unnecessary to record much further information.

TABLE XVIII.—MORAL AND MENTAL CONDITION OF THE OWN PARENTS WHERE SOME DEFINITE STATEMENT WAS MADE

Statement	Legitimate		Illegitimate	
	Father	Mother	Father	Mother
	Percent	*Percent*	*Percent*	*Percent*
Morally defective	63	32	27	28
Mentally defective	12	24	2	14
Both morally and mentally defective	14	27	2	11
Total defective	89	83	31	53
Morally sound	3	6	10	9
Mentally sound	7	10	44	27
Both morally and mentally sound	1	1	15	11
Total sound	11	17	69	47
Total	100	100	100	100
Number reporting	204	178	41	104
Percent of total reporting	80	68	29	74

The cases in which there was definite information as to the moral or mental condition of the parents are probably representative of the whole group. There is a possibility, however, that there was a tendency on the part of the investigators to be more careful to report defects than to mention the fact of normality.

The great extent of moral and mental defect on the part of the legitimate parents is shown by the fact that nearly 90 percent of the fathers and slightly over 80 percent of the mothers were described as "defective." It was often very difficult to make an accurate classification of these characteristics, both because of the meager family record and of the lack of definite moral standards. Subnormal mentality and insanity were considered as mental defects, while crime, abuse and neglect of family, excessive drunkenness, etc., were listed as moral defects.

Still greater difficulties were encountered in classifying the characteristics of the illegitimate parents. The very fact of being the parents of an illegitimate child might be taken as an indication of moral defect, but it was not so considered here unless there was some record of other offenses.

The chief interest in the present comparisons is in the extent of mental defect. It will be noted that 26 percent of the legitimate fathers and 51 percent of the legitimate mothers were mentioned as being mentally defective, according to the judgment of the Society investigators. The corresponding percentages for the illegitimate parents were 4 and 25, respectively. Except for the insubstantial data on the illegitimate fathers, the results as a whole show a very much larger proportion of mental defectives than are found in the general population.

Table XIX is presented to show the extent to which the children of the Home Group had defective parents. While there were

TABLE XIX.—EXTENT OF MORAL AND MENTAL DEFECT ON THE PART OF THE PARENTS OF THE HOME-GROUP CHILDREN (401)

	Mentally Defective	Morally Defective	Both Mentally and Morally Defective
One parent	120	145	85
Both parents	26	73	6

only 26 cases (6 percent) in which both parents were reported as mentally defective, there were 146 children (36 percent) who had at least one parent who was mentally defective. Both of these percentages would quite probably have been greater had complete data been available. As mentioned before, some records contain only the fact that the parent is dead and consequently make no remark concerning his mental or moral condition while living.

A more specific description of the moral and mental defects of the parents will next be given. Twelve of the legitimate children had fathers who were insane, 6 of them being in a state institution. In three other cases the fathers were epileptic. (Epilepsy was not considered as a mental defect, however.) In 57 cases (22 percent) the fathers had been implicated in some crime; 42 had been imprisoned in such institutions as Leavenworth federal prison, Joliet state prison, or county or city jails. Some of the offenses for which they were sentenced were attempted murder, conducting houses of ill fame, rape, adultery, cruelty to wife, larceny, forgery, and bootlegging. Other common moral defects were abuse of family, gambling, and excessive drinking. In 92 cases (35 percent) the father had deserted or refused to support his family. The fathers of 6 of the children had committed suicide.

As noted in Table XVIII, about half of the mothers of the legitimate children were considered mentally deficient. Twenty-one of these children (8 percent) had mothers who were insane, of whom 18 were in state institutions. Out of 84 mothers who were recorded as feeble-minded or of low mentality, 11 were in the Lincoln State School for the Feeble-Minded, 45 others were judged to be of very low mentality, and 6 others were rated as feeble-minded on the Binet Intelligence Test. In 22 cases (8 percent) the mothers had been involved in crimes or offenses leading to imprisonment or placement in correctional institutions. The most common offenses were murder, adultery, disorderly conduct, forgery, and excessive drinking. Many of the mothers were reported as having deserted or neglected their children. Other undesirable characteristics were profanity, shiftlessness, and filthy habits. In 8 cases the mother had committed suicide or been murdered.

Only one instance of possible feeble-mindedness is found among the fathers of the illegitimate children. Six of the fathers had been in a penitentiary or jail for such offenses as violating the Mann act, stealing, army desertion, and attempted murder. One of the fathers had committed suicide. As already noted, the data for this group were meager and possibly somewhat biased, since the cases for which no information was available may have been the worst cases.

Seven of the mothers of the illegitimate children were insane, and 26 of low mentality. Five were in institutions for delinquents, and 24 others were listed as being immoral.

4. Causes of Dependency

This description of the moral and mental condition of the legitimate parents indicates the chief reason for the dependency of the children. In the case of 30 of the 160 families the parents were declared unfit by a court and the children committed to the care of the Society. In 26 other families the immediate cause for dependency was the fact that one or both of the parents was in an asylum or a penal institution. Other reasons for dependency were the desertion of the mother or father (21 cases) and the death of one or both parents. This last reason was not as frequent as might be supposed, however, since in only 10 of the 160 families were both parents dead. In a considerable number of cases either the father or the mother was dead, but this fact alone seldom caused the dependency of the children, since it is evident that it was usually accompanied by some mental or moral unfitness on the part of the living parent.

In 104 out of the 141 illegitimate cases the baby was voluntarily committed to the Society by the mother as a protection to her reputation or because she felt she was unable to care for it properly. In 12 cases the mother had deserted the baby, leaving it with friends or in a boarding home. In 17 cases social or court agencies removed the child because of the mental or moral defect of the mother.

From the facts presented in this section, then, it is apparent that the heredity of the children of the Home Group, especially of the legitimate children, is considerably poorer than that of the general population. Eighty percent of the legitimate children came from families for which specific mention was made of parental defect. It is quite probable that such moral and mental defect existed in still other cases where no specific record was furnished. Probable evidence of defects is given in such reports as "no remark about father, mother incompetent;" "father inefficient and ignorant, mother dead;" "father worthless, mother in the poorhouse."

Though the data on the parents of the illegitimate children are relatively meager, there seems to be little doubt that low mentality and moral defects, other than those related to the fact of illegitimacy, were much less prevalent in this group.

The foster parents were distinctly superior to both of these groups with respect to mental and moral characteristics as well as occupational status and education. It is thus apparent that the group of foster children studied came from inferior homes and had a heredity which was decidedly poor. The homes in which these children were placed were considerably above the general average, as shown by the superiority of the foster parents. If the development of intelligence depends in any degree upon the character of the environment, such a decided change in surroundings should lead to a corresponding improvement in intelligence.

VI. The Home Group

1. The Entire Committed Group

This group consists of the 401 children of the Home Group and 83 other children who were tested in temporary homes. These 83 children had been committed to the Society for several years, but could not be included in the Home Group because they had not spent these years in continuous residence in some foster home. Although their new environment could not be rated, it was no doubt superior to that of their own homes. The entire Committed Group of 484 thus includes children removed from their own homes and placed in an improved environment.

The age distributions in Table XX show that the mean age of commitment for the group was 4 years, 5 months. The mean age of these children when tested in the present study was 11 years, 4 months, indicating an interval of nearly seven years in the new environment.

The distributions of intelligence for this Committed Group are given in Table XXI, separated according to legitimacy. It will be noted that the mean intelligence quotient for the entire group is 95.3, which is slightly below normal. The average for the illegitimate group is considerably higher than that for the legitimate group—a fact which is verified throughout the entire study.

TABLE XX.—AGE DISTRIBUTIONS OF THE ENTIRE COMMITTED GROUP (484)

Central Age	When Committed	When Tested	Central Age	When Committed	When Tested
22–6	—	3	10–6	16	46
21–6	—	2	9–6	20	16
20–6	—	3	8–6	26	35
19–6	—	1	7–6	32	28
18–6	—	11	6–6	31	26
17–6	1	14	5–6	36	20
16–6	—	24	4–6	38*	28
15–6	—	31	3–6	38	8
14–6	5	43	2–6	41	3
13–6	4	42	1–6	31	—
12–6	10	48	0–6	144	—
11–6	11	52*	Total	484	484
			Mean age	4 years, 5 months	11 years, 4 months

*These points indicate the approximate means of the distributions.

While the environment of the 83 children in this group who had not been in permanent homes is different from that of the Home Group children, they were nevertheless included in the study for the purpose of making such comparisons as that between the age of removal from their own home and their present intelligence. Some of the 83 were siblings of members of the Home Group and were tested in order to complete the record of the family.

2. The Home Group

The most important comparisons of this section are made with the Home Group of 401 children. The distributions given in Table XXII show the ages at which the children were committed (removed from their own homes), placed in their present foster homes, and tested in the study.

A striking difference is shown between the commitment ages of the legitimate and the illegitimate children. The mean age for

TABLE XXI.—DISTRIBUTIONS OF INTELLIGENCE QUOTIENT FOR THE ENTIRE COMMITTED GROUP (484)

Intelligence Quotient	Frequency			Percent		
	Legitimate	Illegitimate	Total	Legitimate	Illegitimate	Total
160–164	—	1	1	—	0.7	0.2
155–159	—	—	—	—	—	—
150–154	—	—	—	—	—	—
145–149	—	1	1	—	0.7	0.2
140–144	—	—	—	—	—	—
135–139	2	—	2	0.6	—	0.4
130–134	1	—	1	0.3	—	0.2
125–129	4	9	13	1.2	6.0	2.7
120–124	2	8	10	0.6	5.3	2.1
115–119	9	19	28	2.7	12.7	5.8
110–114	20	13	33	6.0	8.7	6.8
105–109	23	17	40	6.8	11.3	8.3
100–104	45	19	64	13.4	12.7	13.2
95– 99	32	23	55	9.6	15.3	11.4
90– 94	40	17	57	12.0	11.3	11.8
85– 89	38	7	45	11.4	4.7	9.3
80– 84	37	6	43	11.1	4.0	8.9
75– 79	32	4	36	9.6	2.7	7.4
70– 74	25	4	29	7.5	2.7	6.0
65– 69	10	1	11	3.0	0.7	2.3
60– 64	10	—	10	3.0	—	2.1
55– 59	4	1	5	1.2	0.7	1.0
Total	334	150	484	100.0	100.2	100.1
Mean I. Q.	91.5	103.7	95.3			
S. D.	15.3	15.5	16.3			

the legitimate group was 4 years, 9 months, while that for the illegitimate group was only 16 months. Although there is considerable range in the age of commitment for the group as a whole, it will be noted that over half of the children studied were committed before they were 3 years of age. In the case of the illegitimate children, practically 90 percent were committed before this age, while the corresponding percentage for the legitimate children was only 32. This difference in the commitment age of the two groups has a significant bearing on the results brought out in the study.

The fact that not all of the children were placed in their present foster homes immediately after their commitment is shown by a

TABLE XXII.—AGE DISTRIBUTIONS FOR THE HOME GROUP OF 401 CHILDREN

Central Age	When Committed			When Placed in Present Foster Home			When Tested		
	Legitimate	Illegitimate	Both	Legitimate	Illegitimate	Both	Legitimate	Illegitimate	Both
22–6	—	—	—	—	—	—	2	—	2
21–6	—	—	—	—	—	—	1	—	1
20–6	—	—	—	—	—	—	2	—	2
19–6	—	—	—	—	—	—	—	1	1
18–6	—	—	—	—	—	—	8	1	9
17–6	—	—	—	1	—	1	7	4	11
16–6	—	—	—	1	—	1	13	2	15
15–6	—	—	—	1	—	1	18	6	24
14–6	—	—	—	2	—	2	26	5	31
13–6	—	—	—	1	—	1	30	5	35
12–6	1	—	1	7	—	7	31*	12	43
11–6	4	—	4	9	1	10	36	10	46
10–6	9	1	10	14	3	17	23	12	35
9–6	11	1	12	15	—	15	11	2	13
8–6	14	1	15	18	2	20	17	10	27
7–6	23	1	24	12	1	13	13	14	27
6–6	22	1	23	25	2	27	8	17	25
5–6	32	1	33	25	3	28	7	12	19
4–6	30	5	35	36	4	40*	4	20	24
3–6	32	4	36	28	4	32	3	5	8
2–6	27	11	38	20	10	30	—	3	3
1–6	15	14	29	17	14*	31	—	—	—
0–6	40	101	141	28	97	125	—	—	—
Total	260	141	401	260	141	401	260	141	401
Mean Age	4–9	1–4	3–6	5–8	1–7	4–2	12–2	9–0	11–0
S.D.	3–0	1–9	3–1	3–7	2–3	3–10	3–7	4–0	4–0

*These points indicate the approximate means of the distributions.

comparison of the ages of commitment with those of placement. The intervening time (averaging 8 months) was spent in the Society Receiving Home, in a temporary boarding home, or in a foster home that did not prove to be permanent.

The mean age at the time of the test in the present study was 11 years for the entire group. The most important fact to be noted in this connection is the difference of over three years between the legitimate and the illegitimate groups. Another significant result is the great variation in test age for the entire group—the range is from 2 to 22 years, with a standard deviation of 4 years.

The variations in age of commitment and age of placement made possible a study of the effect of such factors upon the present intelligence of the foster child. The great variability in age at the time of the test, however, proved to be a complicating factor. In some of the comparisons, therefore, only certain portions of the group were used.

Table XXIII gives the distributions of intelligence quotients for the legitimate and the illegitimate children of the Home Group. The mean for the entire group is 97.5, which is slightly nearer the

TABLE XXIII.—DISTRIBUTION OF INTELLIGENCE OF THE HOME GROUP

Intelligence Quotient	Frequency			Percentage		
	Legitimate	Illegitimate	Total	Legitimate	Illegitimate	Total
160–164	—	1	1	—	0.7	0.2
155–159	—	—	—	—	—	—
150–154	—	—	—	—	—	—
145–149	—	1	1	—	0.7	0.2
140–144	—	—	—	—	—	—
135–139	2	—	2	0.8	—	0.5
130–134	1	—	1	0.4	—	0.2
125–129	4	8	12	1.5	5.7	3.0
120–124	2	8	10	0.8	5.7	2.5
115–119	8	19	27	3.1	13.5	6.7
110–114	19	13	32	7.3	9.2	8.0
105–109	24	15	39	8.8	10.6	9.5
100–104	37	19	56	14.2	13.5	14.0
95– 99	26	21	47	10.0	14.9	11.7
90– 94	33	15	48	12.7	10.6	12.0
85– 89	29	7	36	11.5	5.0	9.2
80– 84	25	5	30	9.6	3.6	7.5
75– 79	21	4	25	8.1	2.8	6.2
70– 74	16	3	19	6.2	2.1	4.7
65– 69	6	1	7	2.3	0.7	1.8
60– 64	6	—	6	2.3	—	1.5
55– 59	1	1	2	0.4	0.7	0.5
Total	260	141	401	100.0	100.0	99.9
Mean I. Q.	93.9 ± .6	104.1 ± .9	97.5 ± .6			
S. D.	15.0	15.4	15.9			

normal value than that found for the entire Committed Group. It will be noted that the variabliity of the legitimate and the illegitimate groups is practically the same.

The marked superiority of the illegitimate children is shown by their mean intelligence quotient, 104.1, as compared with that of 93.9 for the legitimate group. This difference of about 10 points, however, is reduced to 6.4 points when corrected for age. Such a difference is not surprising in view of the fact that the illegitimate children had a somewhat superior parentage and certain marked environmental advantages, as shown later.

3. Comparison of Home Group with Certain Other Groups

In order to compare the intelligence of this group of foster children with an unselected group of normal children it was decided to use the distribution given by Terman in the standardization of the Stanford Revision of the Binet Scale.[5] In as much as the Terman distribution included only children whose ages ranged from 4 years, 10 months, to 14 years, 2 months, these same limits were used in the distribution shown in Table XXIV below. It was

TABLE XXIV.—COMPARISON OF THE DISTRIBUTION OF INTELLIGENCE QUOTIENTS OF THE HOME GROUP WITH THAT OF TERMAN'S UNSELECTED GROUP OF 905

Intelligence Quotient	Home Group	Two-Year Group	Terman Unselected Group
	Percent	*Percent*	*Percent*
136–145	1.1	0.9	0.55
126–135	2.4	3.5	2.3
116–125	9.8	14.8	9.0
106–115	15.8	20.0	23.1
96–105	25.6	27.8	33.9
86– 95	25.3	17.4	20.1
76– 85	14.0	8.7	8.6
66– 75	4.9	6.1	2.3
56– 65	1.1	0.9	0.33
Total	100.0	100.1	100.18
Number of cases	285	115	905
Mean I. Q.	98.5 ± .6	101.7 ± 1.0	101.5 ± .3
S. D.	15.2 ± .4	15.7 ± 0.7	13.0 ± .2

[5] Lewis M. Terman *et al. The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence.* Educational Psychology Monographs. No. 18. Warwick and York, Inc. 1917. Page 40.

also necessary to alter the class intervals and reduce the frequencies to percentages in order to make the data comparable with those of Professor Terman.

It will be noted that the mean intelligence quotient for the Home Group is 98.5, while that for the Terman group is 101.5. Since the mean age of the latter group was 9-6 and that of the Home Group was 11-2, the observed difference of three points would be reduced to only one point by the corrections from Table XXXVIII. This difference is evidently insignificant.

It will also be noted that the standard deviation for the Home Group is 15.2, as compared with 13.0 for the Terman group. The greater variability of the Home Group is also indicated by the fact that about 6 percent of its members have intelligence quotients of 75 or below, whereas the corresponding percentage for the Terman data is only 2.6. At the upper end of the scale it is found that the Home Group contains 3.5 percent with an intelligence quotient

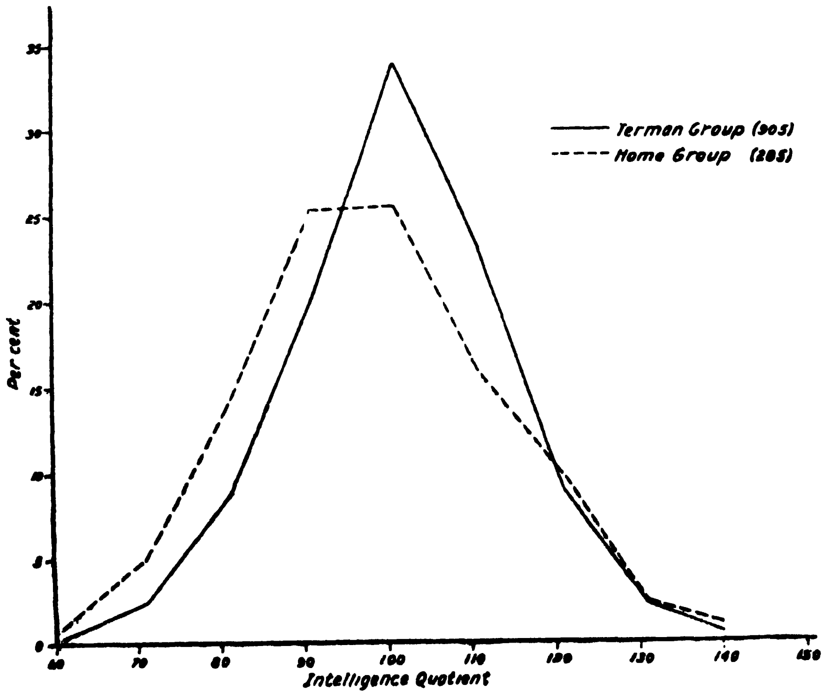


FIG. 1. COMPARISON OF HOME GROUP WITH TERMAN UNSELECTED GROUP

above 135, while the Terman group contains 2.8 percent in this range.

The general agreement of the two groups is shown more clearly in Figure 1. Both curves appear to be fairly symmetrical, but the distribution for the Home Group is platykurtic.

A second comparison with the Terman Unselected Group is furnished by the children of the Home Group who were placed in their foster homes before they were two years of age. The means of the two groups are practically the same, but the variability of the two-year group is somewhat greater than that of the Terman Group. The general resemblance of the two distributions is shown in

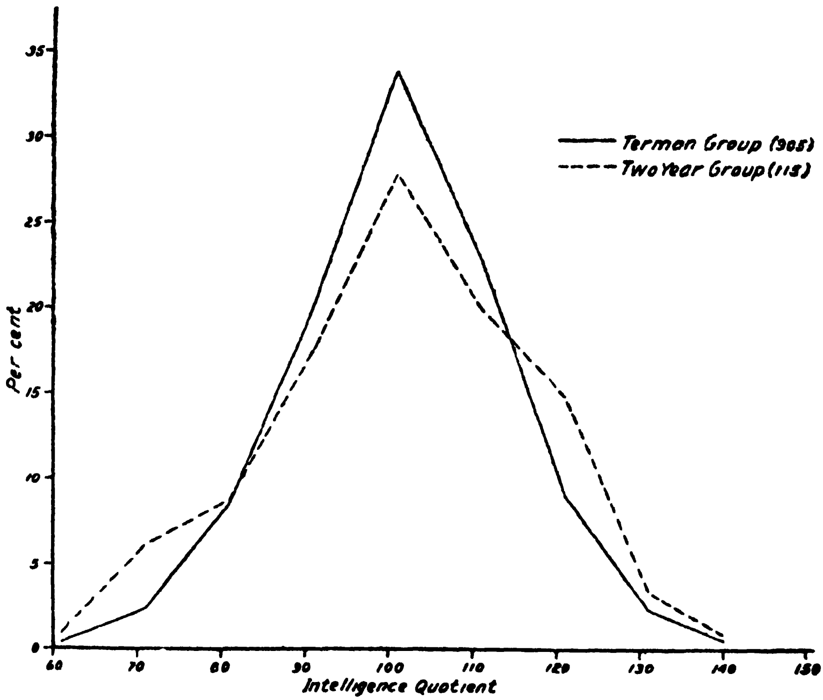


FIG. 2. COMPARISON OF TWO-YEAR GROUP WITH TERMAN UNSELECTED GROUP

Figure 2 (see also Table XXIV). It is apparent that this group of children who have been in foster homes practically all their lives compare more favorably with a group of unselected children than do all the children of the entire Home Group.

These comparisons show that, on the whole, the Home Group is substantially a normal one. To interpret this fact it would be necessary to know whether the degree of intelligence shown is inherent or is due to the improved character of the environment. It seems hardly possible that the children of such inferior parents could form a group so normal in intelligence if heredity is the only factor. If a measure of the intelligence of the children at commitment had been available, the question could be answered directly. Some light may be thrown upon it by considering groups of dependent children who had not yet been placed in foster homes.

Studies made by Stenquist and others,[6] Cobb,[7] and Gordon[8] show that the intelligence of dependent children is usually markedly below normal. This observed inferiority may be due in part, however, to the fact that the groups studied were orphanage or county home children. There may be some tendency for such institutions to retain more of the less intelligent dependent children, thus lowering the general average.

A group of newly committed dependent children which furnish a more direct comparison consists of 247 wards of the Society who were tested by the Institute for Juvenile Research before their placement in foster homes. The mean intelligence quotient for this group was found to be 84.0, the standard deviation 15.9. This average is significantly lower than that of 98 found for the Home Group of the present study. Moreover, the observed difference of 14 points is increased to 17 when the means are corrected for age. The two groups, however, are not entirely comparable. Since this test group of 247 is only a small part of the total number of children committed to the Society and includes only those who were tested, it may not be entirely representative of the entire group of children received during the period. As already explained, tests were given for various reasons, one of these being a suspicion of low mentality. This may have had a tendency to lower the average

[6] Stenquist, Thorndike, and Trabue. "The intellectual status of children who are public charges." *Archives of Psychology,* 5: Sept., 1915, No. 33.

[7] Margaret E. Cobb. "The mentality of dependent children." *Jour. of Delinquency.* May, 1922, 132-140.

[8] Kate Gordon. "Report of psychological tests of orphanage children." *Jour. of Delinquency,* 4: Jan., 1919, 46-55.

for the group. Since this newly committed group contained practically no illegitimate children, the mean of 84 is more comparable with that of 94 found for the legitimate children of the Home Group. In spite of these limitations, however, it appears probable that there is a significant difference between the intelligence of dependent children before and after residence in a foster home.

In this next comparison the difficulty just mentioned in the selection of the children for testing has been eliminated. The group of 137 newly committed legitimate children who were tested in the present study are representative of the type of dependent children now being received by the Society. The children so tested were selected only on the basis of their convenient location in temporary receiving or boarding homes. The group consisted of those who had become wards of the Society during the period of the study and had not yet been placed in foster homes. Since there has been a growing tendency for the Society to accept fewer children of obviously low mentality, the group as a whole is quite probably superior to children accepted in previous years when the children of the Home Group were committed. Table XXV gives

TABLE XXV.—COMPARISON OF NEWLY COMMITTED DEPENDENT CHILDREN WITH THOSE WHO HAVE BEEN RAISED IN FOSTER HOMES

Intelligence Quotient	Newly Committed Group	Legitimate Children of the Home Group	Legitimate Children of Home Group Committed Before Five Years of Age
	Percent	*Percent*	*Percent*
130–139	....	1.2	1.4
120–129	1.5	2.3	4.2
110–119	7.3	10.4	12.5
100–109	10.2	23.1	25.0
90– 99	29.2	22.7	23.6
80– 89	23.3	21.1	19.4
70– 79	16.8	14.2	11.1
60– 69	10.2	4.6	2.8
50– 59	1.5	0.4	....
Total	100.0	100.0	100.0
Number of cases	137	260	144
Mean I. Q.	88.6 ± .8	94.1 ± .6	96.8 ± .8
S. D.	14.7	15.2	14.8
Mean age at test	9–3	12–2	10–4

the distributions of intelligence for this Newly Committed Group as compared with the legitimate children of the Home Group. A separate distribution is given for those of the Home Group who were committed before they were 5 years of age.

The mean intelligence quotient for the Newly Committed Group is 88.6 as compared with 94.1 and 96.8 for the entire and the restricted Home Group. When the usual tests are made it is found that the mean for the new children is significantly lower than that of either of the other two groups. The most striking difference (8.2 ± 1.1) is that between the newly committed children and those who were once similarly dependent but had been placed in foster homes before they were 5 years of age. This difference is increased to 9.5 points when the usual correction for age is made. The difference of 5.5 ± 1.0 between the means for the Newly Committed Group and all the legitimate children of the Home Group becomes 9.0 when corrected for age difference.

If it could be assumed that the Newly Committed Group and the Home Group had the same type of heredity and early environment, the observed difference in intelligence would have to be ascribed to the improved surroundings of the members of the Home Group. As already noted, there may be some possibility that the obviously dull children have been eliminated from both groups. Since there was more evidence of such elimination in the case of the Newly Committed Group, the actual difference in the intelligence of the two groups is probably even greater than that observed.

Further light on this comparison is given by the results of the International Test. In order to make the groups comparable, it was necessary to pair the cases for age. Table XXVI gives the age and score distributions for the 41 pairs. The mean score for the Newly Committed Group was 108.4, as compared with 122.1 for the Home Group. The difference, 13.7 ± 6.6, is only twice its probable error, but the latter has been distorted because of the great variability in the scores. With a more normal distribution of scores the observed difference would probably have been significant.

4. Selection of the Legitimate Children of the Home Group

It has already been shown that dependent children who have been in foster homes for a number of years (mean time 6 years, 10 months), have an average intelligence quotient which is practically normal. This fact appears rather striking in view of the low mental level of the own parents, described in the preceding section. Since some of these parents had children other than those included in the present study, it will be necessary to show that there was no

TABLE XXVI.—INTERNATIONAL TEST RESULTS FOR THE NEWLY COMMITTED GROUP AND THE HOME GROUP

Score and Age in Months	Age		Score	
	New Group	Home Group	New Group	Home Group
210–219	—	—	—	2
200–209	1	1	—	—
190–199	—	—	1	2
180–189	1	1	—	2
170–179	1	1	1	2
160–169	2	2	1	2
150–159	2	2	3	3
140–149	3	3	4	2
130–139	8	8	3	2
120–129	4	5	5	5
110–119	5	4	4	2
100–109	4	4	1	4
90–99	4	4	4	2
80–89	4	3	3	—
70–79	1	3	2	4
60–69	1	—	4	2
50–59	—	—	3	—
40–49	—	—	—	4
30–39	—	—	1	1
20–29	—	—	1	—
Total	41	41	41	41
Mean Age	123.8 months	124.0 months		
Mean Score			108.4±4.2	122.1±5.2
S. D.			39.4	49.2

selection of the brighter children of the family either for commitment or for subsequent placement in a foster home.

The 260 legitimate children of the Home Group came from 160 families. In 75 of these families all of the living children were committed to the Society. In 13 families the records were too incomplete to permit of a conclusive statement. In the remaining 72 families there were known to be other children besides those committed to the Society. These families had in all 366 children, 193 of whom were taken by the Society. The following distribution gives an idea of the disposition or the status of the children who were not so committed:

51 families—the *oldest* children—usually grown, self-supporting, or married
6 families—the *youngest*—kept by the family or friends
6 families—children of varying ages kept by family, basis of selection not stated
4 families—children committed to other placement societies
3 families—children placed with relatives or residents of the town
2 families—children placed in some orphanage

72 families

It is clearly evident that the selection of the children of a family for commitment was not on the basis of their intelligence. In the great majority of cases only the younger children were taken; the older ones had been previously placed in other homes or had reached maturity. In the 57 cases in which the commitment was on the basis of age it is clear that age, not intelligence, was the basis of selection. In the cases in which children were taken by relatives or neighbors of the family it is very improbable that the duller children would have been so chosen and the brighter ones left for commitment to the Society.

Further evidence of the lack of any selection on the basis of intelligence is furnished by the mean intelligence quotients found for two groups of children. One of these groups consisted of the 120 children of the Home Group who came from the 72 families where not all the children were committed to the Society. The second group consisted of the 123 children of the 75 families where all the children were committed to the Society. In both cases the children tested in the study were about two-thirds of the total number committed. The mean intelligence quotient of the first group was 93.7, while that of the second was 92.8, showing that there was no significant difference in intelligence. Had the children of the 72 families been the selected, brighter portion of the children, it would seem that they should average higher in intelligence than a group containing all the children of certain other families. It is important to note that the heredity of these two groups is quite similar, since the proportion of parents mentioned as mentally defective was the same for both groups (24 percent). Hence, it seems fair to assume that the children of these 72 families who were committed to the Society were not selected on the basis of their superior mentality.

The next possibility to be considered is that the Home Group of the present study consists of only the brighter portion of the committed children. The entire number of children committed from the 160 families mentioned above was 404. Of this number, 278 were tested in the study and of these 260 were considered in the Home Group; the remaining 126 were not tested, for the reasons given below.

Reason for not testing	*Number of cases*
Too distant or isolated	45
Not settled in permanent home	15
Test refused by foster parents	14
Had been returned to parents or relatives	14
Died since placement	10
Cannot be located	7
Sent to feeble-minded school	7
Too old, location uncertain	7
Too old, married, ran away	4
In the Navy	1
In school for delinquent boys	1
In school for delinquent girls	1
Total	126

It will be observed that the most frequent reasons for not testing were inaccessibility, lack of permanent foster home residence, and refusal of test. In only 7 cases was there specific mention of feeble-mindedness. As these cases constitute less than 2 percent of the total number of children committed, the effect of selection on the Home Group is slight. The percentage is even smaller than the proportion of the Home Group who have intelligence quotients below 70. In the other cases, those which were inaccessible, refused, or returned to parents or relatives at their request, there is no reason for inferring inferiority of mentality that would have lowered the average of the Home Group had the cases been included.

Some further evidence of the mentality of this group not included in the study is furnished by tests given to 39 of the children by the Institute for Juvenile Research. These tests were given to children, who, upon their commitment, were suspected of being of low mentality. In 11 cases the intelligence quotient was found to be below 70; five of these children were sent immediately to an institution for the feeble-minded. The total percentage of children found to be subnormal by a test at the time of commitment was thus less than 3 percent, which is slightly less than that found in the study for the legitimate children of the Home Group. In the study there was no intentional exclusion of cases in which a previous test had indicated low mentality.

It is further worthy of note that of the five children sent to a school for feeble-minded and the eight others who tested below 70, the average commitment age was 10 years, 3 months, as compared with that of 4 years, 9 months, for the legitimate Home Group. In 9 of these 13 cases the child found subnormal was the oldest of his family. It thus appears that the defective children were the older ones who had been longer in a poor environment.

Of the 278 legitimate children tested in the study 18 were not included in the Home Group for the reasons given below.

In Society Receiving home, returned from foster home.....	8
In present home too short a time..........................	3
In a delinquent institution................................	2
In Society school for older girls.........................	2
Working as a maid in a home...............................	2
In a crippled children's school...........................	1
Total	18

It is evident that the chief reason for omitting these children from the Home Group was their lack of residence in a foster home which could be rated. They were tested in order to complete the data on the members of a family, some of whom were in permanent foster homes. That the group is rather below average mentally is shown by their mean intelligence quotient, 85. Whether this low average is due to their unsettled life in various homes and institutions or whether their failure to find a permanent home is due to their low mentality cannot be definitely determined, although the latter seems more probable. Several of the group presented behavior problems. It is significant that the mean age of commitment for these 18 cases was 8 years, 8 months, which is considerably older than the Committed Group as a whole.

In view of all these facts, it is evident that there has been no marked selection of the children included in the Home Group. This conclusion is supported by evidence from two groups of children. One of these groups consisted of the children from the 75 families where all the children were committed to the Society. The mean intelligence quotient for the 132 children tested out of the 193 who were committed was 91.7. The second group consisted of 72 children from 42 families. The mean intelligence quotient based on tests of all these children was 92.8. Since the average is practically the same whether only a part or all of the children of the families are tested, it seems improbable that the children who were tested in the study represent any selection on the basis of intelligence. The percentage of mentally defective parents in these two groups is also substantially the same, 24 and 21, respectively.

5. Intelligence of Foster Children in Relation to Their Parentage

As already pointed out, the data on the own parents is relatively meager. The most complete information is that concerning the extent of moral and mental defect. The intelligence of the children whose parents are recorded as being mentally or morally defective is shown in Table XXVII. In order to show the relatively greater influence of a good environment when the children are removed from their parents at an early age, the group has been divided into two sub-groups which include those removed before and after five years of age.

The mean intelligence quotient of the 120 children who had at least one mentally defective parent was 92.9. The most significant fact to be noted, however, is that the children who remained with their own parents until they were at least 5 years of age had a mean intelligence quotient of 87.2, as compared with an average of 95.1 for those who were removed before that age. Of the entire 120 children, only 4 (3.3 percent) had an intelligence rating below 70.

The second group in the table consists of 26 cases in which both parents were definitely reported as being mentally defective. The mean intelligence quotient for these children is 81.2, which is considerably lower than that found when only one parent was defective. The contrast between early and late removal is again shown by the means of 86.0 and 78.1, respectively. The average age of commitment for these children was 5 years, 10 months, while that for the entire Home Group was only 3 years, 6 months. Moreover, the homes in which they were placed averaged only 15, indicating that they were much poorer than those of the group as whole,

TABLE XXVII.—INTELLIGENCE OF THE CHILDREN OF DEFECTIVE PARENTS

Mention of Defect on Part of Child's Parents	Number with an Intelligence Quotient Below 70	Percentage with Intelligence Quotient Below 70	Mean Intelligence Quotient
Only one parent mentally defective—			
Removed before 5 yrs. (86)	2	2	95.1
Removed after 5 yrs. (34) ..	2	6	87.2
All cases (120)..........	4	3	92.9
Both parents mentally defective—			
Removed before 5 yrs. (10)	—	—	85.0
Removed after 5 yrs. (16) ..	4	25	78.1
All cases (26)...........	4	15	81.2
Only one parent morally defective—			
Removed before 5 yrs. (87)	1	1	98.3
Removed after 5 yrs. (58) ..	6	10	87.9
All cases (145)...........	7	5	94.1
Both parents morally defective—			
Removed before 5 yrs. (44)	1	2	97.3
Removed after 5 yrs. (29) ..	1	3	88.7
All cases (73)...........	2	3	93.9
One parent morally and mentally defective—			
Removed before 5 yrs. (53)	2	4	96.3
Removed after 5 yrs. (32) ..	3	9	85.3
All cases (85)...........	5	6	92.2
Both parents morally and mentally defective—			
Removed before 5 yrs. (3).	—	—	—
Removed after 5 yrs. (3) ...	—	—	—
All cases (6)............	—	—	—

for which the average was 18.8. It is thus evident that these children were placed in quite poor homes and at a relatively late age. Under more favorable conditions, they might be expected to make an even better showing.

It is significant that only four of these children, both of whose parents were reported to be feeble-minded, were found to be subnormal. These four were legitimate children who had remained with their own parents beyond the age of six. Their intelligence quotients, ages, and home ratings are given below.

Child	Intelligence Quotient	Age When Committed to Society	Age When Tested	Home Rating
A	69	8-9	16-7	13
B*	69	8-10	18-1	9
C*	63	6-9	15-11	9
D	66	8-10	13-1	12

*B and C are siblings placed in the same foster home.

If feeble-mindedness is to be regarded as a recessive character, the offspring of two feeble-minded parents would *all* be feeble-minded, according to the Mendelian law. In the above group, however, only 4 of the 26 children were found to be subnormal and two of these were members of the same family. It should also be noted that these four children were over 6.5 years when committed, were relatively old when tested, had been in very poor foster homes, and tested only slightly below 70. Owing to the nature of the test, it is extremely likely that they would have rated above the level of feeble-mindedness if tested at an earlier age. Although the ratings of the intellectual level of the own parents are, in the majority of cases, only the estimates of the Society investigators, it is probable that these parents in question are correctly classified as feeble-minded. The findings on this group, therefore, appear to indicate that feeble-mindedness is not to be regarded as a unit character which is inherited in accordance with the Mendelian law. On the contrary, it is a trait which is subject to the modifying effect of environment.

In connection with these four children, it is interesting to note that Child A has two younger brothers whose intelligence quotients are now 95 and 98. They were aged 3-6 and 0-9 when committed and were placed in homes rating 26 and 17, respectively. B and C have a brother who was 4 years of age at the time of commitment and was placed in a home rating 20. His intelligence quotient is now 87. A brother to Child D was placed at 13 months of age in the same foster home and has an intelligence quotient of 81.

In the remaining comparisons in Table XXVII it is shown that in every case the earlier separation of the children from their own parents resulted in a higher mean intelligence quotient.

It is of further interest to compare the present intelligence of foster children with that of their older siblings who were not committed to the Society and had been all their lives in a poor environment. The data on these uncommitted children is very scant, but a few comparisons are possible. As before noted, there were 72 families from which only a part of the children were committed to the Society. In the case of 14 of these older children who were not committed there is mention of some mental defect. In Table XXVIII the intelligence of the foster home children has been com-

TABLE XXVIII.—COMPARISON OF FOSTER HOME CHILDREN WITH THEIR SIBLINGS RAISED WITH OWN PARENTS

Children Not Committed		Their Siblings Who Were Committed	
Mental Rating	Age When Siblings Were Committed	Present Intelligence Quotient	Age at Time of Commitment
"Subnormal"	*years* 22	75 89 114	13–11 9–11 6–6
"Considered subnormal"	19	91	7–11
"In subnormal school" "Retarded mentally"	 Oldest Next to oldest	104 88 111	5–5 3–10 0–11
"Mentality doubtful"	9	81 109	7–3 3–1
"Borderline" by test	7	93	1–10
"Not bright. In third grade at 17 years of age"	12	81 69 95 98	10–6 8–0 3–6 0–9
"Considered feeble-minded"	14	85	8–5
In Lincoln State School for the Feeble-Minded	15	85 95	10–2 7–8
Same	9	98	6–7
Same	8	82	3–1
Same, I. Q. of 63 Same, I. Q. of 62 Same, I. Q. of 72	16 15 13	68 86	10–1 7–7 ....

pared with that of their non-committed siblings, by families. In the case of the latter the intelligence is given as a verbal rating or an intelligence quotient determined before or at the time the case was being investigated. The intelligence quotients for the foster-home children are from tests given after several years of residence in a foster home.

It is shown in Table XXVIII that in every case the non-committed children are older than their siblings placed in foster homes. The marked inferiority of these older children is clearly indicated. Since there is no reason to believe that in general the older children of a family are the duller, this difference must be due to the superior environment of the younger children. For conclusive proof of this fact it would be necessary to know the intelligence of the adopted children at their commitment. Lacking this information, however, it seems fair to assume that the committed and non-committed children were of practically the same initial level of intelligence.

Further information concerning the intelligence of the children of defective parents is shown by a comparison of mothers and children. In the case of 35 children of the Home Group there was definite information concerning the intelligence of the mothers. In the case of 20 of these children the intelligence quotient of the mother was available. Table XXIX shows the contrast between the intelligence of these 13 mothers and that of their 20 children. Any significant information known concerning the fathers is included in the table. The mean intelligence quotient of these mothers was 68.1, as compared with 94.9 for the children. When only the defective mothers were considered, their mean intelligence quotient was found to be 56.4, while that of their children was 89.1. A difference of 11.7 in the mean I. Q.'s of the mothers was accompanied by a difference of only 5.8 in the mean I. Q.'s of the children. It is noteworthy that not one of the 11 children of definitely defective mothers had an intelligence quotient below 75.

In addition to the 13 mothers mentioned above, there were 10 others for whom the mental level was definitely known. Nine of these were in state institutions where they were rated as feeble-minded; one was rated as subnormal according to a test, but the intelligence quotient was not stated in the record. Table XXX gives the intelligence quotients of the 15 children of these mothers.

TABLE XXIX.—COMPARISON OF THE INTELLIGENCE OF CHILDREN RAISED IN FOSTER HOMES WITH THAT OF THEIR OWN MOTHERS

Intelligence Quotient of Mother	What is Known of Father	Child is Legitimate or Illegitimate	Child's age at Commitment	Child's Present Intelligence Quotient	Child's Previous Intelligence Quotient
99	Deserted. In Navy......	I	0– 0	119	....
85	Machinist..............	I I	0– 1 0– 1 (twins)	99 104	....
80	Odd jobber. Arrested for violating Mann Act...	I	0– 4	101	....
78	Deserted............... Deserted...............	I I	3– 7 1–10	110 112	 91 (3 years ago)
77	Sailor. Deserted........	I I	2–10 1– 0	99 92	92 (2 years ago)
71*	Express checker.........	I	0– 1	82	70 (2½ years ago)
65		L	0– 3	84	....
63	Common laborer........	L	5– 9	105	106 (2 years ago)
60	Intelligence quotient of 55, day laborer, uneducated................	L	7– 1	76	85 (7 years ago)
57	Drunkard, neglected his family..............	L L L	4–10 3– 0 1– 5	78 87 75	....
54	Two years of high school, moral degenerate, imprisoned for forgery and cruelty to wife....	L L	7– 3 3– 1	81 109	81 (2 years ago) ....
50	Dead..................	L L	5– 8 4– 8	93 90	... .
46	Cuts timber and buys junk, in jail for drunkenness, has a feeble-minded brother.......	L	2–10	102	....
N = 13 Mean I. Q. 68.1			N = 20 Mean I. Q. 94.9		

*This line, drawn to separate the defective mothers, significantly separates the legitimate from the illegitimate children.

TABLE XXX.—INTELLIGENCE OF CHILDREN OF MENTALLY DEFECTIVE MOTHERS

Mother	Father	Child is Legitimate or Illegitimate	Child's Age at Commitment	Child's Present Intelligence Quotient	Child's Previous Intelligence Quotient
In state feeble-minded colony	More intelligent than wife, restless, poor manager	L	9–7	116	110 (6 years ago)
In colony		I	2–5	99	
In colony	Imprisoned for non-support of family	L	4–6	93	
In colony	Low mentality, never completed first grade in school	L L	8–10 1–1	66 81	74 (4 years ago)
In colony		I	0–3	70	54 (3 years ago)
In state hospital for insane, "imbecile"	Unfit parent, in county poor house	L	0–6	98	
In state hospital for insane, "feeble-minded"		I	0–3	67	51 (2 years ago)
In state hospital for insane, "feeble-minded"	Miner, ignorant, simple-minded	L L L	8–0 3–6 0–9	69 95 98	 96 (2½ years ago) 86 (4 years ago)
Rated as a moron in state insane hospital	All right mentally, a drunkard	L L L	14–9 11–5 2–6	79 101 100	 93 (6 years ago)
"Subnormal" according to a test	Chauffeur, immoral	L	2–2	92	
N = 10			N = 15 Mean I. Q. = 88.3		

The mean is 88.3 and only three of the children rate below 70 (69, 67, and 66). These facts corroborate those found for the mothers for whom definite intelligence quotients were known.

Although there is not complete information on the mental level of the fathers of these children, there is specific mention of low mentality in three cases involving six children. The available data would seem to indicate that, as a group, the fathers were below normal in intelligence. It is possible, according to the Mendelian law, for a mentally defective mother to have normal children (meaning "not feeble-minded") in case the father is normal. The indications of low mentality on the part of these fathers, however, are sufficient to make it probable that a considerable number of the children would be of subnormal mentality if heredity were the only factor. As a matter of fact, only three of the children are below 70 in intelligence, while the average of the 26 children with defective mothers is 89.

Slight evidence of the influence of environment in producing this level of intelligence is afforded by the 13 cases in which a pre-test was available; the average gain in intelligence was 6 points. If corrected for age, the gain is increased to over 10 points. Some of the pre-tests were given after a period of residence in the foster home.

6. The Influence of Environmental Factors

It has been shown that the distribution of intelligence for the entire Home Group is very similar to that found for normal unselected children. In view of the inferior heredity of these children and the relatively low intelligence of some of their non-committed siblings, it is extremely probable that there was a distinct improvement in intelligence after their placement in foster homes. If improvement in environment tends to produce a gain in intelligence, there should be a relation between the type of this environment and the present intellectual level of the foster child. In the present section a number of such relationships will be shown by group means and correlations.

In order to interpret the pronounced relationships which were observed, it was necessary to consider the possibility that the more intelligent parents, living in better homes, might have selected brighter children for adoption. This type of selection has been discussed briefly in the third section and will be taken up again at the end of this section.

The relation between the child's intelligence and his foster home rating is shown in Table XXXI. As already noted, the home rating index is made up of ratings on material environment, evidences of culture, occupational standing, education, and social activity. The entire range of this index was from 6 to 30 with a mean of 18.8 and an S. D. of 5.6. In the table this range has been divided into three parts indicating good, average, and poor homes. The corresponding level of intelligence of the foster children in these homes is shown by their mean intelligence quotients.

TABLE XXXI.—INTELLIGENCE OF FOSTER CHILDREN ACCORDING TO GRADE OF HOME
(Entire Home Group of 401)

Type of Home	N	Mean Intelligence Quotient of Foster Child
Good (23 to 30)	114	106.8
Average (15 to 22)	186	96.4
Poor (6 to 14)	101	88.9
Total	401	
Mean, 18.8	....	97.5
S. D., 5.6	....	15.9

The striking relationship noted above is also shown by the correlation of .48 ± .03 found between the child's intelligence and his foster home rating. When the group is divided into legitimate and illegitimate groups, the observed correlations are .43 ± .08 and .47 ± .04, respectively. When the group is restricted to white children five to fourteen years of age (273 cases), the correlation is found to be .49 ± .03.

If the Home Group is limited to those who have been in foster homes four years, the number is reduced to 304. For these cases the correlation with home rating is .47 ± .03 for the entire group, and .43 ± .04 and .45 ± .05 for the legitimate and illegitimate subgroups, respectively.

In Table XXXII the Home Group has been limited to children who had entered their foster homes before they were 5 years of age and had been in those homes at least four years.

This table shows the same general result as the preceding one. It will be noted that there is a difference of 12.5 points between the mean intelligence quotients of the legitimate children in good and poor homes. An approximately equal range is found for the illegitimate children, but at a level about 9 points higher on the scale. The mean intelligence quotient for the combined groups is 99.6, which is slightly higher than that of the entire Home Group. This difference (2.1 ± 1.1) is probably due to the larger proportion of illegitimate children and the earlier placement age in this limited group.

TABLE XXXII.—INTELLIGENCE OF FOSTER CHILDREN ACCORDING TO GRADE OF HOME

(Group limited to those entering homes before five years of age and having spent at least four years in the home)

Grade of Home	Legitimate		Illegitimate	
	N	Mean Intelligence Quotient	N	Mean Intelligence Quotient
Good (22 to 30)	36	101.3	48	109.4
Average (17 to 21)	44	94.5	35	104.4
Poor (7 to 16)	34	88.8	29	96.1
Total	114		112	
Mean, 19.7		95.0		104.4
S. D., 5.4		14.2		14.2

The relation between intelligence and home rating for this group is supported by correlations of $.41 \pm .04$ for the entire group and $.35 \pm .06$ and $.45 \pm .05$ for the legitimate and illegitimate groups, considered separately.

A still more striking difference between the levels of intelligence of children in the three types of homes is shown in Table XXXIII. This group consists only of those who entered their foster homes before they were two years of age. For this entire group the children in the poorer homes had a mean intelligence quotient of 91.3, while those in the good homes had an average of 111.3. This difference of 20 points is considerably greater than those found in the preceding comparisons and is probably due to the greater influence of the home environment at the earlier age. The correlation between home rating and foster child's intelligence for this group was .52 ± .04.

This group contained 111 illegitimate and 45 legitimate children. The difference in the intelligence of the children in the poor and good homes is similar for both groups. It appears that the illegitimate are quite superior in each of the three grades of homes. This superiority is probably due in part to the better heredity of

TABLE XXXIII.—INTELLIGENCE OF TWO-YEAR GROUP ACCORDING TO GRADE OF FOSTER HOME

Type of Home	Entire Group		Legitimate		Illegitimate	
	N	Mean I.Q.	N	Mean I.Q.	N	Mean I.Q.
Good (23 to 30)	55	111.3	10	106.5	45	112.4
Average (16 to 22)	58	102.8	19	98.3	39	104.9
Poor (8 to 15)	43	91.3	16	83.4	27	95.6
Total	156		45		111	
Mean Intelligence Quotient		102.6		94.8		105.7
S. D.		16.0		15.8		15.1
Mean Home Rating		19.6		17.6		20.4
S. D.		5.7		5.6		5.5

the group and to the fact that their foster homes were of a higher grade, even within the given range.

Table XXXIV presents a summary of the correlations mentioned above, together with certain others found for the International Test. As stated before, this test does not yield an intelligence quotient, so that it was necessary to eliminate age by the method of partial correlation. The correlations so found are all somewhat lower than those found with the Binet intelligence quotient. This difference is due in part to the fact that the method of partial correlation does not give results entirely comparable with those based on ratios. In order to compare the results on the Binet Test with those on the International Test, partials were worked out for certain groups between Binet mental age and home rating with chronological age held constant. Thus, in the case of the entire Home Group the correlation between the intelligence quotient and home rating is .48, while that between the International Test score and home rating is only .24. A part of this difference of 24 points is obviously due to the type of correlation used since the partial correlation with Binet mental age is only .36. The same general result is found for the two other groups for which the correlations were worked out.

TABLE XXXIV.—CORRELATIONS BETWEEN CHILD'S INTELLIGENCE AND HIS FOSTER HOME RATING

Group	Binet I. Q.		Binet M. A.*		International Score*	
	N	r	N	r	N	r
Home Group	401	.48 ± .03	401	.36 ± .03	298	.24 ± .04
Legitimate	260	.43 ± .03				
Illegitimate	141	.47 ± .04				
In homes 4 years	304	.47 ± .03				
Legitimate	185	.43 ± .04				
Illegitimate	119	.45 ± .05				
In homes 4 years, entered before 5	226	.41 ± .04				
Legitimate	114	.35 ± .06				
Illegitimate	112	.45 ± .05				
Entered before 2 years	156	.52 ± .04	156	.43 ± .04	80	.28 ± .07
Entered before 9 months	111	.46 ± .05	111	.37 ± .06	58	.30 ± .08

*In these correlations chronological age has been held constant by the method of partial correlation.

The remaining difference between the partial correlations may be due to the different nature of the tests used or to the fact that the International Test could not be given to the entire group and that those omitted were mainly the younger children for whom the highest correlations were found between Binet I. Q. and home rating.

The next comparisons to be made are those between the intelligence of the child and that of his foster parents as measured by the Otis Test. Table XXXV gives the average intelligence of the children for three groups of foster fathers, divided according to their intelligence.

TABLE XXXV.—RELATION BETWEEN THE INTELLIGENCE OF CHILDREN AND THAT OF THEIR FOSTER FATHERS

Foster Father's Intelligence	Child's Intelligence	N
Superior Mean I. Q.—122 Range in I. Q.'s—111 to 136	104.9	68
Average Mean I. Q.—102 Range in I. Q.'s—96 to 110	97.0	61
Inferior Mean I. Q.—87 Range in I. Q.'s—68 to 95	93.8	51
Total	......	180
Mean I. Q. 106 S. D. 16	99.0 15.6	

It is evident from Table XXXV that the more intelligent children were those with the more intelligent foster fathers. The correlation found between the intelligence of child and foster father is .37 ± .04. (N = 180). Very similar results are found when the foster mothers are considered as shown in Table XXXVI. The correlation for this group is .28 ± .04 (N = 255).

TABLE XXXVI.—COMPARISON BETWEEN INTELLIGENCE OF CHILD AND FOSTER MOTHER

Intelligence of Foster Mother	Mean Intelligence Quotient of Foster Children	N
Superior Mean I. Q.—115 Range in I. Q.'s—108 to 132	104.9	64
Average Mean I. Q.—100 Range in I. Q.'s—92 to 107	97.8	91
Inferior Mean I. Q.—83 Range in I. Q.'s—63 to 92	94.3	100
Total	....	255
Mean I. Q., 97.0 S. D. 14.0	98.4 15.1	

The average of the two parents on the Otis Test gives a correlation of .39 ± .04 (N = 169) with the child's intelligence as measured by the Binet Scale. For the 120 cases in which the child had the International Test, the partial correlation (age held constant) of this test score with the mid-parent Otis score was .16 ± .06. It might also be noted that, when the group is limted to children who entered their foster homes before they were two years of age, the correlation between their Binet intelligence quotient and the parents' Otis score is .39 ± .05 (N = 132). This result remains unchanged when the group is further limited to those entering foster homes before 9 months of age.

The correlations between the child's intelligence and the vocabulary test scores of the foster parents are substantially the same as those for the Otis Test. The vocabulary test was designed to give some measure of the general cultural level of the foster parents. The following correlations were found:

	r	N
Intelligence quotient of child with mid-parent vocabulary score	.36 ± .05	146
Intelligence quotient of child with father's vocabulary score	.27 ± .05	152
Intelligent quotient of child with mother's vocabulary score	.37 ± .04	224

Table XXXVII shows the intelligence of the children in relation to the occupational status of their foster fathers. This relationship is also indicated by a correlation of .37 ± .03.

TABLE XXXVII.—INTELLIGENCE OF CHILDREN ACCORDING TO THE OCCUPATIONAL STATUS OF THEIR FOSTER FATHERS

Occupational Class	Mean Intelligence Quotient of Foster Children	N
Professional	106.8	61
Semi-professional and business	101.1	160
Skilled labor	91.6	149
Semi-skilled to slightly skilled labor	84.9	19
Unskilled labor		5
		394

Two additional outcomes which are consistent with those already reported are furnished by two other correlations. The correlation between the intelligence of the child and the personal ratings of the foster parents was .49 ± .03 (for rating blanks see Appendix). The education of the parents and the intelligence of their foster children have a correlation of .42 ± .03.

It is apparent from the facts presented in this section that the intelligence of the children is rather closely related to the environment furnished by their foster homes. From such definite relationships it would appear that their environment had had a direct influence upon their intelligence. Before accepting such a conclusion, however, it will be necessary to consider the possibility that the brighter children were selected for placement in the more superior homes.

7. The Extent to Which Selection is a Factor

Selection of the type mentioned would seem to depend upon two conditions. The first is the probability that parents or Society workers would intend to employ intelligence as the basis of their selection; the second is the possibility that they would be able to do so. Thus, if it were desired to select a child upon the basis of his intelligence, it would be necessary to make a correct judgment of his mental level before such selection could actually operate. Such a judgment, in most cases, would have to be based upon a personal observation of the child and a reference to his family history, if such were available.

The general circumstances of placement will be discussed next to show to what extent such selection is likely. This will be followed by the presentation of some evidence from the study concerning the extent of selection with given groups.

Reference to the application blanks filed by prospective foster parents and inquiries made in the study indicate that the items most frequently considered were sex, age, and nationality of the child desired. Among other qualifications mentioned that of health was most in evidence. Occasionally, parents asked for a child of a given complexion or for a well-behaved child. Examples of other requests are "not illegitimate," "of good parents," "a real boy," or a "normal child." When any requirement regarding mentality was made, it was only that the child be of normal ability. Such requests were made by inferior as well as by superior parents. The application blanks, of course, do not tell the whole story, but they indicate the qualifications that were in the mind of the parents when they first asked for a child. The circumstances of the actual choice, however, show that these preconceived ideas concerning qualifications were of less importance than the personal, and often unexplainable, attractiveness of a certain child. In many cases the entire procedure was of so sentimental a nature that there seems to be little probability of intentional selection on the basis of mentality. In some cases a family was so desirous of having a child immediately that the number of available children from which to make a choice was quite limited. Sometimes parents wished to adopt a child immediately to take the place of one who had died; others wanted one before the Christmas season or before going on their summer vacation. There are a number of instances in which a family took a child for a temporary period only and later were unable to part with it. The whimsical nature of some of the requests is shown by such requests as "a child with big brown eyes," "a blonde with curly hair," or "a boy that would like a big collie dog we have." There is an instance in which a mother suddenly decided to adopt a baby because she thought it would look "so sweet" in a coat that she could make from some left-over pieces of white fur. The child adopted now has an intelligence quotient of 72.

As the health and personal attractiveness of the child were the chief bases for choice where several children of the sex and age desired were available, it would seem that intelligence could hardly have been an important factor. In the many cases in which the child taken was the only one considered or even seen by the foster parents, the only possible selection would be that made by the Society representatives. It is possible, of course, that these workers may have refrained from offering to a superior family a child who in their judgment was of inferior mentality. It is not likely, however, that they would withhold an available bright child from a somewhat inferior family. As the number of young children available for placement has usually been much less than the number of applications, the probability of such selection is greatly reduced.

The majority of the children in the present study were placed in their foster homes a number of years before the days of scientific social work. The volunteer workers in the earlier days were actuated mainly by religious motives, many of them being ministers whose chief concern was to find a home for homeless children. The present workers are specially trained in social service and make efforts to secure all possible information that will enable them to "fit the child to the home." Such an adjustment, however, depends as much upon the behavior and personality of the child as upon his intelligence.

Even though the parents and Society workers had wished to select a child on the basis of mentality, it still remains to be shown whether they would have been able to do so with any degree of accuracy. The large proportion of the children who were quite young at the time of their commitment (over 60 percent less than 4 years of age) would make almost impossible any valid rating of intelligence based only upon personal observation. Some basis for estimating the level of intelligence of the child would be furnished by the family history records. The information in these histories is relatively meager with respect to the mental condition of the parents. In only about half the cases is any such information given and then only in the form of statements as to whether the parents seemed normal or mentally defective.

It is quite probable that the more recent Society workers, trained in social service, and with the more complete family records available, would exercise some degree of selection in the placement of the children who are under their charge. They would probably hesitate to place a dull child or one of quite defective parents in a superior home. The effect of such selection, however, could hardly be great enough to account for the high correlations found in the present study, since the children concerned were not recent placements and their family histories were relatively meager. Since definite information as to the mental condition of the parents was available in only the small percentage of cases where mental tests had been given, any valid estimate of the intelligence of the children would seem very improbable. It should further be noted that only rarely were the foster parents told the entire history of the child in question.

Some evidence on the question of selection is furnished by a consideration of the facts found for certain groups. It has already been noted that 74 children of the Home Group had mental tests before their entrance into foster homes. In these cases definite rating of the child's mental ability was known by the Society placement workers and usually by prospective foster parents. The intelligence quotient, however, often did not have the correct significance for either group. There are evidences in the records that it was considered as a grade comparable with school marks. An intelligence quotient of 90, for example, would be considered superior, while one of 85 would be thought of as average. Another misconception was that this score increased as the child grew older. A remark from the letter of an early Society worker reads: "We had him tested at the Institute for Juvenile Research here in Chicago and found that he has an intelligence rating of 85, which is *very good for a boy of his age.*"

Selection evidently existed with this Pre-Test Group, since the correlation between the intelligence of the children and the rating of the home in which they were adopted was .34. As already shown, this value would appear to represent the maximal amount of selection possible. The significant fact to be noted, however, is that this value was found to have increased to .52 after about 4 years of residence in the foster home. This would indicate that the effect of the new environment had been considerable, even with children at the mean age period of 8 to 12 years.

If the judgment of the child's mentality were based upon personal observation, the reliability of such a rating would be less in the case of the younger children. This would imply less selection with groups of very young children and a consequently smaller correlation between the child's present intelligence and his foster home rating, if selection were an important factor. Consideration of a group of children who were placed quite young, however, shows that the correlation is fully as great as with the entire Home Group. In the case of the 156 children of the Home Group who were placed before they were two years of age, the correlation between intelligence and home rating is $.52 \pm .04$, as compared with $.43 \pm .03$ for those placed after that age. It should be noted that this group contained 111 (70 percent) illegitimate children, while the percentage of such children in the Home Group was only 35. In view of the fact that the records concerning the parents of the illegitimate children are relatively incomplete, selection based upon such records would be much less possible with this two-year group than with the Home Group as a whole. It is evident, therefore, that where selection on the basis of an observation of the child's mentality is least possible, the observed correlation with home rating is still high.

In order to determine the extent to which there is selection on the basis of the child's family history a study was made of a group of 59 cases for which there was practically no information on the parentage. The correlation of the present intelligence of these children with their foster home rating was found to be .51 ± .06, showing that selection on the basis of the family history was relatively ineffective. It is further worthy of note that the mean age at which these children were placed was 2 years, 4 months, so that the possibilty of selection according to observed intelligence was also slight.

It should be noted that the mean home rating of this group is 18.6, which is practically identical with that of the entire Home Group. This would indicate that superior as well as inferior parents were willing to take young children with practically no information concerning their antecedents, which is further evidence that selection is not an important factor.

The question arises [9] whether there may have been a selective placement on the basis of race or nationality. Thus, negroes or South-European children may have been placed with families of their own race and intellectual level. Such selection would produce a correlation between intelligence and home rating because of racial differences.

All the children in the present study were placed in Protestant homes. Of the 401 children of the Home Group, 34 were colored or partly colored, and were placed in corresponding homes. Only three of the remaining 367 children were of South-European descent. They were placed with American parents of North-European ancestry.

In order to determine the effect of race selection, the negro children were omitted from the Home Group, and the correlation between intelligence and home rating computed. This coefficient was

---

[9] The authors are indebted to Professor Terman for this suggestion.

found to be .48 ± .03, which agrees to the second decimal place with that found when the negro children were included. It is apparent, therefore, that selection according to race had no effect upon the observed correlation.

In order further to investigate the possibility that the correlations were affected by race as a spurious factor or by irregularities in the scale as applied to younger and older children, certain of the correlations were recalculated, including only white children of from five to fourteen years of age, inclusive. The comparison of these correlations with those previously reported for the whole group is as follows:

Pre-Test Group. Home rating with I.Q. before and after foster home care.
Whole group (n = 74): before .34; after .52.
Restricted group (n = 51): before .36; after .57.

Limited Sibling Group. Intelligence of siblings separated. Oldest with youngest in total group.
Whole group (n = 125) .34.
Restricted group (n = 63) .44.

Foster-Own Group. Intelligence of own with foster child.
Whole group (n = 40) .34.
Restricted group (n = 17) .65.

Home Group. Parents' I. Q. with child's I. Q.
Whole group (n = 167) .39.
Restricted group (n = 112) .47.

Two-Year Group. Child's I. Q. with home rating.
Whole group (n = 156) .52.
Restricted group (n = 104) .50.

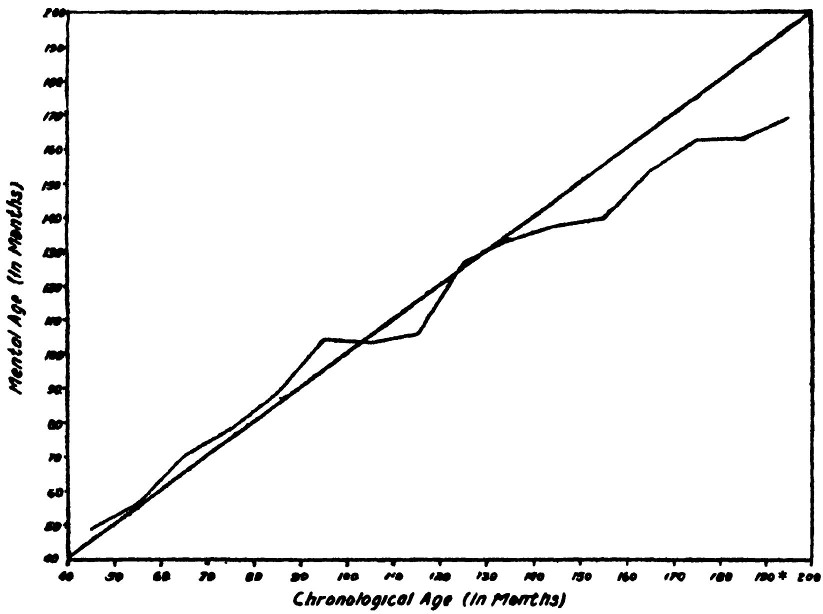
It is obvious that the correlations are not significantly affected by the racial composition of the group nor by the defective standardization of the scale at the upper and lower ages.

Selection according to age is shown by the fact that there is a correlation of − .27 ± .03 between home rating and age of placement for the entire Home Group. In order to eliminate any influence this selection might have upon the correlations between intelligence and home rating, the factor of age of placement was held constant by means of partial correlation. The observed correlation of .48 is thus reduced to .43. For the legitimate children only, there is a correlation of $-.17 \pm .04$ between the rating of the foster home and the age of the child taken. When age of placement is held constant with this group, the observed correlation of .43 between intelligence and home rating is reduced only to .41.

In this discussion regarding circumstances of placement an attempt has been made to determine the extent to which selection could affect any comparisons with home rating. It has been stated that, in order for selection to be of importance, it would be necessary that the foster parents or the Society workers should intend to choose upon the basis of mentality. Such an intention could not be carried out unless it were possible to make a correct judgment of the intelligence of the child from a personal observation and from such family records as were available. When two groups of children were considered for which such conditions for selection were least possible, it was found that the correlations with home rating were practically the same as those obtained for the whole group. While some general selection doubtless exists, it has been shown that it could not account entirely for the close relationships found between the foster child's present intelligence and the character of his environment.

8. Relationship Between Intelligence Quotient and Chronological Age

In a group of normal individuals it would be expected that the relationship between mental age and chronological age would be linear, at least up to 16 years of age. For the Home Group of the present study, however, the curve formed by plotting the mean mental ages according to chronological age is seen to drop considerably below the standard line, as shown in Figure 3. Up to about 8 years of age the curve of means is above the standard, indicating intelligence quotients of above 100. For children over 8 years of age the mental age curve begins to drop below the standard until the mean intelligence quotient of the 15-year-olds is only 89.4. The drop in intelligence quotients for these years is not so pronounced as the mental age curve would appear to indicate from a superficial examination, since the quotient is less affected by a given difference at the higher age levels. The correlation between the intelligence quotient and chronological age for the Home Group is $-.34 \pm .03$. If the legitimate children only are considered, the correlation is $-.29 \pm .04$. For the entire Committed Group the correlation is $-.38 \pm .03$.



* All cases with a chronological age above sixteen years are included in this interval.

FIG. 3. MENTAL AGE CURVE FOR HOME GROUP COMPARED WITH STANDARD

This drop in mental age and in intelligence quotient is so considerable as to need explanation. It might be due to the character of the group studied or to the nature of the test. It was noted that most of the older children were those who had entered their foster homes relatively late—the correlation between age of commitment and age at test was $.53 \pm .02$ for the Home Group. It might be that these children were not able to reach the level of intelligence of those who entered their foster homes at a much younger age. If this were true, it would explain the fact that the older children test lower. If age of commitment is held constant, the correlation between intelligence quotient and age at test is reduced to $-.23 \pm .03$ for the entire Home Group and to $-.24 \pm .04$ for the legitimate children alone. It would thus seem that the drop in intelligence is partly, but not entirely, due to this characteristic of the group.

In order to show how the drop depends upon the age of commitment, separate curves were made for the children who entered their foster homes before they were two years of age and for those who entered after that time. As shown in Figure 4, the curve for the group committed early appears to coincide fairly well with

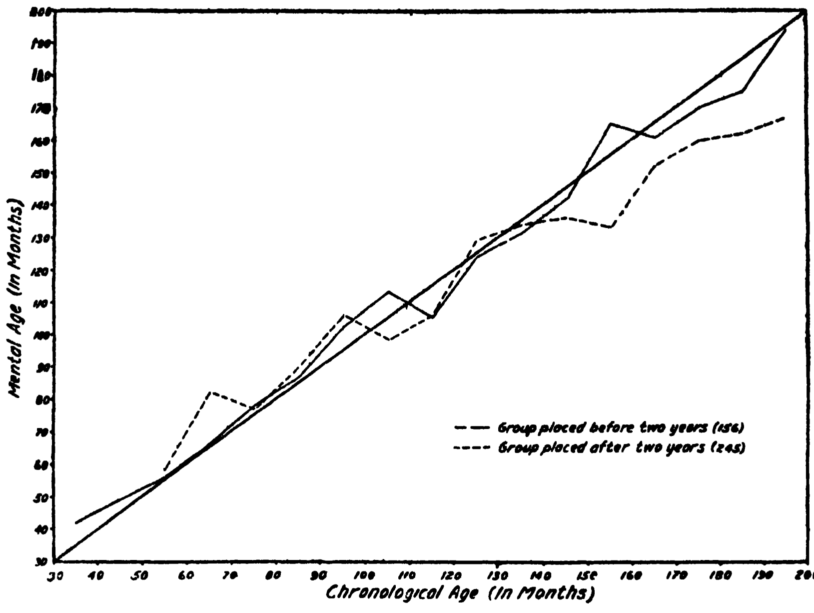


FIG. 4. MENTAL AGE CURVES FOR GROUPS PLACED BEFORE AND AFTER TWO YEARS OF AGE

the standard, as the means are slightly above normal up to about 9 years of age and only a little below normal at the older ages. The correlation between intelligence and age for this two-year group is only $-.12 \pm .05$. In the case of the group committed after two years of age, the drop is more marked for children older than 11 years. It thus appears that a part of the drop in I. Q. is due to the age of commitment.

In order to see if the type of home affected this drop in any way, the group who entered their foster homes before they were

5 years of age was divided into the better and the poorer half of homes. These curves appear to have about the same slope as shown in Figure 5. The correlations between I. Q. and age for the two groups (limited to those who are at least 6 years of age) were $-.23 \pm .06$ for the better home group and $-.20 \pm .06$ for the poorer home group. It would seem, therefore, that the type of

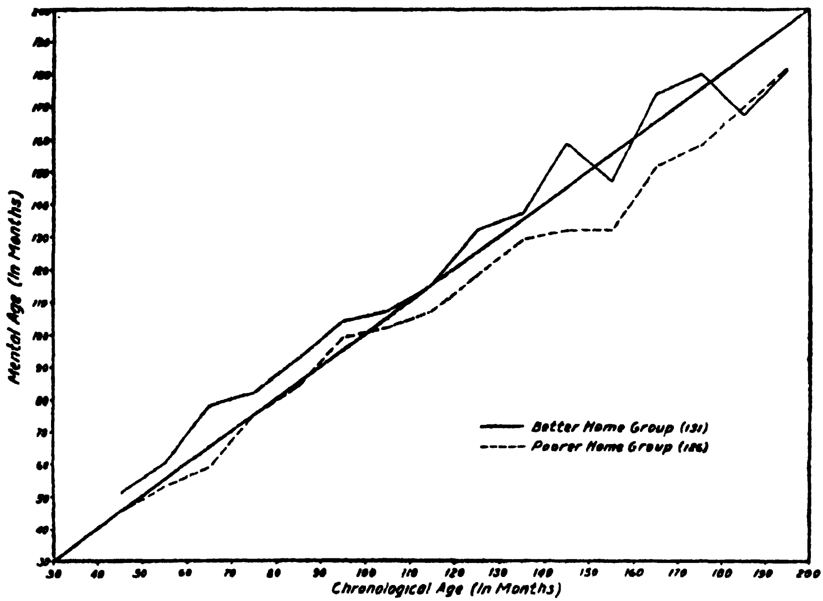


FIG. 5. MENTAL AGE CURVES FOR GROUPS IN BETTER AND IN POORER HOMES

home apparently has little effect on the observed drop of the intelligence score with age.

The same tendency for the older children to test lower was found in the case of children tested upon their commitment to the Society. Figure 6 shows the mental age curves for two such groups, one the group of 247 children tested by the Institute for Juvenile Research, the other the Newly Committed Group of 142 tested during the present study. Both show a marked drop with age. The correlation between intelligence and age for the former group is $-.30 \pm .04$, which is reduced to $-.21 \pm .04$ if those above 14 years of age are eliminated. For the Newly Committed Group, the correlation is $-.39 \pm .05$, all cases considered. In the case of these two groups the drop might be due to the fact that relatively dull children were being considered, as the average intelligence quotients were 84 and 89. There may be a special tendency for dull children from a poor environment to fall increasingly below normal. Another possible explanation is that the older children

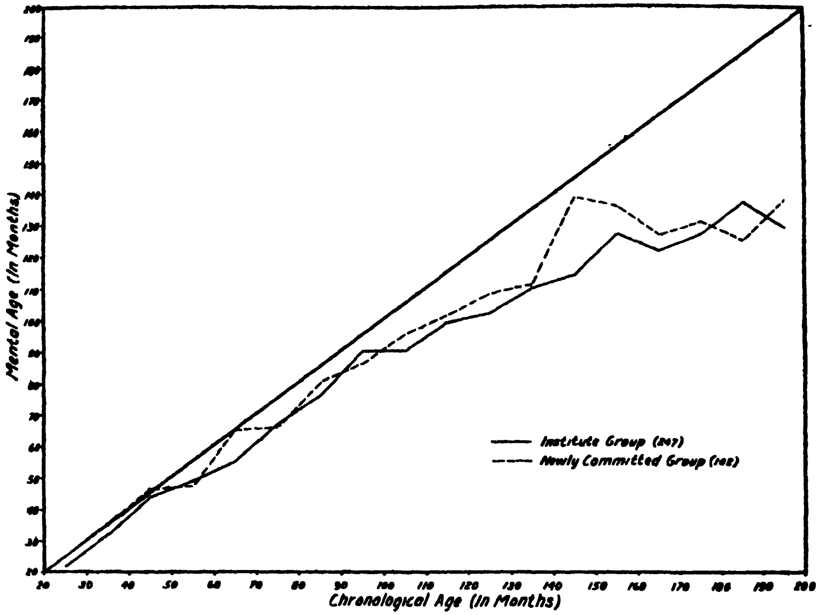


FIG. 6. MENTAL AGE CURVES FOR INSTITUTE GROUP AND NEWLY COMMITTED GROUP

might represent a select group from which the younger and brighter had been eliminated, perhaps previously placed in foster homes. In the discussion of selection, however, it was shown that there was no appreciable tendency of that sort.

A final possibility is that the drop in the intelligence quotient which has been observed with all the groups may be due to the nature of the test itself. In order to investigate this possibility, it was necessary to secure some normal group of unselected children. It was thought that this would be furnished by the Terman group of 905 unselected children referred to in the first section of this chapter. The mean intelligence quotients for the different ages (5 to 14 years only[10]) were worked out from the percentage distributions given by Terman. These are compared in Figure 7 with those found for the Home Group. The drop in I.Q. for the Terman group appears to be almost as great as in the case of the Home Group of foster children.

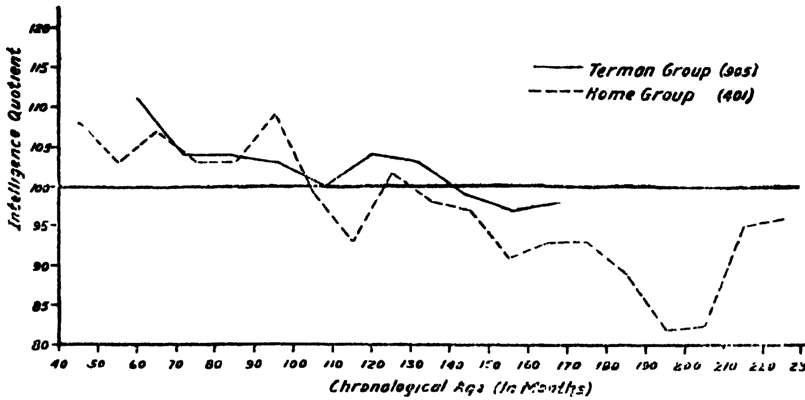


Fig. 7. Change in Intelligence Quotient with Age for Terman Group and Home Group

The correlation between chronological age and I.Q. was worked out for the Terman group, using the percentage distributions given for each of the ten age-levels. The coefficient was found to be $-.24 \pm .02$, which is clear evidence that intelligence quotients based upon the Stanford Revision are markedly affected by age. It is interesting to note that the corresponding correlation for the Home Group, limited to the same age range, 5 to 14 years, is $-.29 \pm .04$ ($N = 301$). If age of commitment is held constant, the latter coefficient is reduced slightly. It would thus seem that the observed drop for the Home Group (age-range 5 to 14 years) is

[10] Terman points out that the 14-year-olds are a somewhat selected group, since some of the brighter children of that age had been promoted to the high school and were therefore not tested. He also states that the 5- and 6-year-old children may be selected, but finally includes all three ages in his group of "905 unselected children." *The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence.* Chapter II.

not significantly greater than that found for Terman's unselected group.

Table XXXVIII gives the mean intelligence quotients[11] for the various age groups used by Terman. The smoothed values were obtained from the regression line and are those which have been used throughout the study to correct roughly for age differences of groups. The drop in intelligence as measured by the Stanford Revision is clearly evident.

A still further limitation in the use of the Stanford Revision arises in testing children who are above 14 years of age. Since the highest mental age possible is nineteen years, six months, no 14-year-old, however bright, can make an intelligence quotient above 139; no 15-year-old, above 130; no person 16 or over, above 122. It is possible that this limitation in the upper range affected some of the children tested in the study, since about one-fourth of the

TABLE XXXVIII.—MEAN INTELLIGENCE QUOTIENTS FOR TERMAN UNSELECTED GROUP OF 905 CASES DIVIDED ACCORDING TO CHRONOLOGICAL AGE

Central Age in Years	Number of Cases	Mean Intelligence Quotient	Smoothed Values of Means from the Regression Line
5	54	111.0	107.6
6	117	104.1	106.4
7	93	104.3	105.2
8	98	102.9	104.0
9	113	100.2	102.8
10	87	104.1	101.6
11	79	102.6	100.4
12	83	98.8	99.2
13	98	97.4	98.0
14	82	98.2	96.8
Mean		102.0	

[11] These means were computed from the distributions given by Terman on pages 33-37, *op. cit.* They are correct for the percentages and intervals given, but a number of discrepancies were found in the distributions and medians given in the report. For example, Terman gives the median for the entire group as 99, but from the distribution given it would appear to be 101. The median of the 5-year-olds, given as 102, is quite evidently 109. Several other errors in medians and percentages occur.

Home Group were at least 14 years of age. One comparison that would be considerably affected by this limitation is that between the intelligence of the legitimate and illegitimate children, since nearly one-third of the legitimate group were at least 14 years of age when tested, while less than one-seventh of the illegitimate children were in that age range.

It would also appear that the Stanford Revision is somewhat unreliable at the younger ages, since Terman's group furnished mean I.Q.'s well above 100 and the coachability of some of the tests at these ages is obvious. Approximately 20 percent of the Home Group were less than 7 years of age when tested. The percentages for the legitimate and illegitimate groups were 8 and 40 percent, respectively, which no doubt accounts in part for the relatively high scores of the illegitimate children.

In order to study the change in I.Q. with age for other groups, superior children who had been in a good environment all their lives were considered. Curves showing the relation of the intelligence quotient to age were made for 673 children of the University of Chicago Elementary School and for a group of 187 children tested under the direction of Professor Bird T. Baldwin at the University of Iowa. These curves are shown together in Figure 8.

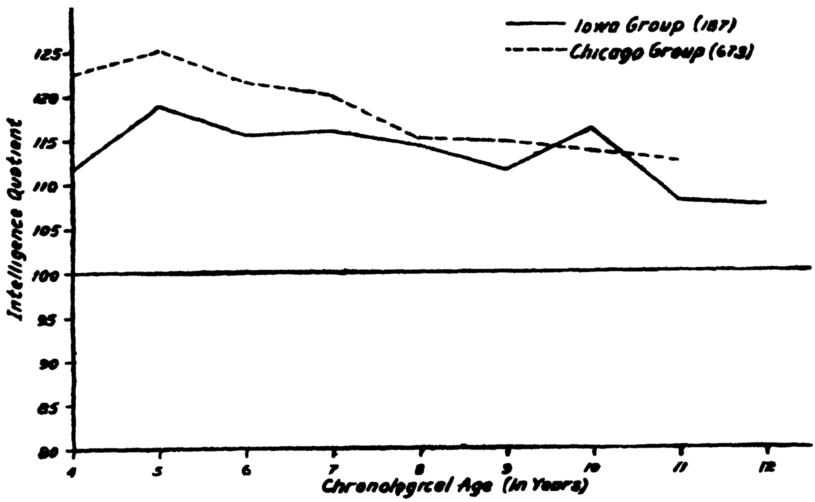


FIG. 8. CHANGE IN INTELLIGENCE QUOTIENT WITH AGE FOR CHICAGO AND IOWA GROUPS

They indicate the same drop with age as that found with normal groups. The drop is expressed by the correlations of $-.28 \pm .02$ for the Chicago group and $-.21 \pm .05$ for the Iowa group. These values are very nearly the same as those found for the Terman unselected group ($r = -.25$) and are evidence of the fact that, even with a superior group, intelligence quotients from the Stanford Revision diminish with age.

Negative correlations between I.Q. and age have been found in a number of other studies of dependent children. Thus Miss Elderton[12] obtained a relation of $-.31$ for a group of orphanage children tested with the Stanford Revision. She concluded that either the tests were at fault or that residence in an orphanage decreases the intelligence of the inmates.

In a study of children who were public charges, Stenquist and others[13] used the Goddard Revision of the Binet Scale. They found the same tendency for the older children to be increasingly below the normal in intelligence. Possible explanations offered were that the brighter older children had been eliminated from the group because they were chosen for placement in foster homes or because of their earning power, or that there is some special tendency for children of this sort to fall behind with advancing years.

The results presented in this section appear to show that the pronounced drop in I.Q. with age may be accounted for by the special character of the group studied and the nature of the test. When the age of entrance into the foster homes is restricted to the earlier years, the drop is considerably reduced, but is still in evidence, owing to the imperfect standardization of the test.

Any comparisons made between the Stanford-Binet intelligence quotient and such factors as length of time in foster home and age of entrance into home are complicated by the drop in I.Q. in the later ages. The relationships between these factors are so involved that the method of partial correlation has been used in an attempt to disentangle them.

---

[12] Ethel M. Elderton. "A summary of the present position with regard to the inheritance of intelligence." *Biometrika,* 1923, pp. 378-408.

[13] Stenquist, Thorndike, and Trabue. "The intellectual status of children who are public charges." *Archives of Psychology,* 5 :Sept., 1915, No. 33.

9. Intelligence as Related to Length of Time in Foster Home

It might be expected that there would be some relationship between the child's present intelligence and the length of time he has spent in his foster home. Some indication of such a relationship is shown in Table XXXIX. While the zero-order correlations are negative for each group, they become positive when age tested is held constant. Since two of the four coefficients are more than three times their probable errors, there is some evidence of a slight correlation between intelligence and the number of years spent in the foster home. The correlation is not in evidence in the zero-order coefficients because of the relation between I.Q. and age discussed in the preceding section.

TABLE XXXIX.—RELATION BETWEEN PRESENT INTELLIGENCE OF THEFOSTER CHILD AND THE LENGTH OF TIME HE HAS SPENT IN THE FOSTER HOME

Group	Correlation Between Binet I. Q. and Length of Time in Home	
	Zero Order Coefficient	Partial (Age tested Constant)
Committed Group (n=484)	−.02±.03	+.13±.03
Home Group (n=401)	−.05±.03	+.08±.03
Sibling Group (n=185)	−.11±.05	+.06±.05
Two-Year Group (n=156)	−.09±.05	+.19±.05

Supporting evidence of the relationship shown in Table XXXIX is given by the partial correlation (age constant) between the International Test score and the length of time in the foster home environment. This coefficient, .18 ±.04, must be regarded as significant, since it is over four times its probable error.

The slight value of these correlations may be due to the fact that intelligence improves very rapidly at first upon a change in environment and soon reaches its maximal level. This form of regression would tend to reduce correlations involving time. Some evidence of this rapid development in intelligence within the first few years after placement is shown in the case of the Pre-Test Group. The children who were tested before entering their foster homes and then retested several years later showed considerable gain in intelligence. The group who were tested for the first time after several years of residence in their foster homes and then retested later failed to show any significant gain (change = $+2.9 \pm 1.8$, when corrected for age).

Another, and perhaps better, measure of environmental influence is given by the percentage of the child's life spent in some foster home. The correlation between this percentage and the child's present intelligence is $.34 \pm .03$ for the Committed Group, showing that the more intelligent children are those who have spent the most of their lives in a good environment.

In view of these correlations, it would seem that there is a tendency for high intelligence to be associated with the length of time the foster home has exerted its influence or the percentage of the child's life spent under such influence.

10. Intelligence as Related to Age of Placement

If environment tends to improve mental capacity, it might be expected to have the greater influence during the earlier and more plastic period of the child's life. It would, then, be expected that, other things being equal, the earlier in life the child is brought under this improved environment, the higher would be his present intelligence. Evidence that there is some relation between intelligence and age at commitment is shown in Table XL.

The means for the combined legitimate and illegitimate groups drop from 102 to 88, indicating that the brighter children are those who were committed at an early age. This result is substantiated by correlations between intelligence and age of commitment. For the entire Committed Group, the correlation is $-.44 \pm .02$. The coefficient for the legitimate group ($N = 330$) is $-.35 \pm .03$, while that for the illegitimate group ($N = 150$) is $-.24 \pm .05$. If only those children who are committed before five years of age are considered, the correlations are reduced to $-.01 \pm .04$ and $-.06 \pm .05$, respectively, which would seem to indicate that children of all ages up to five years may be equally influenced by an improved environment.

These results were based on the entire Committed Group. In order to study the bearing of the home rating factor upon the relationships found, it was necessary to work out corresponding

TABLE XL.—DISTRIBUTION OF INTELLIGENCE FOR THE ENTIRE COMMITTED GROUP ACCORDING TO COMMITMENT AGES

Intelligence Quotient	Committed Before 2 Years		Committed From 2 to 5 Years		Committed After 5 Years		Total	
	Legitimate	Illegitimate	Legitimate	Illegitimate	Legitimate	Illegitimate	Legitimate	Illegitimate
160–164	—	1	—	—	—	—	—	1
155–159	—	—	—	—	—	—	—	—
150–154	—	—	—	—	—	—	—	—
145–149	—	1	—	—	—	—	—	1
140–144	—	—	—	—	—	—	—	—
135–139	1	—	1	—	—	—	2	—
130–134	—	—	—	—	1	—	1	—
125–129	3	8	1	1	—	—	4	9
120–124	—	6	2	2	—	—	2	8
115–119	1	16	5	3	3	—	9	19
110–114	4	12	8	1	8	—	20	13
105–109	4	16	6	—	13	1	23	17
100–104	7	16	21	3	17	—	45	19
95– 99	6	17	9	4	17	2	32	23
90– 94	9	12	12	3	19	2	40	17
85– 89	5	4	10	2	23	1	38	7
80– 84	6	3	9	2	22	1	37	6
75– 79	4	4	7	—	21	—	32	4
70– 74	3	3	3	—	19	1	25	4
65– 69	—	1	1	—	9	—	10	1
60– 64	2	—	1	—	7	—	10	—
55– 59	—	—	—	—	4	1	4	1
Total	55	120	96	21	183	9	334	150
Mean I.Q.	95.2	105.1	97.1	102.5	87.7	—	91.5	103.7
Mean I.Q. for combined groups	102.0		98.1		87.7		95.3	

correlations for the Home Group children only, using age of placement instead of age of commitment. The correlation between intelligence and placement age for this group is $-.30 \pm .03$ ($N = 401$). When the factor of home rating is held constant, the correlation is reduced to $-.20 \pm .03$. This reduction is due to the fact that there is a tendency for the better homes to adopt younger children, as shown by a correlation of $-.27 \pm .03$ between home rating and age of placement.

Since the illegitimate children have a somewhat superior heredity and are placed quite young in relatively superior homes, it is possible that this correlation found for the group as a whole might be due to some other factor than that of age of placement. Consequently, the correlation was worked out for the legitimate children separately. The relation between intelligence and placement age for these children alone was found to be only $-.17 \pm .04$ ($N = 260$), which would seem to indicate that a part of the correlation of $-.30$ found for the entire Home Group was due to the superior intelligence of the illegitimate children. Furthermore, when home rating is held constant with the legitimate group, the correlation is reduced to $-.09 \pm .04$. This last value is hardly significant, showing that age of placement may be a negligible factor in the case of the legitimate children.

It might be finally noted that the correlation of .48 found between intelligence and home rating for the Home Group is reduced to $.43 \pm .03$ if age of placement is held constant.

11. Comparison of Legitimate and Illegitimate Groups

From the comparisons already made, it is evident that the intelligence of the illegitimate children is superior to that of the legitimate. The mean intelligence quotients, home ratings, and ages for the two groups are compared in Table XLI. The difference of 10.2 points in intelligence in favor of the illegitimate children is evidently due in part to the grade of their foster homes, to their age at placement, and to their test age, since the table indicates their advantage in each of these respects.

TABLE XLI.—COMPARISON OF LEGITIMATE AND ILLEGITIMATE GROUPS

Group	N	Intelligence Quotient		Foster Home Rating		Mean Age	
		Mean	Standard Deviation	Mean	Standard Deviation	When Placed	When Tested
Entire Home Group.........	401	97.5	15.9	18.8	5.6	4–2	11–0
Legitimate.......	260	93.9	15.0	17.9	5.4	5–8	12–2
Illegitimate......	141	104.1	15.4	20.3	5.2	1–7	9–0

The home rating of the illegitimate children was 2.4 ±.4 higher than that for the legitimate group, indicating a distinctly better environment. Another important difference between the two groups is shown in their mean ages of placement which indicate that the illegitimate children were placed quite young—more than four years earlier than the legitimate children. In view of the nature of the Binet Test, the difference of 3 years, 2 months, in the age at which the test was taken indicates a further advantage for the illegitimate children. According to the table of corrections used, almost 4 points of the observed difference in the mean intelligence quotients can be accounted for by the test. This would leave 6.4 points as the measure of the difference in actual intelligence of the two groups.

In order to determine whether the combined effect of these three factors can account entirely for the observed difference in the intelligence of the two groups, an attempt was made to hold all of them constant by using matched cases. It was possible to find 55 pairs of legitimate and illegitimate children (committed before 2 years of age) for which the factors were substantially equal. The results of the matching are shown in Table XLII.

The difference in the mean intelligence quotients for these matched groups is 4.7 ± 1.9 which is less than three times the probable error and therefore hardly significant. Therefore, it may be concluded that the intellectual superiority of the illegitimate chil-

TABLE XLII.—COMPARISON OF LEGITIMATE AND ILLEGITIMATE CHILDREN BY 55 MATCHED PAIRS

Factor	Legitimate	Illegitimate
Home rating		
Range	8 to 29	8 to 30
Mean	17.75	17.69
Age of commitment (in months)		
Range	0 to 23	0 to 22
Mean	9.3	6.9
Age at test (in months)		
Range	46 to 200	44 to 213
Mean	117.5	116.1
Mean intelligence quotient	95.2±1.4	99.9±1.3
Standard deviation	15.9	14.6

dren is due chiefly to such environmental factors as type of foster home and age of commitment. The superior heredity of the children, then, appears to have but a slight effect upon their present intelligence.

VII. The School Achievement of Foster Children

It has already been shown that the Home Group with their inferior heredity were substantially normal in intelligence. In the present chapter the school progress of these foster children will be studied in order to determine whether their achievement is up to the usual standards.

The age-grade placement of the 348 children of the Home Group who were old enough to be in school is shown in Table XLIII. Fourteen of these 348 were not in school at the time of the study, but they were included in the table as in the last grade attended. It should be noted that all of these 14 had completed at least the seventh grade before dropping out of school. The table shows that 69 percent of the children were in the normal grade for their age, while 11 percent were accelerated and 20 percent were retarded.

In Table XLIV the group has been limited to those who entered their foster homes before they were six years of age, in order to show the progress of those whose entire school life has been spent in the improved environment. The percentage of normally placed children in this group is 75, while 14 percent are accelerated and only 11 percent are retarded.

The school progress of these foster children compares very well with that for the country as a whole. In the Gary survey, made in 1918, there is a report of the age-grade placement of the children of four cities.[14] Percentage distributions from this study are compared in Table XLV with those found for the foster group.

A comparison of the percentages for the foster groups with those for the four cities shows that the former have a much larger proportion of children who are in the normal grade for their age, while the percentage of retarded children is considerably smaller.

---

[14] A. Flexner and Frank P. Bachman. *The Gary Schools, General Account.* 1918. New York. General Education Board. pp. 181.

TABLE XLIII.—AGE-GRADE PLACEMENT OF 348 FOSTER CHILDREN

Grade		AGE IN YEARS 5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Total	Accelerated	Normal	Retarded
	Kg.	9		1																10	0	9	1
	I		20		1															21	0	20	1
	II		1	20		3														24	1	20	3
	III			5	23		5													33	5	23	5
	IV				1	9		3	4											17	1	9	7
	V				1	1	22		9	2										35	2	22	11
	VI						8	36		8	3	1								56	8	36	12
	VII							7	26		5	6	1							45	7	26	12
	VIII							1	4	20		5	2	1						33	5	20	8
	IX									5	23		5							33	5	23	5
	X											10		1	2					13	0	10	3
	XI										1	2	9		1					13	3	9	1
	XII													5						5	0	5	0
College	I													1	5					7	1	5	1
College	II															1				1	0	1	0
College	III																1			1	0	1	0
College	IV																	1		1	0	1	0
Total..																				348	38	240	70
																	Percent				11	69	20

TABLE XLIV.—AGE-GRADE TABLE OF CHILDREN PLACED IN FOSTER HOMES BEFORE SIX YEARS OF AGE

Grade	AGE IN YEARS																		Total	Accelerated	Normal	Retarded
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22				
Kg.	9		1																10	0	9	1
I		20		1															21	0	20	1
II		1	19		1														21	1	19	1
III			4	19		5													28	4	19	5
IV				1	6		2	2											11	1	6	4
V				1	1	17		4											23	2	17	4
VI						6	26		4	1									37	6	26	5
VII							6	15		1									22	6	15	1
VIII							1	3	14		3								21	4	14	3
IX									4	12		2							18	4	12	2
X											5								5	0	5	0
XI										1	2	6							9	3	6	0
XII													3						3	0	3	0
College I													1	2					3	1	2	0
College II															1				1	0	1	0
College III																1			1	0	1	0
Total..																			234	32	175	27
	Percent																			14	75	11

TABLE XLV.—COMPARISON OF SCHOOL PROGRESS OF FOSTER CHILDREN WITH THAT OF CHILDREN OF FOUR CITIES

City	Percent Accelerated	Percent Normal	Percent Retarded
Dubuque	18	49	33
Rockford	21	42	37
Rochester	9	53	38
Gary	22	40	38
Entire Home Group	11	69	20
Those of Home Group placed before they were 6 years of age	14	75	11

It is, therefore, apparent that the foster children made better progress in school than did those in the four cities studied.

It should be noted that the children of the Home Group were in a great variety of schools, varying from the one-room rural school to first-class private schools. In spite of the different standards of grade placement, however, it was found that the percentages did not vary greatly when the schools were classified as shown below.

	N	Accelerated	Normal	Retarded
Chicago and suburbs	213	10	74	16
Other cities and towns	75	13	66	21
Rural	60	12	56	32

In order to study the effect of age of placement and type of foster home upon school progress, the groups were classified as shown in Table XLVI. The most striking facts brought out in

TABLE XLVI.—SCHOOL PROGRESS ACCORDING TO AGE OF PLACEMENT AND TYPE OF FOSTER HOME

Placement Age	Better Homes N=161			Poorer Homes N=187		
	Percentage			Percentage		
	Accelerated	Normal	Retarded	Accelerated	Normal	Retarded
Before 2 years	23	75	2	11	74	15
Between 2 and 5 years	14	74	12	4	76	20
After 5 years	13	74	13	2	51	47

the table are that the percentage of accelerated children is relatively large (23 percent) in the case of those placed very young and in the better homes, while the percentage of retarded children is large (47 percent) for those placed at over 6 years of age in the poorer homes. The advantage of early placement in good homes is clearly evident.

A measure of school achievement was made up of combined ratings on age-grade placement and the relative standing of the child in the given grade. Normal age-grade placement was rated as 2, one grade accelerated as 3, one grade retarded as 1, two grades accelerated as 4, two grades retarded as 0, etc. If the child were doing average work in the grade, he was rated as 2, if superior work as 3, if inferior as 1, if very inferior as 0. The ratings on these two points were added for the "school achievement index." The normal index would thus be 4, indicating average work in the normal grade. A correlation of $.70 \pm .02$ was found between this school achievement index and the child's intelligence.

A distribution of these achievement indices according to placement age is shown in Table XLVII. It will be noted that the average index of the children placed before 2 years of age is 4.2, while that for those placed after 6 years of age is only 3.3. The close relationship between school achievement and early placement is shown by the coefficient of contingency, $C = .34$.

TABLE XLVII.—DISTRIBUTION OF SCHOOL ACHIEVEMENT INDICES ACCORDING TO PLACEMENT AGE

School Achievement Index	Placed Before 2 Years of Age	Placed at From 2 to 5 Years of Age	Placed After 5 Years of Age	Total
7	2	1	—	3
6	6	5	—	11
5	26	17	14	57
4	62	63	51	176
3	14	21	20	55
2	4	12	14	30
1	—	—	11	11
0	1	—	4	5
Total	115	119	114	348
Mean Index	4.2	3.9	3.3	3.8

The distribution of school achievement indices for the better and poorer homes separately is shown in Table XLVIII. The mean indices of 4.2 and 3.4 for those in the better and poorer homes, respectively, verify results already found. The correlation between school achievement and foster home rating is .34 ±.03.

TABLE XLVIII.—DISTRIBUTION OF SCHOOL ACHIEVEMENT INDICES ACCORDING TO TYPE OF HOME

School Achievement Index	Better Homes (Ratings 19 to 30)	Poorer Homes (Ratings 6 to 18)	Total
7	3	—	3
6	9	2	11
5	38	19	57
4	83	93	176
3	22	33	55
2	7	23	30
1	1	10	11
0	....	5	5
Total	163	185	348
Mean Index	4.2	3.4	3.8

Thus, it will be seen that the group of foster children studied had made more than normal progress in school. This may have been due in part to the special interest taken in the child's school work by the foster parents. The present level, however, could hardly have been attained unless the children had shown the necessary ability and application.

VIII. THE BEHAVIOR OF THE HOME GROUP

While the main problem of the study was to investigate the intelligence of foster children, some inquiry was also made regarding their behavior. The information concerning conduct was obtained chiefly through the questioning of the parents and teachers, supplemented by the observation of the child by the field worker. While there was no attempt to obtain a detailed record of the behavior difficulties, nevertheless some statement as to their character and seriousness was obtained in every case. It is possible that in some instances the parents may have given a biased report, but it is fairly certain that few, if any, cases of serious misconduct were overlooked.

Some of the most common types of behavior difficulties found are given in the following list:

Difficulty	*Number of instances*
Is disobedient, wilful, or incorrigible	22
Steals	17
Lies	16
Runs away, stays out at night, walks streets	12
Is domineering, quarrels, fights	12
Plays pranks at home, church, or school	12
Strikes parents	7
Masturbates	6
Plays truant from school	6
Destroys property	3

It should be noted that the total number of instances is far in excess of the number of children involved, since most of the children who were behavior problems exhibited several difficulties.

In only 32 cases out of the 401 in the Home Group was the misbehavior considered of a serious nature. While this number is only 8 percent of the total, it may not be entirely representative of foster children in general, since children exhibiting extreme misconduct might not have been in permanent foster homes and therefore not included in the study. This group of 32, however, includes some who had just been removed from their foster homes on account of misconduct.

Of these 32 cases, 13 were considered quite serious behavior problems. Five of these 13 children had been returned to the Society unreservedly, while five others had been placed under their temporary supervision. Eight of the 13 were girls; 5 were boys. The ages of the group ranged from 11 to 17 years. The mean placement age was 5 years, 1 month, which is somewhat greater than that of the entire Home Group. It is of interest to note that only one of these 13 serious cases was that of an illegitimate child.

In the case of the other 19 children, the difficulties were considered somewhat less serious. Of this group 5 were girls; 14 were boys. The mean placement age was 5 years, 8 months, while the present ages of the group ranged from 9 to 18 years. Two of the 19 were illegitimate children.

In other cases minor behavior difficulties were reported (chiefly for younger children), but these were not studied since it was felt that those reported represented only a small proportion of the children with such characteristics. Since it was difficult to determine just what constituted misbehavior in these problems with the younger children, it was thought advisable to exclude them from the summary.

In most of the 32 cases of serious misbehavior the difficulty was confined to the home or friends of the family. In 10 cases, however, the problem had become more generally known. Two of the boys had been sent to the St. Charles School for Boys (a state corrective institution), while five others had come to the attention of school truant officers, city police, or Juvenile Court representatives. The remaining three had had serious trouble in school.

The chief types of misbehavior for the 13 more serious cases were stealing, lying, and incorrigibility. In the less serious cases the difficulties involved minor instances of stealing and lying, school truancy, resistance to authority, and unsocial attitudes.

The mean I.Q. of the 32 children who were behavior problems was 89.0. The mean of the serious cases was 84.4, while that for the less serious cases was 92.2. Although the groups are far too small for any conclusions, it is evident that the group of behavior cases averaged slightly lower than the entire Home Group.

The home ratings for the 32 behavior cases average 17.9, which is not significantly lower than that of the entire Home Group. There is little evidence, therefore, that the behavior of the children is directly related to the type of foster home in which they have been raised.

In order to study the relation of misbehavior to the heredity of the children, the cases were grouped according to the parents' defects, as shown in Table XLIX. As noted before, young children were not considered in the behavior summary since it was felt that their difficulties might not be correctly interpreted, and also because they were not yet old enough to exhibit the more serious types of misbehavior. The percentages given in the table are, therefore, worked out on the basis of the children who were at least 9 years of age at the time of the study. It should be noted that, when this basis is used, the 32 cases of behavior difficulties represent 12 percent of the total group, rather than the 8 percent previously mentioned.

Of the 268 members of the Home Group who were at least 9 years of age, 100 (37 percent) had one or both parents mentally defective. Only 12 of these 100 children exhibited behavior difficulties; only three were considered as quite serious.

Out of the 145 children (at least 9 years of age) who had morally defective parents, 23 (16 percent) showed behavior difficulties. When this group was divided into those committed before and after 5 years of age, it was found that 20 percent of those

TABLE XLIX.—THE BEHAVIOR OF CHILDREN OF DEFECTIVE PARENTS

	N	Number of Children Exhibiting Misbehavior		Percent of Children Exhibiting Misbehavior	
		Serious	Less Serious	Serious	Less Serious
One or both parents mentally defective...	100	3	9	3	9
Committed after 5 years	42	0	4	0	10
Committed before 5 years..............	58	3	5	5	9
One or both parents morally defective....	145	10	13	7	9
Committed after 5 years	70	4	10	6	14
Committed before 5 years...............	75	6	3	8	4

committed after 5 years of age were behavior problems, while only 12 percent of those committed before that age were so considered. In the case of the mentally defective parents the corresponding percentages were 10 and 14, respectively.

It will be noted that, of the 32 children who were behavior cases, 23 (72 percent) had morally defective parents. Of the entire Home Group, 54 percent had morally defective parents. It would seem, therefore, that heredity had played some part in the behavior of these 32 children. It is important to note, however, that, of the 145 children with morally defective parents, only 16 percent were themselves behavior problems. Environment, therefore, must have been an influential factor in the good behavior of the remaining 84 percent who also had defective parents.

In 12 of the 13 serious behavior cases some information was available concerning the parents. The facts regarding these parents and the types of misbehavior exhibited by their children are shown below.

Parents	*Child*
1. Father a day laborer. Worthless. Deserted his family. In jail for wife abandonment. Mother weak in mind and body. In state insane hospital.	1. Girl. I.Q. 86. Steals. Lies. Defiant of authority. Stubborn. Boisterous.
2. Father a carpenter. Deserted his family. Mother of a bad reputation. Accomplice in a robbery. Poor housekeeper. Left school after 4th grade. Two sons are delinquent; one in Parental School, other under supervision of the Juvenile Court.	2. Girl. I.Q. 96. Steals money. Deceitful. Destructive of property. Mean-spirited. Unreliable.
3. Same as No. 2.	3. Girl. I.Q. 78. Sibling to No. 2. In same foster home. Same difficulties. More stubborn and sullen.
4. Nothing known about the father. Mother was murdered in a fight.	4. Boy. I.Q. 58. Steals. Stays out nights. Problem in school. Pranks. County judge has considered committing him to reform school.
5. Father dead. Mother a domestic. Two sons are in reform school.	5. Boy. I.Q. 101. Steals. Not dependable. Sullen. Mischievous.
6. Same as No. 5.	6. Boy. I.Q. 106. Sibling to No. 5. In same foster home. Runs away for several days at a time. Plays truant from school. Sullen. Mean disposition.
7. Father a day laborer. Deserted his family. Mother immoral. Imprisoned for making threat to kill.	7. Girl. I.Q. 69. Unmanageable. Impudent. Quick-tempered. Lazy. Unsocial. Strikes parents.
8. Father a day laborer. Deserted his family. Heavy drinker. One-eighth Indian. Mother depraved. Unfit to care for her children.	8. Boy. I.Q. 82. Steals money. Lies. A fighter. Disliked by boys. Mean disposition. Has threatened his foster mother's life.
9. Father in jail for bootlegging. One-half Indian. Mother immoral. Home dirty. General moral tone of home bad.	9. Girl. I.Q. 71. Steals. Lies. Masturbates. Walks streets. Unmanageable.

10. Father a day laborer. Drinks. Neglects his family. Mother died from overwork. Father married a young girl.	10. Girl. I.Q. 81. Impudent. Defiant of authority. Lazy. Vulgar. Man-crazy.
11. Father a miner. Drinks. Mother feeble-minded. Immoral. Neglects children.	11. Girl. I.Q. 91. Steals at school. Lies. Masturbates. Restless. Clownish.
12. Father dead. Mother "not very bright. Broken down or insane."	12. Boy. I.Q. 75. Steals. Stays out all night. Unmanageable. Quick-tempered. Struck his foster mother. Sent to state reform school at request of foster parents.

It is evident that in nearly every case there is some evidence of moral defect on the part of the parents. It seems fair to conclude, therefore, that heredity has some share in the misconduct of these 12 children.

IX. Summary

The main problem of this investigation was the measurement of the effect of environment upon the intelligence of foster children. The detailed procedure by which this measurement was made has been described in the foregoing pages. This summary includes only a brief statement of the more important comparisons made and the conclusions which seem to be warranted.

1. A group of children were tested before placement and then retested after several years of residence in a foster home. A comparison of their ratings on the two tests gave evidence of a significant improvement in intelligence (as measured by intelligence test scores). A study of certain sub-groups showed that the children in the better foster homes gained considerably more than did those in the poorer homes. Furthermore, the children who were tested and adopted at an early age gained more than those adopted at a later age. These facts appear to indicate that an improvement in environment produces a gain in intelligence.

2. A comparison was made between the intelligence of siblings (brothers and sisters) who had been reared in different foster homes. The correlation between their intelligence was found to be lower than that usually found for siblings raised together. The usual coefficient of fraternal resemblance is about .50, but it was found that for siblings separated before either of the pair was six

years of age the correlation was only .25. When the comparison was made for those whose foster homes were of different grade, the correlation was found to be only .19. These facts make it appear that a part of the resemblance between siblings reared together is due to the influence of a similar environment.

3. A group of siblings was divided into two groups by putting into one group the member of each pair who was in the better foster home and into the other group the one in the poorer home. The mean I.Q. of the group in the poorer homes was found to be 86, while that of those in the better homes was 95. An analysis of the conditions of adoption made it seem unlikely that there was any marked tendency for the brighter member of a pair of siblings to be taken into the better foster home. A random formation of two groups from pairs of siblings would give groups of equal intelligence. The superior intelligence of the siblings in the better homes appears, therefore, to give evidence that the character of the home affects the child's intelligence to a marked degree.

4. Two unrelated children reared in the same home were found to resemble one another in intelligence. The correlations between the intelligence of such unrelated pairs ranged from .25 to .37. This resemblance is probably due for the most part to the similarity of their environment.

5. The available information on the own parents of the foster children indicated that a large percentage were of defective mentality. If heredity were the only factor in the determination of intelligence, it would be expected that their children would be decidedly below the average. It was found, however, that their mean I.Q. was practically equal to the standard for children in general. Only 3.7 percent rated below 70, and these were those placed at relatively late ages. These facts seem to point quite clearly to the influence of environment upon intelligence.

6. In the case of 26 children studied, both parents were rated as feeble-minded. If intelligence were inherited according to the Mendelian law, all of these children would be feeble-minded. It was found, however, that only four had an I.Q. below 70 and these only slightly below. The average I.Q. of 81 for these 26 children is higher than would be expected according to the Mendelian law, but is considerably below that of the entire group of children

studied. These facts appear to indicate that heredity and environment are both influential factors in the development of intelligence.

7. The school progress of the children studied compares very favorably with that of the children in several large school systems.

8. In various groups comparisons were made between the intelligence of the children and the grade of foster home in which they had been reared. For the entire group of 401, the correlation between home rating and intelligence was found to be .48. The correlation between the intelligence of the children and the intelligence of their foster parents as measured by the Otis test was found to be .37. In the case of the children who had been tested before adoption, an initial correlation of .34 with home rating was raised to .52 after a period of residence in the foster home. These correlations would indicate that the character of the home is an important factor in the development of the child's intelligence.

9. The influence of the home is further shown by the fact that there is a correlation between early placement and intelligence and a slight relationship between the child's intelligence and the length of time he has spent in the foster home.

10. A large percentage of the children studied had parents who were morally defective. In spite of this poor heredity, however, few cases of serious misbehavior were found among the foster children. It seems probable, therefore, that environment has been an important factor in determining their conduct.

11. In interpreting certain data of the study it was necessary to know whether the apparent effect of good home environment could be accounted for by a selection of initially bright children by superior foster parents. The analysis of the data led to the conclusion that selection was not a large factor in the relationships. In the extreme cases in which the children were given a mental test before adoption, the correlation between this initial I.Q. and home rating was .34. In over eighty percent of the cases no test was made before adoption. Furthermore, a survey of the circumstances of adoption indicates that the intelligence of the child is not usually taken into account as a major consideration. Finally, for certain groups in which it would have been least possible to estimate the intelligence of the child the correlation with home rating is as high as for other groups.

X. Appendix

There are exhibited herewith samples of certain materials used in this investigation and referred to in preceding sections. These samples include the vocabulary test, and forms for collecting information on the material environment, characteristics of the parents, ratings of the parents, and characteristics of the child. No attempt has been made to conserve the exact size of type or spacings of the original materials.

Two specimen correlation tables are also exhibited.

1. Vocabulary Test

Name . . . .

Illustrations of what you are to do:

beast	means the same as	afraid . . . bird . . . animal . . . large
active	means the same as	joyful . . . lively . . . actual . . . strong
confidence	means the same as	power . . . misery . . . secret . . . trust

annual	capable . . . usual . . . yearly . . . new
beneficial	courageous . . . truthful . . . helpful . . . efficient
nourishment	refusal . . . food . . . novelty . . . gardening
executive	administrative . . . national . . . pertinent . . . elective
ponderous	surly . . . weighty . . . magnificent . . . clumsy
enumerate	increase . . . count . . . pronounce . . . explain
alteration	quarrel . . . succession . . . worship . . . modification
convalescent	assemblage . . . recovering . . . coiled . . . reminiscent
extinct	quenched . . . flavoring . . . happy . . . large
ambiguity	yellowness . . . aspiration . . . civility . . . uncertainty
eminent	impending . . . foreigner . . . renowned . . . kind
salutation	charity . . . greeting . . . deliverance . . . cleanliness
chaotic	confused . . . pure . . . old . . . dangerous
respiration	end . . . sweating . . . prayer . . . breathing
humidity	comedy . . . moisture . . . fear . . . mankind
aesthetic	drug . . . austere . . . sportmanslike . . . beautiful
obnoxious	offensive . . . obscure . . . forgetful . . . stubborn
hypothesis	pretender . . . insanity . . . supposition . . . trance
aspiration	weariness . . . ambition . . . breathing . . . closing
soliloquy	exaggeration . . . hardness . . . monologue . . . anxiety

corroborate	purify . . . improve . . . rust . . . verify
precocious	advanced . . . dangerous . . . comical . . . valuable
dilate	predict . . . weaken . . . enlarge . . . tell
asphyxiation	burning . . . repairing . . . operation . . . suffocation
prosaic	old . . . poetical . . . commonplace . . . interesting
genealogy	literature . . . pedigree . . . liberality . . . pleasure
excise	pardon . . . surplus . . . exertion . . . duty
metamorphosis	deadening . . . transformation . . . metallic . . . separation
inundate	count . . . violate . . . dig . . . flood
extant	degree . . . old . . . existing . . . stretch
enigma	secret . . . plot . . . falsehood . . . riddle
obelisk	jewel . . . pillar . . . corpulent . . . rectangular
facetious	jocular . . . sophisticated . . . critical . . . imaginary
incinerate	imply . . . cut . . . name . . . burn
condiment	quality . . . seasoning . . . terms . . . flattery
bisque	chance . . . bread . . . smoke . . . soup
heresy is a term used most in	botany . . . cookery . . . religion . . . medicine
depreciation	poetry . . . music . . . surgery . . . finance
etching	business . . . art . . . medicine . . . music
allegory	literature . . . physiology . . . surveying . . . law
rarebit	painting . . . grammar . . . athletics . . . cookery
libel	music . . . law . . . botany . . . forestry
adagio	architecture . . . religion . . . medicine . . . music
perihelion	athletics . . . law . . . astronomy . . . poetry
tapestry is a	food . . . cloth . . . statue . . . science
vermillion	sickness . . . color . . . herb . . . animal
artichoke	cloth . . . vegetable . . . animal . . . vine
vitamine refers to	laws . . . religion . . . food. . . . flowers
deciduous	money . . . trees . . . laws . . . weather
corolla	crowns . . . pictures . . . flowers . . . engines

2. Material Environment

Residence—rural town city population—
house apartment owned rented

Distance from town— from school— from church—

1. Neighborhood—good medium poor
2. Building and Grounds

Number of rooms—large medium small

Number in family—(children— adults—)
Grounds about home—large medium small none
Porch—large medium small none
Light and ventilation—ample fair inadequate

3. Furnishings—luxurious good adequate poor
Selected and arranged—tastefully indifferently in bad taste
Comforts and conveniences—furnace electric lights bath running water sleeping porch telephone automobile ( )

4. Care and Upkeep
Interior—very clean and orderly satisfactory dirty and disorderly
Exterior—well kept in fair condition neglected

5. Evidences of Culture
Library—large medium small
Books—best mediocre harmful
Magazines
Newspapers—
Pictures—best ordinary poor
Music—best ordinary cheap
Radio Victrola piano other musical instruments

3. The Parents

	Husband	Wife
Nationality—		
Age (in 1925)—		
1. Occupations. Present—		
Former—		
2. Education		
3. Social Activity		
Church activities—		
Lodges and clubs—		
Offices and positions—		

4. Use of Leisure Time—lectures musicales plays movies church entertainments parks auto trips visiting reading sports—
Interests and Hobbies—

Remarks:
Reason for taking child—
Basis of selection—looks health brightness disposition likeableness other reasons—

4. Ratings of Parents

Congeniality and Harmony—
Cöoperation in Training of Child—

	Husband	Wife
Intelligence—		
Disposition—		

Refinement—
Speech—
Personal Appearance—
Kindliness—
Personal Interest in Child—
Discipline—

Remarks:

Test Scores

Vocabulary Test—
Otis Intelligence Test—

5. The Child

Name   Foster Parents
Born——————   To Society——————   Taken into home——————

1. School Achievement
   Age of entering—
   Grade (1924-25)—   Name of school—
   Scholastic standing—
   Progress—rapid   normal   slow
   Attendance—regular   irregular
   Grades repeated—   Grades skipped—
   Subjects most easily mastered—
   Subjects that cause difficulty—
   Interest in school work—
   Study habits—
2. Extra-School Educational Advantages
   Parents' encouragement and stimulus in school achievement—
   Special training by parents   by others
   Reading to child   telling stories   helping with lessons
   music   drawing   dancing
3. Interests and Talents
   Special talents—
   Interests—
   Amount and type of reading—
   Recreation—
4. Social Activity
   Associates—
   Attendance at church meetings—
   Activities in other organizations—
5. Behavior   Home   School
   Special problems—
   Attitude toward parents—
   Attitude toward teacher—
   Attitude toward playmates—
   Amount of punishment—
   Type of punishment—

Remarks:

Work done by child—

Changes in child since coming into foster home—

Special efforts to secure improvement—

Scores in Previous Intelligence Tests

	Date	*C.A.*	*M.A.*	*I.Q.*
Stanford-Binet Tests—				
International Test—				

Information concerning child's own parents:

The two following correlation tables give examples of the detailed relationships which are represented by the coefficients of correlation reported in the preceding discussion. The correlation between the child's intelligence quotient and the rating of the foster home is .48 ± .03, and the correlation between the intelligence of siblings who have been separated four years or more is .25 ± .06.

Relation Between Intelligence of Siblings Who Have Been Separated Four Years or More

(Double Entry Used.)

Intelligence Quotient	Intelligence Quotient															
	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–94	95–99	100–104	105–109	110–114	115–119	120–124	125–129	Total
125–129		1		2				1								4
120–124																
115–119					1		1	1		3	2	1				9
110–114					1	1	4	2	1	3	1	4	1			18
105–109				1		1	1	2	2	2		1	2			12
100–104	1		1	1	3	4	5	5	3		2	3	3			31
95– 99			2	2	1		4	4	4	3	2	1				23
90– 94				3		1	2	6	4	5	2	2	1		1	27
85– 89	1	1	4	2	2	4		2	4	5	1	4	1			31
80– 84			1	3	7	2	4	1		4	1	1				24
75– 79			1	4	6	7	2		1	3		1	1			26
70– 74		1	1	4	4	3	2	3	2	1	1				2	24
65– 69				1	1	1	4		2	1						10
60– 64	2	2		1			1								1	7
55– 59		2					1			1						4
Total ..	4	7	10	24	26	24	31	27	23	31	12	18	9		4	250

RELATION BETWEEN CHILD'S INTELLIGENCE AND THE RATING OF HIS FOSTER HOME (Home Group)

Child's Intelligence Quotient	Foster Home Rating																									Total
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
160–164																							1			1
155–159																										
150–154																										
145–149																								1		1
140–144																										
135–139																				1		1				2
130–134																						1				1
125–129											1		2		3			1	1	2		1			1	12
120–124									1				1	1		1	2						2	1	1	10
115–119							1		1	1			2	4	1	1		3	2	3	2		3		3	27
110–114		1							3	1	1	3		2	3	1	3	3	2	1	1	3	2	2		32
105–109							2	1	2	2		3	3	2	3	5	1	3	2	1	3	3		2	1	39
100–104				1		1	2	1	6	3	2	5	5	2	6	3	1	6	1	3	1	2	2	1	2	56
95– 99	1			1	1			1	4	5	6	4	3	4	2	2	1	1	3	1	3	1	2	1		47
90– 94			1		3	2	6	2	3	1	4	3	2	3	1	3		4	2	4	1	2	1			48
85– 89		3			2		5	2	3	1	1	4	3	2	4		2	1		1	1			1		36
80– 84				2		1	4	1	1	1	5	5	3	4		1						2				30
75– 79						1	7	1	1	5	2	2			2			1	1	1	1					25
70– 74			1		3	1	1	2	4	2	1	3	1													19
65– 69				1			1	2		1		1			1											7
60– 64			1	1	1					1		1			1											6
55– 59												1		1												2
Total.....	1	4	3	6	10	6	29	13	29	24	23	35	25	25	27	17	10	23	14	18	13	16	13	9	8	401

CHAPTER X

THE RELATIVE INFLUENCE OF NATURE AND NURTURE UPON MENTAL DEVELOPMENT; A COMPARATIVE STUDY OF FOSTER PARENT-FOSTER CHILD RESEMBLANCE AND TRUE PARENT-TRUE CHILD RESEMBLANCE [1]

Barbara Stoddard Burks
Stanford University

I. Introductory

To what extent are ordinary differences in mental level due to nature and to what extent are they due to nurture?

Few scientific problems have been the subject of so much speculation and controversy as the specific one with which this study deals. This is probably attributable to two facts: the practical and theoretical significance of the problem itself, and the extreme difficulty of gathering evidence which cannot be applied with more or less plausability to the support of either the nature or the nurture hypothesis.

The result has been that, since the appearance in 1869 of Galton's *Hereditary Genius*, nearly every study published in the field

[1] *Acknowledgments.* This study was made possible by a grant from the Commonwealth Fund, and by supplementary grants from the Thomas Welton Stanford Fund for Psychological Research, and Mr. Max Rosenberg.

A study of the type here presented necessarily depends to a large degree upon coöperative effort. Grateful acknowledgment is made to Miss Mary E. Brusie, representing the Native Sons and Native Daughters Central Committee on Homeless Children, and to Mrs. Jessie Jordan, formerly Superintendent of the Children's Home Society of California, both of whom appreciated the need for the study when it was first projected and did all they could to make it possible.

The writer is also greatly indebted to the superintendents, principals, and their representatives in the following school systems and schools for their kindness in making accessible certain files and other information: Berkeley, Hawthorne and Whittier schools; Glendale, Eugene Field school; Los Angeles, Foshay Junior High, Main St., Normandy, Twenty-Eighth St., Santa Barbara, Western Ave., and West Third St. schools; Palo Alto, Walter Hays school; San Francisco, Agassiz, Argonne, Grant, Harrison, Hawthorne, Horace Mann, Jefferson, Madison, Pacific Heights, Peabody, Rochambeau, Roosevelt, Sunnyside schools.

To Miss Florence Bathgate and Miss Zena O'Connor, who for many months cheerfully executed exacting field work; and to Mr. Albert

has been seized upon by both the hereditarians and the environmentalists and interpreted as favorable to the point of view of their own school. The high incidence of genius in certain strains and of feeble-mindedness in others; the consistent decrease in familial correlation coefficients for psychical traits as more and more remote degrees of relationship are considered; the stronger resemblance found between twins than that between ordinary siblings; the approximate constancy of intelligence level when measured in the same individuals over intervals of time; the marked differences in average intelligence which have never failed to appear in studies of groups who were dissimilar in racial, occupational, educational, or social status—all these phenomena might conceivably be due either to hereditary or to environmental forces or to both at once. Hence, none of them offers evidence which can be regarded as crucial.

However, the fact that no experiment has yet served to uncover in unambiguous terms the relative contributions of ordinary nature and nurture differences to differences in mental level does not invalidate the experimental data which have been accumulating for nearly sixty years. In revealing tendencies, suggesting profitable

---

Kurtz, who performed the major portion of the computations and otherwise assisted in the statistical work, a message of warm gratitude is conveyed. To the hundreds of families who served as subjects, giving their time freely to a program that could offer them nothing of personal gain, the deepest appreciation is expressed.

The writer is also indebted to Professor Kelley, whose book on *Statistical Method* served as an important reference in the application of statistical treatment, and who generously took the time to derive for us a formula for inferring the reliability of an abbreviated test when the reliability of a complete test is known.

The following individuals also made valuable contributions of time and thought: Mrs. Annie Baird, Elizabeth Briggs, Mrs. Frances W. Burks, Mrs. Laura Crawford, Jack Dunlap, Mrs. Dortha Jensen, Theresa E. McCarthy, Dr. Maud A. Merrill, Catherine E. Moriarity, Mary E. Mullin, Pauline M. Nott, Ada Swortzel.

Finally, it was Professor Terman's large contribution in time, constructive criticism, and guidance that "saw the study through." Led by his encouragement, the writer's first impulse to undertake the study became an actuality. He secured the funds by which it was carried on, located the field workers who assisted in gathering the data, and kept in closest touch with every plan and every development. If the study should conform in a measure to his principles of scientific moderation and intellectual honesty, may he feel in some degree repaid.—*B. S. B.*

avenues of approach, and in actually indicating the existence of the problem itself, their contribution has been very great. Furthermore, they provide a substantial structure of fact which requires merely the impact of one additional experimental step to become invested with definite meaning. This additional step, by some means not yet employed, must isolate the effects of heredity and environment; then, in the light of our new knowledge, the wide factual background of previous data will assume its proper scientific perspective.

This step might be accomplished in one of two ways. The first would require the experimental control of either nature or nurture so that the effects of one or the other could be singly observed. The second would require the application of mathematical techniques to data collected in the ordinary way.

The second method offers us more than the first, ultimately, for in addition to isolating nature-nurture effects, it promises to yield an explanation of the actual mechanics underlying mental heredity. A start has been made towards developing such a scheme of mathematical analysis in the work of R. A. Fisher (3),[2] who has already applied his techniques to the study of inherited physical traits with extremely interesting results. However, the method he employed will probably be inapplicable to the study of mental traits without important modifications, and is consequently at present out of question for students of mental heredity. This leaves the first, or empirical, method as the only possible attack at present.

The investigation in hand approaches the aspect of the problem which concerns heredity and *home environment* through a comparison of mental test resemblances obtaining between parents and their children on the one hand, with those obtaining between foster parents and their foster children on the other. Thus, it seeks to evaluate the effects of nature and of home nurture through a study of two kinds of familial resemblance, one of which is dependent upon nurture influence alone, and the other upon a combination of both nature and nurture influences. Through its use of foster parents and their foster children as subjects, it applies to its purpose the end results of the social experimentation which is going on in many homes all about us.

[2] Numbers in parentheses following mention of contributors refer to the numbered bibliography at the end of this chapter.

It should be emphasized at this point that whatever tendencies and conclusions can be found in this study are valid only for populations as homogeneous in racial extraction, social standards, and educational opportunities as that from which our subjects are drawn. The distribution of homes of the children studied in this investigation was probably nearly as variable in essential features[8] as homes of the general American white population (though somewhat skewed toward a superior level). It was not as variable, however, as if the homes of southern negroes, poor mountain whites, or Philippine Negritoes had been included; and consequently, home environment cannot be expected to have as large a proportional effect upon the mental differences of the children we studied as though they were being reared in families unselected as to race or geographical location throughout the world.

Reference should also be made to the educational opportunities of the children examined, which were good. (All children were living in California communities.) If the children had varied considerably in educational opportunity, so that a number of them had as limited amount of schooling as that, for example, of Gordon's English canal-boat children, and if, in addition, home environment and educational opportunity had been correlated, it would have been quite difficult to separate the effects of the two upon the mental variability of our children. In this study, not only is the possible complication of differences in educational opportunity averted, but the confusing issue of possible cumulative effects of schooling is averted as well, since the measuring instrument used—the Stanford-Binet test of intelligence—was standardized upon California school children covering the same age range as our children, who themselves had undergone a cumulative educational process.

Other factors causing real or apparent impairment in mental ability, such as language handicap, deafness, pathological trauma (as from spastic birth paralysis, lethargic encephalitis or other diseases leaving permanent mental deficiency) were also ruled out.

Thus, the study is based upon children homogeneous as to race and educational opportunity; sufficiently homogeneous in health

[8] This seems probable because the variability in intelligence of both the control and foster children coming from these homes is as large as that of unselected children.

and physique to avoid confusion; and about as variable in hereditary endowment and in home environment (including kindred social mores) as white children of ordinary communities.

The study does not purport to demonstrate what proportions of the *total* mental development of an individual are due to heredity and to environment. Biologists have frequently pointed out the futility of attempting such a demonstration, since *any development whatever would be impossible without the contributions of both nature and nurture.* But if we direct our attention to the contributions of ordinary differences in heredity and ordinary differences in environment to *mental differences* (*i. e.*, I. Q. variance), it is possible to draw some significant conclusions. The causes which affect human differences, rather than the causes which condition the absolute developmental level of the human species have, after all, the more vital bearing upon social and educational problems.

Given a group of school children such as our subjects (which surely are representative of the largest single element in the American juvenile population), it will later be seen that the data gathered in this investigation lead to the conclusion that *about 17 percent of the variability of intelligence is due to differences in home environment.* It will further appear that the best estimate the data afford of the extreme degree to which the most favorable home environment may enhance the I. Q., or the least favorable environment depress it, is about 20 I. Q. points. This amount is larger, no doubt, than some of the firmest believers in heredity would have anticipated, but smaller than the effects often attributed to nurture by holders of an extreme environmentalist's view. To the writer, these results constitute an important vindication of the potency of home environment. But even more significant appear to be the implications of these basic results, *e. g.*, that *not far from 70 percent of ordinary white school children have intelligence that deviates less than 6 I. Q. points up or down from what they would have if all children were raised in a standard (average) home environment;* that, while home environment in rare, extreme cases may account for as much as 20 points of increment above the expected, or congenital, level, heredity (in conjunction with environment) may account in some instances for increments above the level of the generality which are five times as large (100 points).

The methods by which such conclusions as these are reached are set forth in detail in VI (Main Results) and VIII (Interpretation and Conclusions). The reader who wishes to bear in mind a general picture of the type of treatment to which the data were subjected as he reads the preliminary chapters upon selection of subjects, methods devised, field work schedules, etc., is invited to turn to VI and VIII at once, reserving the preliminary sections till afterward.

II. Historical

1. Previous Studies

In addition to the present study, there are several investigations in this field which have attempted to go beyond the limits of ordinary biometry. Besides a few reports of the effect of environmental changes upon single individuals, there are at least three comprehensive studies which should be mentioned in this connection. The first is Dr. Gordon's investigation of orphanage siblings (4). It would appear to be significant that the siblings in a standardized institutional atmosphere which she tested on the Stanford-Binet showed an average correlation similar to that usually found among sibling pairs reared under ordinary home influence (about .50). But unfortunately, the study is rendered somewhat ambiguous by the racial heterogeneity of the children and by the fact that the sib pairs had lived together for unknown, varying lengths of time before being admitted to an orphanage.

Miss Theis' report, "How Foster Children Turn Out" (17), completed for the State Charities Aid Association of New York in 1924, yields facts some of which have a bearing on our problem. The report deals with 910 children who were placed in foster homes by the State Charities Aid, and who were eighteen years old or over at the time of the investigation. The data, collected by well-conceived family-study schedules, include much valuable information concerning the outcome of foster home care and are naturally of intense interest to child-placement agencies. No test data were reported, but Miss Theis and her staff of field workers secured information for each former ward upon his "capability to manage himself and his affairs with ordinary prudence," and his "ability to take formal education." These two criteria showed no significant relationship to the cultural level of the homes in which the children had been placed, but classifying the homes as to the *type of care* that had been given the children, the author is led to the following significant conclusion: "Of the children who had excellent care, approximately 87 percent developed into 'capable' subjects; of those who had average care, 80 percent are 'capable'; and of those who had poor care, 66 percent are 'capable.' "

Miss Theis' definition of "capability" implies moral as well as intellectual stamina, and consequently these results cannot be carried over directly to the problem with which we are engaged. Moreover, her subjects varied greatly in age at time of foster-home placement—the average age was eight, and not more than 35 percent were under five. This fact greatly complicates interpretation from the nature-nurture standpoint. As the report frankly states in the introduction (by Homer Folks): "We cannot disentangle the factor of inheritance from that of early life with the children's own parents and the environment provided by them."

The third investigation is that by Gertrude Hildreth on "Resemblances of Siblings in Intelligence and Achievement" (7), published in 1925. Miss Hildreth compared the resemblances found in several groups of siblings who had been tested by the Stanford-Binet. She found that siblings who were being reared in a large New York Hebrew orphanage (which admits children at the age of five or over), and that siblings who had been separated from each other for an average interval of four years, resembled each other to about the same degree as siblings being reared in their own homes.

These data strongly suggest an interpretation of resemblance based upon inheritance, but are admitted by their compiler to be inconclusive. The selective influences at work were unsusceptible of measurement, and the possibility that home and parental influences during the early formative life of the orphanage siblings may have accounted for at least part of their later resemblance was not ruled out.

Evidently, an investigation in which all possible disturbing factors have been eliminated, or recognized and allowed for, has been a pressing need. As far back as 1913, Richardson (13) criticized Pearson for his unproved assumption that the correlations found in his biometric studies represented uncontaminated hereditary coefficients. He suggested the desirability of making a study of poor children adopted into rich homes to measure the potency of environment. Poyer, in his book on *Problèmes Généreaux de l Hérédité Psychologique* (12) made the suggestion that, to separate the influence of heredity and environment, it would be necessary that, for a long series of generations, every child in a family should be separated from his relatives and placed in an entirely different environment, without contact with his parents. . . . This sort of experiment never has been made and never will be made.''

It is not easy to see why Poyer thought that to study one family over a number of generations would be preferable to the study of two generations of many families, but his statement of the necessity for the type of approach used in the present investigation is clear and compelling. Doubtless, our type of study would have been undertaken long ago by some one of the many who must have sensed its need if the practical difficulties connected with it had not appeared prohibitive.

2. Formulation of the Project

The practical difficulties just referred to were of such a nature that it seemed desirable to proceed cautiously, and indeed to stand ready to abandon the entire project if it promised not to be feasible. It was necessary to gain the consent of child-placement organizations to select from their files cases that would meet the requirements laid down for the investigation. Such consent could be obtained only after convincing the executive committees of the need for the kind of data we proposed to gather, and after assuring them that the identity of the families visited would be held strictly confidential, and would be quite submerged in the statistical treatment of results.

Once permission had been secured to visit the foster children with the approval of the organizations that had placed them, each family case still remained an unknown term, which might react favorably or unfavorably to the request we planned to make. As all our cases were *adopted* foster children, and hence had passed outside the supervision of the placement agencies, there was no pressure that could serve to encourage coöperation from the foster parents except an appeal (through a letter from the agencies which was carried by the field visitors) to whatever friendly gratitude they might feel toward the organization that had given them their child, and an effort to arouse a broad humanitarian interest in a study designed to give us a "better understanding of children's development." Usually, too, there was a fear on the part of the foster parents to overcome, because in at least 65 percent of our cases the child had never been told that he was a foster child. In these cases the foster parents nearly always hesitated about co-operating until they were fully assured that nothing in the field visitor's program could possibly cause the child to suspect the fact of his adoption.

During the first months of field work it was also necessary to recognize that the program was on trial with the child-placement agencies, that permission to proceed had been granted only provisionally and would have to be withdrawn if any considerable number of the foster parents were found to resent what might appear to them an intrusion.

Accordingly, in the fall of 1923, only a small, tentative beginning was made. Results accumulated slowly, as the center of our foster child population was several hours journey from Stanford and I was able to give only occasional week-ends to field work. However, by the end of the summer of 1924 data from approximately twenty cases had been gathered, and not more than one or two families visited had declined to coöperate. The feasibility of the method seemed well established, and the child placement agencies definitely committed themselves to the program. The time appeared to be ripe to seek funds for securing two or three full-time field assistants who would make it possible to carry out the study on a wider scale and to gather data for a control group of true parents and true children.

In the fall of 1924 Professor Terman applied for, and received, a subvention from the Commonwealth Fund, which was supplemented by a gift from Mr. Max Rosenberg of San Francisco, and by considerable financial assistance from Stanford University.

Immediate steps were then taken to obtain the services of two trained field workers. We were fortunate in securing Miss Zena C. O'Connor, a candidate for the doctorate at Teachers College, Columbia;

and Miss Florence Bathgate, who had completed two years of graduate work in psychology at the University of California. Both assistants had had practice in the use of the Stanford-Binet scale, both had specialized in the field of educational psychology, and both demonstrated a rare tact and skill in meeting the exacting human demands of the field visiting. Miss O'Connor was engaged in field work for fifteen months, and Miss Bathgate for twelve months. The writer also, working part time, spent the approximate equivalent of nine months in the field. The data from about thirty percent of the foster cases, and fifteen percent of the control cases, were collected by her.

III. Methods Employed

1. Approach

The program for family study required four to eight hours of a field worker's time per family. Much of the testing and interviewing had to be done at night to suit the hours when the fathers of the children could be at home.

It was our invariable rule to make no first approaches by telephone, as it seemed probable that our chances of gaining the interest and coöperation of families would be far better by personal interview. Consequently, much time was lost in attempted calls when the family were out, away from town, moved, etc. This condition, linked with the wide areas it was necessary to travel, and the difficulty of dove-tailing appointments at unusual hours with any degree of efficiency, resulted in slower progress than we had at first contemplated. From two to three completed cases weekly was the ordinary average per field worker.

2. Schedule

The items of our family case schedule were these:

1. Stanford Binet Test, administered to parents and children.
2. A home-information blank, containing an adaptation of the Whittier Scale for Home Grading and a culture scale of our own, filled out by field assistants.
3. Rating of the child on ten character and temperament traits made independently by the two parents.
4. Personal information blank filled out by each parent.
5. Woodworth-Cady questionnaire (to test emotional stability) filled out by children ten years old or over.
6. Information was also obtained from the files of the placement agenices, in the case of the foster group. This included heredity (if known), age at placement, age at adoption, national descent, etc.

The Stanford Binet Test and record booklet are so well known as to require no description here.

The nature of the Whittier Scale and the culture scale are made evident in the section later on wherein the scoring standards for these scales are described.

The ten traits upon which the parents rated their children were: (1) will power and perseverance; (2) cheerfulness and optimism; (3) musical appreciation; (4) sense of humor; (5) permanency of moods; (6) leadership; (7) sympathy and tenderness; (8) conscientiousness; (9) originality; (10) general intelligence. The traits were selected from a large number of traits used in connection with the Stanford study of gifted children; and ratings were made upon a seven-category graphic rating scale, as reproduced in *Genetic Studies of Genius,* I (15).

The personal information blank filled out by each parent called for data upon the following points: birthplace; occupation; highest school grade reached; special interests, hobbies or accomplishments; positions of honor, trust, or recognition which have been held; distribution of time during the day (at home or away from home); children's hobbies or interests; occupations which parents think may be suitable for child in future; where child spends his leisure time; discipline of child. In addition, the blank filled out by the mother asked for information upon the kind and amount of home reading done by the child at various ages; the home instruction or attention received by the child in such matters as reading or writing, story-telling to child, number work, or nature study; and the private tutoring received by the child (in music, dancing, or other subjects).

The Woodworth-Cady questionnaire—reproduced in full in *Genetic Studies of Genius,* I, pp. 500 ff. (15)—is a questionnaire of 85 questions designed to sift out psychotic tendencies. A number of questions are inserted as 'padding' to lull the suspicions of the subject as to the purpose of the test. Samples of the questions are:

"Do your teachers generally treat you right?"
"Did you ever have a nickname you didn't like very well?"
"Are you happy most of the time?"

3. Administration and Scoring

The Stanford Binet Test was administered in the standard way, with two exceptions as to procedure.

The testing was carried down only to a level at which all tests but one were passed, and up to a level at which all tests but one were failed. This modification, known as the "lopped Binet," often saves fifteen to thirty minutes testing time, but has the effect of slightly lowering the reliability. The lowered reliability was later allowed for in correcting coefficients of correlation for attenuation.

The other exception consisted of slight changes in the wording of the directions of some of the tests on the lower levels when the test was administered to adults (but not when it was administered to children). These were accomplished, we believe, in such a way as to retain the psychological significance of the test intact. They were made only because a small proportion of the parents proved to have intellectual pretentions beyond their mental level, and occasionally showed irritation when it was necessary to carry the test down to items specifically framed for children of immature experience. Asking a man or woman, for example, who had earned and expended money for years: "If I were to buy 4 cents worth of candy and should give the storekeeper 10 cents, how much money would I get back?" was likely to cause a little self-consciousness in the subject. Instead, we asked: "If you bought 4 cents worth of something and gave the storekeeper 10 cents, how much money would you get back?" Slight though our changes were, they can be defended only on a policy of expediency in the absence of a test of the Binet type standardized upon adults.

In scoring the Binet Test, allowance was made for the fact that a few of the brightest adults exceeded the limits of the test, by applying a correction worked out in connection with the Stanford study of gifted children (15). As described in *Genetic Studies of Genius*, I, p. 42, the method represents an attempt to correct the scores "to correspond to what they would have been had the scale been more nearly adequate in the upper range." It is based upon the average scatter of successes in levels above the basal mental age, and involves "the following additions of months to the mental age score for those passing various numbers of tests out of the total of twelve tests in year-groups 16 and 18:"

Tests passed in 16 and 18..	5	6	7	8	9	10	11	12
Number of months to add..	3	6	9	12	15	18	21	24

The scoring of the Whittier Scale for Home Grading was based upon the standards published in the Whittier Manual (18). The items of the scale were rated from 1 to 5 with the exception of *Necessities, Parental Conditions,* and *Parental Supervision,* which were rated from 1 to 6 because the Whittier standard of 5 did not seem adequate for a number of our cases. The total Whittier index is the sum of the ratings on the five individual items. In the case of our culture scale, all five items were rated from 1 to 5. The total culture index is again the sum of the individual ratings.

Brief descriptions of the types of conditions which were assigned the different numerical ratings are presented herewith. In the case of the Whittier items, the descriptions are taken verbatim out of the Whittier Manual (18). The scoring descriptions represent the standards agreed upon by the field workers at the time their visiting began.

Scoring of Whittier Home Scale

1. Necessities.

1 point. Wages of driver of small express and transfer wagon. Old ragged dirty clothes. Little food, very plain. Three small rooms in basement of cheap tenement house. Hardly bare necessities. Old, cheap, broken, wooden chairs and tables. No pictures or decorations. Bare floors. No comforts or improvements.

2 points. Income of painter, work somewhat irregular. Cheap, plain clothing, small frame bungalow, bare necessities, plain wooden furniture, no pictures, cheapest fixtures for light and cooking.

3 points. Wages of pressman in tile factory. Sufficient food and clothing. Small frame house poorly finished both outside and inside. Furnishings sufficient in quantity, plain, serviceable, cheap. Cheap pictures, calendars, and family photographs for ornaments. No carpets or rugs. Oil stove. Electric lights, no ornamental fixtures.

4 points. Income, salary and tips of head waiter in a large hotel. Clothing neat, well-kept, apparently made to last. Good table set. Half modern bungalow. Furniture good quality, plentiful. Wicker and reed chairs, piano, rugs, good pictures. Rather poor lighting from windows, but modern electric fixtures. Running water, modern sanitary conveniences. Rear porch bedroom, couch in living room.

5 points. Architect, well-to-do. Well-dressed. Table ware indicates abundant food. Large modern bungalow, frame construction, well finished. Furniture fine quality, plentiful. Fine carpets, rugs and

pictures. Modern conveniences, built-in cupboards, electric fixtures, plumbing.

6 points. Conspicuously superior to the level receiving 5 points. Seldom given to any home. Denotes unusually luxurious living conditions.

2. Neatness.

1 point. Everything filthy dirty. Empty cans and trash in rooms. Absolutely no order or arrangement. Food, dirty dishes, etc., piled any old way on table and chairs, falling on floor. Small, crowded area way, dirty and bare. No lawns or flowers. Buildings bare and need paint. No attention to home.

2 points. Rooms and furniture dirty. Part of furniture stacked in corner. No arrangement for the rest. Yard rather disorderly. Lawn grown up and poorly cared for. Few evidences of attention to the home.

3 points. Rooms fairly clean. Little or no order. Yard bare and dirty with a few tumble down sheds. House in need of paint, steps and porch need repairing. Lawn not kept clean, little attention given to external appearances.

4 points. Rooms clean, but dark, closed and stuffy most of the time. Furniture neatly arranged and kept in good order. Exterior cleanliness good. House somewhat in need of paint. Lawn well-kept. Considerable attention given to home when possible.

5 points. Interior clean and sanitary. Furniture neatly arranged, good order. Yard and grounds clean, no outbuildings. House well-kept. Yard clipped close, small, neat garden. General neatness good. Considerable attention apparently given to care of home.

3. Size.

1 point. Two very small rooms in basement of three-story tenement. Cook, eat and 'live' in one room. Propositus, two other children and parents. No yard, only a small area way.

2 points. Five-room bungalow. Medium sized rooms. Tent in backyard. Convenient, but not sufficient for so large a family. Propositus, five brothers and sisters, ages 2 to 16, mother and father.

3 points. Small three-room house. Medium sized rooms. Rather close quarters. Propositus, mother and step father. Very small unkept yard.

4 points. Seven rooms, all rather small. Two-story house. Rooms convenient, although small. Propositus, three younger children, mother and step father.

5 points. Seven rooms, two-story house. Good sized rooms. Plenty of room conveniently arranged. Two adults, father and mother, propositus and younger sister. Rather small front yard, good open porch. Large back yard as city yards go.

4. Parental Conditions

1 point. Mother makes small living by housework and washing. Parents separated, father away with another woman, mother not remarried. Mother at home evenings and some days.

2 points. Parents both probably normal. Father blacksmith for a large ranch. Continual trouble when parents together. Parents separated, alleged immoral conduct on each side. Boy officially with father, who was at work all day. Boy left much alone.

3 points. Father planing mill hand, fairly successful when working, but an invalid from excessive drinking. Mother probably normal. Harmony most of time except when father's drinking interferes. Either father or mother work out during day, mother when she can get work, father when in good enough condition.

4 points. Father a painter, in good health. Mother probably normal. Harmonious most of the time. Mother nags father some on account of irregular work. No separation. Father away at work during day.

5 points. Father normal, has average success as carpenter. Mother keeps home in fair condition. So far as known, there is harmony between the parents. Mother at home all of the time, father away at work most of day. (In practice we never assigned a rating as high as 5 to this item if either parent tested with a mental age below 12-0).

6 points. Conspicuously superior to the level receiving 5 points. Both parents superior on the mental test, and *exceptionally* harmonious in their relations.

5. Parental Supervision

1 point. Little interest from either parent. Practically no supervision. Father a consumptive, drives an old wagon. Mother epileptic. Boy left to run streets day and night.

2 points. Lack of discipline because parents away large proportion of time. No partiality (in dealing with children) as far as known. Parents good, hard-working people. Spent evenings on streets.

3 points. Mother not well. Father interested but lazy. Discipline lax. Evidence of favoritism and better treatment of girls. Father lazy and easy going, but of good habits.

4 points. Father apparently interested in welfare of boys. Fairly good control. Equally fair treatment as far as known. Father a colored preacher. Good habits and reputation.

5 points. Parents interested in health, education and welfare of children. Kind and intelligent discipline. Complete fairness as far as known. Parents of good reputation and character, good example to children. Children kept at home evenings as a general rule.

6 points. Care given the children and provision made for their welfare very exceptional.

Scoring of Culture Scale

1. Speech

1 point. Average vocabulary level of parents less than twenty-five (on Binet vocabulary test); or average level of parents less than fifty *and* a less-than-average mean rating of parents on the other items of 'speech.'

2 points. Average vocabulary level of parents less than sixty and most of other items of 'speech' not above average. Or average vocabulary level of parents higher than sixty but lower than seventy, and most of other items below average.

3 points. Average level of parents' vocabulary between sixty and seventy and most of other items average. Or average vocabulary level of parents somewhat above seventy, but other items chiefly below average.

4 points. Average level of parents' vocabulary above seventy and other items average or above. Or average level of parents' vocabulary far above seventy, but other items average and below.

5 points. Average level of parents' vocabulary eighty or above and practically all the other items above average.

2. Education of Parents

1 point. Average of parents' schooling is 1, 2, or 3 grades.
2 points. Average of parents' schooling is 4, 5, or 6 grades.
3 points. Average of parents' schooling is 7, 8, or 9 grades.
4 points. Average of parents' schooling is 10, 11, or 12 grades.
5 points. Average of parents' schooling is above 12th grade.

Schooling, as used above, refers only to academic schooling. Business college and nurses' training, etc., are not counted in computing average grade completed, but are given weight in border-line cases (where the parental average lies between two of the class groups above). When the schooling for only one parent is known, this item is not rated.

3. Interests of Parents

There are five steps to describe the level of both father's and mother's interests. The classes underlined on the culture scale are determined by reference to the parents' information blanks, which ask for special interests and hobbies, and for positions of honor and trust which have been held by the parents. In general, we try to assign a numerical rating to "interests of parents" which would correspond to an average of the ratings which we assign to each parent. When only one parent has filled out an information blank, we use the field worker's estimate of the other parent. When the field worker has not rated the parent who filled out no information blank, we assign a rating to the item as a whole based upon just one parent.

1 point. No interests are noted for either parent.

2 points. Interests are few and of an unintellectual sort—reading the funny paper, going to movies.

3 points. Ordinary pursuits, sewing, gardening, etc.

4 points. More intellectual pursuits, administrative positions of a minor sort such as president of a lodge or P.T.A., a moderate interest in literature or art, etc.

5 points. Very high type of interests. Important contributions of time to public welfare, some intellectual pursuit or study (aside from mere gaining of a livelihood).

4. Home Library

These ratings are based chiefly upon the number of books in the home library (exclusive of number of books in child's library, quality of books, number and kind of magazines). But in cases lying near the borderline of the classes described, number of books in child's library, quality of books, and number and quality of magazines are decisive factors in determining the rating. As a general guide:

1 point. Less than 10 books.
2 points. Less than 50 books.
3 points. Less than 150 books.
4 points. Less than 500 books.
5 points. 500 books or over.

5. Artistic Taste

1 point. No musical instrument, no pictures or only those of the most inartistic type, no taste in furnishing of house.

2 points. Possibly a musical instrument, but furnishings and pictures distinctly inharmonious, and trashy ornaments, such as kewpies and gaudy bric-a-brac scattered about.

3 points. Rather nondescript, nothing offensive to the eye, usually some kind of musical instrument, though the Victrola records and piano music are not of a high type. Pictures are cheap prints poorly framed. Photographs of the family are usually abundant.

4 points. Noticeable effort to make home somewhat beautiful. Rugs and hangings blend well. Nearly always a musical instrument provided with fairly good music. Furniture and pictures do not clash, no trashy ornaments about, and family 'photos' are absent or present in very moderate numbers.

5 points. Effect of interior is beautiful. Fine taste used in blending the colors of rugs, hangings, furnishings, and pictures (though these things are not necessarily of the more expensive sort). Musical selections for piano or Victrola are from standard composers (though a little popular music or jazz may be included as well). Ornaments selected with discrimination and well arranged.

IV. Selection, Location, and Coöperation of Cases

1. Foster Group

a. Criteria of selection. The following criteria were satisfied in selecting cases for the study from the files of the Native Sons and Native Daughters of the Golden West Central Committee on Homeless Children, and the Children's Home Society of California:

1. Children were placed in their foster homes before the age of 12 months. (The average age of placement of our group proved to be 3 months, 2 days.)

2. Children were legally adopted—not merely cared for in free or boarding homes.

3. Children were between 5 and 14 years, inclusive, at the time of the investigation.

4. Foster parents were white, non-Jewish, English-speaking, and American, British, or north-European-born.

5. True parents (so far as was definitely known) were white, non-Jewish, Americans, British, or north-Europeans.

6. Children were placed in the home of a married couple, both members of which were alive and living together at the time of the investigation.

7. Cases must be accessible to the three centers—San Francisco Bay region, Los Angeles, and San Diego.

These criteria require little discussion. The first one was laid down (1) to insure that each child had lived in the environment of a single home from early infancy, and (2) to avoid the type of selective placement that might easily have been exercised if the children had been old enough at the time of their adoption to give clear evidence of their mental potentiality.

The second criterion confined the study to children who were being reared as though they were the actual offspring of the foster parents.

The third confined it to children within a range for which the I. Q. is fairly comparable at all ages.

The fourth and fifth criteria enabled us to avoid the confusion in results that would ensue from a foreign language handicap in any of the subjects who were tested, and precluded the possibility of an adventitious resemblance between foster parents and children due to the practice by placement societies of matching foster parents and children for racial descent.

The sixth insured that all the children should be homogeneous in having both a paternal and a maternal influence; and the seventh merely made the study administratively feasible.

The criteria were met practically to the letter. When returns for the field work were in, several case records were found which failed to meet all the criteria and had to be eliminated. Such failure could always be accounted for in one of three ways:

1. Incompleteness of the abbreviated file from while the prospective cases were noted (*i.e.*, omission of some essential item regarding nationality, age of placement, etc.)

2. Overlooking on the part of the field workers of some item of information which should have ruled out the case in question from the start.

3. Change of condition described by the sixth criterion since the Society's record of the case was closed.

Cases of the following types were omitted from the calculations and summaries which are presented:

Case a—Foreign language spoken in the home.

Case b—Child was over a year old when placed.

Case c—Foster father deserted several years previously.

Case d—Child whose true father was Mexican was placed with a couple of which the wife was Spanish-American.

Case e—Child was only 4-9 when tested.

Three cases were kept, however, in which the age of placement of the child was afterward found to be in excess of 12 months, but to such a small extent that their inclusion did not seem detrimental to the study. These were:

Case f—Child was 14 mos. 20 days, when placed.

Case g—Child was 13 mos. 30 days, when placed.

Case h—Child was probably 15 mos. 17 days, when placed (record vague).

Since, in case h, we have no tests of the foster parents or supplementary environmental data, the I.Q. of this child has been used only in the I.Q. distribution and nowhere else.

In addition to cases a to e, it was necessary to delete:

Cases i and j because the Binet tests of the children concerned were below our standard of completeness;

Case k because the child concerned had had infantile paralysis which, according to the foster parents, had left her with a secondary feeble-mindedness;

Case l—only the test of the foster father discarded, because he was French Canadian and had a distinct language handicap despite years of residence in this country;

Case m—only the test of the foster father discarded because he was too deaf to take it under standard conditions;

Cases n and o—only the tests of the foster fathers discarded because they fell below our standard of completeness;

Cases p and q—only the tests of the foster mothers discarded for the same reason as that in cases n and o.

A few cases were kept, however, in which tests did not fully meet our standard of completeness, but fell below the standard by such a small amount that it did not seem expedient to deplete our material by eliminating them. Our testing procedure, it will be recalled, employed the complete Stanford-Binet Scale, but required that the testing be carried down only to an age level at which all items but one were passed, and up to an age level at which all items but one were failed. In a few instances it happened that, owing to limitations upon the time that the individuals tested were able or willing to give, the field visitors resorted to the expedient of administering only the abbreviated form of the test (*i.e.*, the starred items) or to carrying the testing only to the levels at which *two* items were failed or passed. Such abbreviations were unauthorized, but were made in situations in which it appeared to be a choice between obtaining an incomplete test or none at all. In no case did we retain in the study tests employing the starred items which did not extend down to a level at which *all* items were passed and up to a level at which *all* items were failed. In no case did we retain tests in which the 'lopping' of required passes or failures was so severe as to provide a basic level having more than two failures or an upper level having more than two passes, nor tests which had been lopped to this degree at both the upper and the lower level. However, the following tests were retained which were incomplete, but to an extent *less* than that described immediately above: two Binet tests of foster fathers, seven Binet tests of foster mothers, two Binet tests of foster children.

In scoring these tests, credit was given each test element on a basis of total credit allowed for a given mental-age level, divided by the number of elements administered at that level. In the few cases of tests retained which were lopped slightly more than our standard permitted, levels below the lowest level administered were counted as passed; levels above the highest level administered were counted as failed, and no effort was made to apply a correction. This no doubt introduced slight errors into the scores in question, but as it is very rare for a subject to gain or penalize himself by as much as 6 or even 4 months on levels beyond the boundaries of such lopping, the errors on these few tests can hardly have disturbed our results for the entire group to an appreciable degree.

In addition to the cases mentioned above, a case was also retained in which the foster child was found afterward to be a few days below our age standard, *i.e.*, 4 years, 11 months, 8 days; and one case was retained in which the foster parents had separated, but had done so only a few months before the tests were made of the foster mother and child.

The foster parents in all cases for whom we have accurate data met the criterion regarding nationality, but the data regarding the true parents of the foster children are sometimes less clear upon this point. The records of the societies, as a rule, noted in connection with the case histories, not the true parents' birthplace, but only their *descent,* which referred sometimes to their native country, sometimes to the nationality of one or both of their parents, and occasionally to the nationality of one or more of their more remote progenitors. Which of these possibilities happened to be the case was seldom stated. Of the 140 true fathers and 146 true mothers for whom we could obtain a record of "descent," 64 percent and 66 percent, respectively, were indexed as of "American" descent; but 5 percent and 7 percent, respectively, were recorded as descended from national stocks which we sought to eliminate through our criterion of selection. In the group of true fathers and mothers combined 2 were of Greek descent, 6 of Italian, 2 of Mexican, 1 of Portugese, and 7 of Spanish.

It would have meant a serious loss if it had been necessary to discard each of these cases which were inadvertently included either because of an oversight on the part of the field workers or because of an ambiguity in the records from which cases were listed. Since the purpose of the nationality criterion was to avoid selective placement based upon racial descent, and since it was probable that a number of the cases above were born of diluted, rather than of pure, racial strains, it was decided to retain them if it could be shown that their inclusion had no significant tendency to alter the results.

Accordingly, we computed correlations between the children's I.Q.'s and the mental ages of the foster parents with the cases in question left in the group (with the exception of case d, deleted for the reason previously given), and again with the cases in question taken out. Taking out 12 foster-father and foster-child pairs reduced the correlation by .034; taking out 14 foster-mother and foster-child pairs increased the correlation by .006. As these amounts were slight and in opposite directions, the cases in question were retained. In view of the fact that the group contained no south-European or non-Caucasian foster parents, even though the true parents of a few of the children were south-Europeans; and that only the placement of a child of a given nationality or race with foster parents of the same nationality or race could give rise to the ambiguous results against which we were guarding in this connection, the retention of these cases seems quite justified.

b. Location and coöperation. To secure figures on proportion of total cases listed which were located, which gave full coöperation, which refused to coöperate, etc., the original lists were consulted and each case assigned to one of the categories in the table below. 'Outside' cases not located through the organizations (of which we had 12) were, of course, not included. A small group of cases listed after the study was under way, of which the original work sheets could not later be found, was not included either. Nevertheless, the following figures are undoubtedly representative for the group as a whole and are based upon the majority of the cases listed.

TABLE I.—LOCATION AND COÖPERATION OF CASES

	Listed	Not Located	Located	Full Coöperation	Partial Coöperation	Refusals	Unavailable
Northern California	265	127	138	80	10	15	33
Southern California	372	249	123	78	6	8	31
Total	637	376	261	158	16	23	64

Of 637 cases, 261, or 41 percent, were located; of 261 located cases, 64, or 24 percent, proved to be unavailable for reasons listed below; of 195 available cases, 158, or 80 percent, gave full coöperation, and 16, or 8 percent, gave partial coöperation, yielding 174, or 88 percent, who coöperated to the extent of giving us tests on at least one parent and child. (In our entire group we had 204 cases which coöperated to that degree.)

The 12 'outside' cases previously mentioned were, in the main, foster siblings of our cases, and had been secured by the foster parents through sources other than the two placement agencies with which we worked. They included one or two 'volunteer' cases as well.

TABLE II.—REASONS FOR UNAVAILABILITY

	Number of cases
One foster parent dead	15
Foster parents separated or divorced	10
"No time"	10
Foster child died	8
Sickness in family	4
Home in inaccessible region	3
Field workers told by organization secretaries not to visit case	3
Part of family away	2
Mother works	2
Deafness of foster parent	2
Child had been returned to organization because of feeble-mindedness	1
Possible secondary feeble-mindedness of foster child	1
Child was feeble-minded and in an institution	1
Test of child was invalidated by having foster parents come in and talk during test	1
Case was known by field visitor to be too confidential to approach	1
Total	64

2. Control Group

a. Function of Control Group. The function of the array of families comprising parents and their true offspring which we have termed the 'Control Group' should be clearly defined at this point. The group serves two significant ends. It permits an estimate of the strength of mental *heredity,* after the strength of environment has been evaluated in the foster group; and it furnishes a most important check upon the validity of methods used.

Regarding the first end, more will be said in the final section. With regard to the second end, it is easy to see how indispensable the Control Group really is. If our test data and environmental data had been obtained only for the Foster Group alone, the low 'environmental' correlations reported in a subsequent section of this study could not have been said with any assurance to represent the actual limits of the type of influence we sought to measure. The unanswerable criticism could have been made that our methods might simply be unadapted to measuring the force of environment, and that better methods in the future might contravert our findings. But here we have a control group for which data were gathered by the same field workers and by the same procedure as that employed for the Foster Group. In marked contrast to the results for the Foster Group, we shall find in the Control Group that parent-child mental resemblances are of about the same magnitude as those ordinarily found in the case of hereditary *physical* traits. Such results will offer a solid basis upon which to interpret results.

The point was made in an earlier section that our conclusions are valid (at least in numerical terms) only for populations resembling in important respects the ones tested. It follows that if the results from the Foster Group and the Control Group are to furnish a valid comparison, the two groups must be 'matched' in a very rigid sense with respect to all the factors that could directly or even remotely influence the results. The effort which we made to secure such matching is manifest in the criteria that follow for selecting the Control Group. That the criteria were successful in attaining good matching will be seen in fifth section, headed "Composition of Groups."

b. Criteria of selection. Control cases were chosen by the following criteria:

1. Children of the Control Group were matched with those of the Foster Group for age, sex, and number of five-year-olds who had had no kindergarten attendance.

2. Control families were matched with foster families for locality, type of neighborhood, and occupational field of the father.

3. Non-Jewish, white American, British, or north-European families who spoke English were taken.

4. Both parents were alive and living together at the time of the investigation.

5. Only one child per family was tested (though families were not selected with respect to size).

Only 50 percent as many control cases as foster cases were selected, since our resources and time were becoming limited. As it turned out, the correlations in the Control Group, despite smaller numbers, had no greater probable errors for the most part than the corresponding correlations for the Foster Group, because the control correlations were in general so much larger.

As a possible source of cases, we considered the advisability of seeking the coöperation of parents who had an application for a foster child pending with one of the California child-placement agencies, and at the same time had a true child. Several such cases were looked up, who gave coöperation readily, but it soon became evident that to secure enough cases through this source would require the field visitors to travel prohibitive distances. Accordingly, it was arranged to select cases from the files of the public schools in the general localities represented by the Foster Group. The listing of subjects was as impartial as we were able to make it. At each school coöperating, the procedure was:

1. Select two or three names under each letter in the alphabet to avoid siblings.
2. Take about twelve cases for each age group.
3. Ask principal to check list for divorces, step children, race, etc.
4. Ascertain occupations of fathers from the children.
5. Locate a few children who have five-year-old siblings who have never attended kindergarten.
6. Check list with teachers.
7. Get letter of introduction from principal to parents.

This scheme gave us a working list far greater than the total number of cases we intended to gather, and allowed us sufficient leeway to insure good matching of controls with fosters in all the criteria laid down.

The Control Group, as finally completed, consisted of:

Cases selected through public schools	93
Cases of 'true children' in our foster families who were tested while our foster records were being compiled	6
Cases of 'true children' of parents with applications pending for a foster child	3
Other cases of families interested in project	3
	105

In the Foster Group all the tests were made by the three field workers mentioned, with the exception of one 'outside' case contributed by Miss Elizabeth Briggs, who was doing field work on another Stanford project, and a test of one foster child contributed by a Berkeley school. In the Control Group tests which were not made by one of the three field workers were administered as follows:

Tests of children in Palo Alto, by students trained by Dr. Merrill, of Stanford University	4
Test of child contributed by a school	1
	5

It will be recalled that several cases of the Foster Group for which we had considerable data had to be dropped because of failing to meet our criteria. Only one such case had to be discarded in the Control Group (a case in which the test of the child was too incomplete to keep). The Control Group, however, had a larger quota than the Foster Group of tests which fell below our standard of completeness, but were retained nevertheless because they fell below to so slight a degree. Of these there were 11 tests of fathers, 12 tests of mothers, and 9 tests of children.

c. *Location and coöperation.* With the Control Group, in which certain delicate factors peculiar to the Foster Group were absent, the computed figures for full and partial coöperation were 94 percent and 98 percent. These figures, while undoubtedly representing a greater readiness to coöperate than was found in the Foster Group, are probably a little high, since in the control field work the visitors sometimes abandoned a case on the list if, after a call or two, it promised to consume an unusual amount of time to find the family at home, arrange for an appointment to give tests, etc. It is likely that some of the cases which were thus discarded might have proved to be 'refusals' if they had been pressed.

TABLE III.—COMPARISON OF NUMBERS AND LOCALITIES IN BOTH GROUPS

	Foster	Control
Los Angeles (including Glendale and Pasadena)	83	46
San Francisco	43	37
Oakland	29	—
Berkeley	15	2
Rural (in Southern California)	11	3
San Francisco Bay cities	10	13
Alameda	9	3
San Diego	8	1
Southern California beach towns	6	—
	214	105

The failure of the Control Group to meet its 50 percent quota in some of the localities is accounted for by the fact that we selected in certain large centers, as Los Angeles and San Francisco, not only school districts which would match the districts in which we had located foster children in those cities, but also districts which would match those in nearby towns from which some of our foster cases had been drawn. Thus, we matched most of our San Diego and 'beach towns' quota in similar Los Angeles districts, our Oakland quota in similar San Francisco districts, and our Berkeley and 'bay cities' quota in Palo Alto (itself a 'bay city').

A final count of Binet tests given in the entire study (aside from those deleted for reasons explained) gives the figures shown in Table IV.

TABLE IV.—TOTAL NUMBER OF SUBJECTS TESTED BY STANFORD-BINET

	Foster	Control
Children	214	105
Parents (counting only once the foster parents of 22 pairs of foster sibs)	342	206
Total	556	311
Foster Group plus Control Group (minus 12 parent tests appearing in both groups)	855	

V. VALIDITY OF RESULTS

The extent to which the results of this investigation can be considered valid and conclusive is dependent upon two aspects of the study which should now be discussed. The first aspect concerns the soundness of the data themselves; the second aspect concerns the possible presence or absence of disturbing factors that would tend to render the results ambiguous or cover up relationships which we sought to evaluate.

1. Soundness of the Data

An attempt has been made to apply objective criteria to an appraisal of this point.

(1) One check involved the comparison of tests made by the different field workers. Although each of the three workers was trained in Binet procedure, each had had her training at a different institution, and it was conceivable that a personal equation might have entered to distort the results. The following comparisons were therefore made of average I.Q.'s of children tested in the Bay region and in Los Angeles county. The slight differences found do not appear to be significant.

	Bay Region Mean I.Q.	N	Los Angeles County Mean I.Q.	N
Bathgate	———	—	109.2	39
Burks	103.4	65	———	—
O'Connor	106.7	33	109.4	50
Difference	3.3		0.2	
Difference divided by P.E. of difference	1.6		0.1	

(2) Careful precautions were taken in gathering data, in scoring blanks, and in working up the results statistically, to insure accuracy. The most important scatter diagrams and correlation coefficients were checked by having them repeated independently.[4] The writer checked for errors in the scoring of the Binet tests made by all three field workers from the verbatim responses noted in the record booklets. Dr. Maud A. Merrill of Stanford kindly triple-checked twelve of the Binet blanks, so that we might have further assurance that the scoring was standard. In no case did any suggestion of hers regarding scoring amount to an appreciable change in mental age for the test in question. The chronological age and the I.Q. (M.A./C.A. ratio) were checked for the test of every child. (M.A.'s, not I.Q.'s, were used in the case of the adults.) The writer also rescored the Whittier and culture items on each home information blank in order that all the cases might have ratings made by a single standard.

(3) Finally, the test scores were considered from the standpoint of chronological age. It has been shown by Willoughby (19) and by others that in group tests of the Alpha or Beta type there is a negative correlation between performance and chronological age after an age of approximately thirty. Willoughby's data do not indicate whether this relationship is due to the selection of his subjects, to decreased speed of reaction, or to a decline in the actual functions tested. Material of the Binet type would not seem *a priori* to be so subject to age decline as material used in

[4] In the Foster Group the following correlation scatters and computations were checked: child's I.Q. with M.A. of foster father, with M.A. of foster mother, and with mid-parent M.A. In the Control Group the following correlation scatters and computations were checked: child's I.Q. with M.A. of father, with mid-parent M.A., with Whittier index, with culture index, and with family income. These were found to have been computed correctly the first time.

group tests. However, this point was investigated by computing the correlations between chronological age and mental age for our present groups, and then partialling out the chronological age from the correlations between parents' and children's intelligence. The correlation between the children's I.Q.'s and chronological age was also computed, because inverse relations have been found by some investigators. The following correlations were found between intelligence and chronological age (intelligence measured as M.A. in adults and as I.Q. in children):

TABLE V.—CORRELATIONS OF INTELLIGENCE WITH CHRONOLOGICAL AGE

	Foster		Control	
	r	P. E.	r	P. E.
Fathers	−.23	.05	−.03	.07
Mothers	−.00	.05	.18	.07
Children	−.10	.05	.09	.07

Applying the partial correlation technique to free the parent-child correlations from the effect of age did not affect the correlations beyond one point in the second decimal place. Accordingly, the factor of age was not considered in the subsequent correlations involving intelligence.

2. Possible Disturbing Factors

There are three types of disturbing factors which might enter in to cloud the significance of the results. The first is possible selective placement of the foster children, which, if practiced to any considerable extent, would result in adventitious resemblances between foster parents and children. The resemblances would then be no less ambiguous than those between true parents and children which have been reported in various investigations.

The second type would result in a distorted distribution of the I.Q.'s of foster children, and might come about through a possible 'weeding out' of undesirable children, or through some unknown selective influence with respect to the type of child who comes under the guardianship of the two placement agencies coöperating with us. If such were the case, then any positive deviation of the average I.Q. of the Foster Group as a group from normal might be due both to good environment and to selection, in unknown proportions.

The third type is a possible failure upon our part to secure conditions in the Control Group which are truly comparable to those in the Foster Group. Only if the two groups are comparable in essential respects does the Control Group provide an adequate check upon the Foster Group.

3. Selective Placement

It is natural to suppose that an effort would be made by the child-placement agencies to fit the home to the child. Through interviews with the placement secretaries, we found this to be the case. A number of factors seem to have operated. Racial boundaries were sharply drawn; and the religious faith (whether Protestant, Catholic, or Jewish) of the child's true parents was always observed in selecting a foster home. National descent also figured to a considerable degree, and an attempt was usually made to have the child's general physique and coloring match those in the family of which he was to become a member. Finally, the agencies made an attempt to place the children with foster parents whose cultural background was somewhat similar to that of the true parents.

Through our criteria for listing prospective cases, we succeeded in eliminating some of the phases of selective placement mentioned above. By taking only children who had been placed before twelve months of age, we barred those in whom a clear prognosis of future development could have been made, and who could therefore be 'fitted' to homes directly upon such a basis. This criterion also insured an ordinarily constant environment for the children almost from the dawn of their intelligence, thus avoiding the confusion which would have been introduced if a significant part of their care and training had been given by their true parents. (It will be recalled that the average age of placement for our group was only three months.)

Our specifications regarding the true parents and foster parents of the children eliminated selective mental resemblance due to race or nationality, since only white subjects were used, and the only nationalities included were those which in other investigations have been shown to differ little in average level when measured on a large scale by mental tests.

The placement of children according to religious faith, physique, or coloring could scarcely give rise to spurious mental resemblance (especially since children of south-European or Jewish descent were excluded). That leaves us only the possibility of selection based upon cultural status to consider.

Investigation of the degree to which this factor was operative was approached in several ways.

First, the occupations of all the true fathers of our foster children for whom we had adequate information were assigned Barr[5] ratings by Mr. Kurtz and myself. Similar ratings were also assigned to the occupations of the foster fathers as listed *at the time application for a child was filed* with the placement agencies. Not all the occupations of fathers and foster fathers in our group were represented by the 100 occupations listed on the Barr Scale, but estimated Barr Scale values were assigned to the occupations not represented on the scale. The objectivity of the ratings was good; 29 random pairs of independent ratings of true fathers by Mr. Kurtz and myself correlated .92 ±.02, and 47 random pairs of our independent ratings of foster fathers' occupations (at time of applying for foster child) correlated .91 ± .02. But the correlations that were found between Barr ratings of the children's true fathers and certain other variables were exceedingly low. They may be summarized thus:

Barr Ratings of True Fathers	r	P.E.	N
With Barr ratings of foster father............	—.02	.07	86
With Whittier rating of foster home (at time of our investigation)....................	.01	.07	91
With culture rating of foster home (at time of our investigation)	—.04	.07	83
With child's I.Q. (at time of our investigation)	.07	.07	91

[5] The Barr Scale, described fully in *Genetic Studies of Genius,* I, pp. 66ff., comprises the combined judgments of thirty raters upon the grade of intelligence which each of 100 representative occupations demands on the average. Scores on the Barr Scale are expressed in point values, and vary from the occupation "hobo," which is arbitrarily given a rating of 0, through average occupations such as metal finisher or plasterer (8.02 and 8.04), to that of inventive genius (20.71).

These results were negative in so far as evidence for selective placement was concerned. However, this approach had taken account only of the slight information we had regarding the children's fathers and left their mothers out of the picture. We next attempted to get a better estimate of each child's prognosis as an infant by utilizing all the information we had upon the heredity of each. Utilizing the data copied from the files of the placement agencies, two raters, Mrs. Dortha Jensen and I, made independent estimates of the probable future I.Q. of a child having given each hereditary background. There were 158 cases for whom such estimates were made. Information regarding the remainder of our 214 cases (which included a number of foundlings and abandoned children) was too meagre to justify any estimate whatever. According to the fullness or paucity of data, the two raters indicated that their estimates of the expected I.Q.'s were based upon data that offered "certain," "fairly certain," "rather uncertain," or "very uncertain" grounds for estimate. The estimates were made without knowledge of which children were being judged and were consequently based solely upon our information regarding their parentage. Estimates recorded as "certain" were based upon data that gave the occupational or social status and sometimes the educational status of both parents. Estimates recorded as "very uncertain" were based upon information usually regarding only one parent and that only suggestive to a slight degree. Separating the children into groups for whom the estimates were certain, fairly certain, rather uncertain, and very uncertain, we obtained the correlations with the children's actual I.Q.'s, foster fathers' mental ages, and foster mothers' mental ages shown in Table VI.

The 158 pairs of ratings by the two raters correlated to the extent of .64 ±.03. This measure of reliability does not indicate, of course, how close to the actual I.Q.'s lie the estimates of fallible judges based upon fallible data; but rather how closely two fallible judges can agree as to the significance to be attached to the limited data available upon the children's true parentage. While the figure, .64, is high enough to show that the two raters approached the problem of estimating the I.Q.'s with something like the same point of view, it is not high enough to have much value for prediction (its alienation coefficient is .77). The average correlation of I.Q.'s estimated by the two judges with the mental ages of the foster fathers was only .01, with the mental ages of the foster mothers only .14, and with the actual I.Q.'s of the children only .18. The correlation between foster parents' intelligence and children's I.Q. to be expected from selection alone would be the product of

TABLE VI.—ESTIMATED I. Q. OF FOSTER CHILDREN CORRELATED WITH OTHER VARIABLES

Variable	Certainty of I. Q. Estimate				Total Ratings	Rater
	Certain	Fairly Certain	Rather Uncertain	Very Uncertain		
Actual I. Q.						
r	.36	.24	−.03	.23	.19	B
P. E.r	.12	.09	.08	.14	.05	
N	24	47	65	22	158	
Actual I. Q.						
r	−.04	.14	.23	.75	.18	J.
P. E.r	.19	.08	.08	.11	.05	
N	13	69	69	7	158	
Foster Father's M. A.						
r	−.02	−.22	−.03	−.29	−.01	B.
P. E.r	.15	.11	.09	.14	.06	
N	21	36	57	20	134	
r	.14	.01	.11	−.15	.03	J.
P. E.r	.18	.09	.09	.29	.06	
N	13	60	56	5	134	
Foster Mother's M. A.						
r	.47	.05	....	.48	.16	B.
P. E.r	.11	.10	.08	.11	.05	
N	22	46	64	21	153	
r	.27	−.02	.10	.73	.11	J.
P. E.r	.17	.08	.08	.12	.05	
N	13	67	66	7	153	

the correlation of estimated I.Q. with actual I.Q., times the correlation of estimated I.Q. with foster parents' mental age; namely, .00 in the case of the foster fathers and .02 in the case of the foster mothers. (Sewall Wright method; see reference 20.) Consequently, any correlations actually found in the study between foster parents and foster children can probably be considered as due to environment.[6]

When one considers the steps of judgment necessary in the routine of 'fitting a child to a home,' it is little wonder that the correlation coefficients of selective placement are negligible. A placement secretary, on the basis of rather meager information furnished by a mother who relinquishes her infant, estimates a child's mental potentiality. Our medium-high coefficient of reliability of estimate suggests that, if another placement secretary should make such an estimate, it would tend to vary somewhat from that of the first placement secretary. And our low correlations between estimated I.Q.'s and children's actual I.Q.'s suggest that the data available on the child's heredity are not adequate for any valid prognosis of his future mental development.

In choosing a suitable foster home for the child, the placement secretary has, to aid her, the report of a field agent who visits the homes of applicants for children, and usually one or two personal interviews with one or both of the prospective foster parents. From these she forms her estimate of their financial, moral, religious, and mental status. Our results indicate that her estimate of financial and moral status of the foster parents is probably more valid than that of their mental status.

With no means of accurately gauging the infant's mental potentiality, and with a fallible estimate of the mental status of the foster parents, the placement secretary now does her best to match the foster parents and child from as many angles as possible, including that of mental level. The result, as our data show, is an almost total lack of systematic correspondence between the child's mental potentiality and the mental level of his foster parents.

[6] It is possible, of course, that in some of the cases more information was obtained by the placement societies regarding the child's true parents than appeared in the case records on file. Through personal interviews with mothers relinquishing their children, the society representatives may have reached a better appraisal of the mother's capacity than was possible for us with the limited number of data recorded in the organization files. But it should be noted (Table VI) that even the cases for which the estimates seemed to be based upon reliable data tend to show low correlations between estimated I.Q. and foster parent intelligence. Also, the considerable number of children for whom the hereditary background is almost unknown would diminish still further the possible effects of selective placement throughout the total group.

Through the course of the investigation it became more and more apparent that *intellectual promise carries little weight with either the placement societies or prospective foster parents.* This implies that what knowledge the organizations possessed regarding the children's mental heredity was of little avail in making the placements. The following tabulations suggest, at least, that the foster parents themselves made little attempt to choose their foster children for their mental potentialities. The application blanks of both societies request the applicants to state the age, sex, and appearance of child desired, and leave ample space for description. It is true that the wording of the blank does not specifically include the item of mentality, and that this matter may later be

TABLE VII.—KIND OF CHILD DESIRED, AS SPECIFIED ON APPLICATION BLANKS

Trait or Quality Specified	Frequencies		
	N. S. N. D.* (108)	C. H. S.† (58)	Total (166)
Age	105	55	160
Sex	88	49	137
Complexion	33	19	52
Eye color	23	13	36
Hair color	22	10	32
Health	21	2	23
Nationality	6	2	8
Parentage	7	....	7
Mentality	3	1	4
"Normal"	3	....	3
"Good features"	2	1	3
Race	2	1	3
Legitimate orphan	2	....	2
Size	2	....	2
Refinement	1	1	2
"Attractive"	....	2	2
"Large eyes"	1	....	1
Curly hair	1	....	1
Twins	1	....	1
Number of Items Specified—			
1	9	7	16
2	35	22	57
3	38	17	55
4	18	7	25
5	3	4	7
6	4	....	4
7	1	1	2

*Native Sons and Native Daughters Central Committee on Homeless Children. †Children's Home Society.

taken up by applicants in personal interviews with the child-placing agents. Nevertheless, it seems significant that so few applicants made any mention of qualities other than the objective ones suggested on the blanks.

4. Possible Distortion of Intelligence Distributions

One's immediate observation on scanning the data upon the true parents of the foster children is: "This group seems to represent a fair cross-section of the population at large." Among the true mothers we find clerks, high-school girls, domestics, teachers, college girls, nurses; among the true fathers we find mechanics, lawyers, truck-drivers, students, salesmen, army and navy men. The classification in Table VIII is an attempt to give the clearest possible picture, consistent with brevity, of the vocational fields from which the true parents were drawn (in cases for which such information was available). The average age of the fathers was 29 years, of the mothers 23 years.

TABLE VIII.—OCCUPATIONAL DISTRIBUTION OF TRUE PARENTS OF FOSTER CHILDREN

	True Fathers		True Mothers	
	N	Percent	N	Percent
I. Professional (lawyers, engineers, etc.)	8	6.7	....	....
II. Teachers	....	....	2	2.6
III. Business owners or managers	8	6.7	....	....
IV. Commercial employees (clerks, bookkeepers, stenographers, etc.)	18	15.1	22	28.6
V. Sales work	9	7.6	....	....
VI. Rancher or farmer	6	5.0	....	....
VII. Skilled labor (carpenters, mechanics, etc.)	31	26.1	38 (VII–IX)	49.4 (VII–IX)
VIII. Semi-skilled (truck drivers, barbers, etc.)	6	5.0		
IX. Unskilled labor	8	6.7		
X. Students (college, high school, grammar school)	9	7.6	15	19.5
XI. Army and navy (probably not officers)	16	13.4	....	....
Total	119	99.9	77	100.1

Although the parents entering this tabulation represent only about half of the total number, they possibly offer a fair indication of the social level of the entire group. It will be recalled that almost no average difference was found between the I.Q.'s of foster children who were foundlings and of those whose ancestry was known. Fathers who do not enter the table were generally listed as "unknown," or "occupation unknown." Mothers who do not enter were more often listed as "at home."

In the case of the fathers, the proper class to assign to each occupation is generally clear, but in the case of the mothers, the three classes—skilled, semi-skilled, and unskilled labor—have been combined because of the difficulty of deciding to which one a given occupation belongs. Domestics are outstanding in this grouping—there were 16. Others there included are four factory workers, four telephone operators, six nurses, and two laundry workers.

The mean of the Barr ratings of 91 true fathers[7] for whom we had sufficient data to make a rating was 9.6 ± 0.3. The standard deviation was 3.1 ± 0.2. This mean value is slightly higher, and the standard deviation slightly less than the corresponding values for the adult male population at large in the localities from which the fathers were chiefly drawn (approximately 7.9 and 3.4).[8]

It is possible that true fathers whose occupational level could not be estimated would have either raised or depressed the average and standard deviation of the Barr ratings; but this seems unlikely, because the average I.Q. of children for whose fathers we have Barr ratings is within 1 point of the average for the entire Foster Group. It seems probable that the congenital level of offspring of the group of fathers and mothers that are represented in our study would not be far from average (or 100 I.Q.).

Other evidence, more or less indirect, was brought to bear upon this general question. The possibility was considered that whatever heterogeneity was present with respect to the source of foster children might, in spite of the apparently unselected distribution of occupations, introduce factors that would load the average level of the children to one side or the other. Are legitimate children,

[7] Many fathers for whom we had the general occupational class, such as "sales work," could not be rated on the Barr scale because there was no way of knowing whether their type of work ranked fairly high, as insurance salesmanship, or fairly low, as selling over the counter in a small shop. We made no attempt, either, to assign Barr ratings to students or to army and navy men.

[8] The latter values were computed in connection with the Stanford study of gifted children to compare with similar statistics upon the fathers of gifted children. See *Genetic Studies of Genius,* Vol. I, p. 71.

who because of the death or desertion of one or both parents, and because no relatives step forward to care for them, more likely or less likely to be subnormal than illegitimate children?

In our group (exclusive of 'outside' cases,) causes of dependency were found as shown in Table IX.

TABLE IX.—CAUSES OF DEPENDENCY

	N	Percent
Illegitimate	145	71.8
Foundlings (probably illegitimate)	14	6.9
	159	78.7
Legitimate, one of parents dead	10	4.9
Legitimate, parents separated or divorced	9	4.4
Legitimate, parents could not or would not provide for child	11	5.4
	30	14.7
Unknown or doubtful	13	6.4
Total	202	99.8

The discussion farther on of the significance of the average I.Q. level found in the entire group of foster children will not be anticipated by reporting at this point the mean I.Q.'s of children from these sub-groups. It is sufficient here to note that the only sub-group showing a significant departure in mean level from the mean of the entire group was the sixth (unknown or doubtful). This sub-group is small and is classified 'unknown or doubtful' because of incompleteness in the records of the placement organizations. It seems safe to assume that, for our purposes, the entire group of foster children can be treated as a homogeneous one.

Our next line of approach considered the possibility of intellectual differences among mothers of illegitimate children with respect to the type of young woman who relinquishes and the type who keeps her child. This approach was made through the coöperation of the matrons of two large 'rescue' homes in San Francisco which give assistance to pregnant girls. First, the plan was to consult the files of these homes for the years in which children in our foster group were born; but each home had recently come under new management, and the files prior to this (1921 in one home, and 1923 in the other home) were too incomplete to be of service. Accordingly, the 1921 files for the first home and the 1923 files for the second home were used to compile short case histories of the girls in these homes who had given birth to infants. These histories usually contained information upon the age, occupation, and education of each girl, and occasionally provided data upon her family background and comments upon her mental status. The disposition of the infant (whether adopted, relinquished, kept by mother, or deceased) was also noted. Fifty-seven such histories were compiled. The fifty-seven mothers were then ranked independently by Mrs. Jensen and by me in order of the probable intelligence level. The two series of rankings correlated (Spearman rank-order method) $.92 \pm .01$. The average rankings of *both* judges were then used to determine the percent of illegitimate mothers placing children for adoption who equalled or exceeded mothers of the entire group in estimated intelligence. This proved to be 59 percent, or 16 of 27, of mothers whose children were relinquished and placed for adoption.[9] This proportion suggests a slightly stronger tendency in the brighter than in the duller mothers to relinquish their illegitimate children, but, in the light of the small numbers and the unknown degree of validity of our rankings of the mothers, it is inconclusive.

Next, the files of the County Clerk's office in San Francisco were consulted with a view to estimating the likelihood that children received and placed by our two coöperating agencies were 'unselected' dependents. Adoption papers for one representative year (1922) were examined with respect to the age and source of the children adopted. Of 117 adoptions, 37 had been adoptions of children under two years of age.[10] Of these, only twelve had been relinquished through our two placement agencies; 11 had been obtained through other organizations (chiefly the Associated

[9] The mothers whose children were placed for adoption do not represent all the mothers who relinquished their children, for of course children who died or who became permanent wards of the state because of disease or obvious subnormality were not included in the group placed for adoption.

[10] While our study deals only with children placed under one year, most of the children were not legally adopted for six months to a year later. Accordingly, children were considered who were recorded as adopted under two years of age.

Charities); and 14 had been relinquished to the adoptive parents directly or through private individuals. For these 14, the following sources and causes of dependency were noted:

Mother died, father relinquished	5
Relinquished or abandoned by parents	5
Illegitimate, relinquished by mother	2
Father deserted, relinquished by mother	1
Received through a physician	1
	14

Since the causes for dependency in this group are similar to those found for our cases, and since (as was previously mentioned) no significant difference could be found in our group between the average I.Q. level developed by groups of children coming through such sources, there is little reason to believe that in intelligence they represent any more selection or any less selection than our cases.

Still another possibility of distortion considered is the veto sometimes exercised by the placement agencies with respect to the infants considered unplaceable for one reason or another. If an infant shows a positive Wasserman test or presents symptoms of any congenital disease, if either of his parents is known to be feeble-minded, or if the infant himself exhibits identifiable stigmata or symptoms of under-par development, he is not placed for adoption—at least not until he is cured or is old enough to be given diagnostic tests. Such a policy doubtless raises the average level of children who are placed in adoptive homes.

In San Francisco, the agency which makes provision for dependent children who are unplaceable is the Associated Charities. Miss Moriarity of this agency kindly coöperated with me in making an analysis of the illegitimate children under one year of age who met in all respects our criteria for the selection of foster cases and who had been helped by the Associated Charities during a one-year period (1921). There were 134 such children, 78 of whom were placed for adoption before they were twelve months old (generally through the Native Sons and Native Daughters Committee). Of the remaining 56, 6 died, and 50 were either not placed for adoption at all or were not placed until after reaching twelve months of age. Sufficient information was available upon 30 of the 50 who were not placed to provide some basis for estimating their probable I.Q. development. A sample case, for example, yielded the following data: "Mother feeble-minded, committed to Eldridge (state institution). Child lived in a boarding home until age of four, when she was pronounced normal by a psychologist (I.Q. 91) and placed for adoption." Information for this case is a little more complete than for most of the others, and it was only rarely that test data upon the children were available. Nevertheless, the conclusion is probably justified that the 30 children whose potential I.Q.'s were rated by me were on the average somewhat inferior in mental endowment. The average estimated I.Q. for these cases was 91.

Obviously, if from a fairly normal distribution of I.Q.'s, a number of cases are removed chiefly from the lower half, the effect is to raise the average level of cases remaining. It is likely that a distortion in this direction has occurred in our Foster Group. This possibility gains support from the fact that the distribution of I.Q.'s actually found in the Foster Group is somewhat skewed toward the upper end. It is impossible to estimate the precise effect upon I.Q. level that this factor has. There is at least the possibility that the *total* group of parents whose children become eligible for foster homes or for custodial care are themselves skewed toward the lower end of the intelligence range, but that through judicious elimination of unpromising candidates for adoption, such skewness fails to appear in the group of parents of our cases. Support for this view is found in the apparently representative distribution of the occupations of the true parents of our cases.

Finally, the possibility was investigated that selective coöperation of the foster parents approached might have introduced a source of error. Beside the cases refusing any coöperation whatever, there were cases in which (1) the foster parents permitted a test of the child but would not coöperate themselves; (2) the foster mother and foster child coöperated, but not the foster father; and (3) the foster father and foster child coöperated, but not the foster mother.

In order to apply our data upon these cases to the question of selective coöperation, it is necessary to assume that whatever selection as to mental level resulted from the total refusals of some families to coöperate is revealed at least to some extent in the cases which partially refused to coöperate.

(1) In five cases for which we have a test of the child, but not for either foster parent, the average I.Q. of the foster child was 109.

(2) In 24 cases in which the foster mother and child coöperated, but not the foster father, the average mental age of the foster mother was 15 – 8, and the average I.Q. of the foster child was 101.

(3) In two cases for which we have tests from the foster father and child, but not from the foster mother, the average mental age of foster father was 17 – 2, and the average I.Q. of the foster child was 92.

The mental levels of the foster parents and foster children in the main study are expressed by the mean values M.A. of foster fathers 17 — 0, M.A. of foster mothers 15 — 8, and I.Q. of foster children 107.

Since these values are close to the mean values just cited for groups (1) to (3), there seems to be little evidence that our refusals (approximately 15 percent) materially influenced the final results in so far as mean intelligence of cases is concerned.

An additional source of speculation, of course, is that there may have been a selective influence present in refusals that would actually alter the coefficients of resemblance between foster parents and children, even though it did not affect the mean levels of the tested groups. Such a disturbance would result, for example, if bright foster parents whose children happened to have turned out dull showed a stronger tendency than other foster parents to refuse to coöperate. We do not have data through which this possibility can be objectively ruled out, but we do have the subjective judgment of the field workers that refusals were more often connected with a sensitivity about the fact of the child's adoption than with any other sentiment. The judgment that refusals were not closely associated with discrepancies between parent and child intelligence is rather borne out if we recall that in the Control Group, in which certainly there were a number of cases of fairly wide discrepancy between parent and child intelligence, the number of refusals was almost nil.

As a matter of interest, to find out whether the group of foster parents serving as subjects were representative of the entire group of foster parents (meeting our criteria of selection) with whom children have been placed by our two agencies, the mean and standard deviation of Barr ratings of foster fathers were computed for four groups as shown in Table X.

According to these ratings, the families who were located represent a slightly superior selection.

TABLE X.—BARR RATINGS OF FOSTER FATHERS AT TIME OF APPLICATION

	Mean	S. D.	P. E. of mean	N
Cases coöperating	11.5	2.60	.13	195
Cases located*	10.6	2.65	.11	245
Cases unavailable	10.8	3.09	.30	51
Cases refusing to coöperate	11.6	2.36	.46	12

*Certain occupations are more migatory than others. For example, we have only one chauffeur located on the list of seven whom we attempted to locate. We located none of the three ministers on our list.

The interpretation to be drawn from the preceding pages must take the form of a judgment, rather than of an objective conclusion. A variety of evidence, suggestive rather than conclusive, has been presented with the hope that a reasonable assumption regarding the potential I.Q. distribution of the foster children may be indicated from the composite. The reader, on the basis of his own judgment, will doubtless infer, as I do, that the average I.Q. of our

TABLE XI.*—INTELLIGENCE DISTRIBUTION OF CHILDREN, IN I. Q.

	Foster	Control		Foster	Control
175–179	....	....	105–109	32	15
170–174	....	....	100–104	32	13
165–169	....	....	95– 99	27	5
160–164	1	....	90– 94	16	5
155–159	1	1	85– 89	7	....
150–154	1	2	80– 84	2	....
145–149	1	1	75– 79	3	2
140–144	....	2	70– 74	2	....
135–139	3	7	65– 69	1	....
130–134	7	3	60– 64	....	....
125–129	8	10	55– 59	....	....
120–124	16	13	50– 54	1	....
115–119	24	18	45– 49	....	....
110–114	28	8	40– 44	1	....
Mean				107.4	115.4
S. D.				15.09	15.13
N				214	105

*See also Figure I.

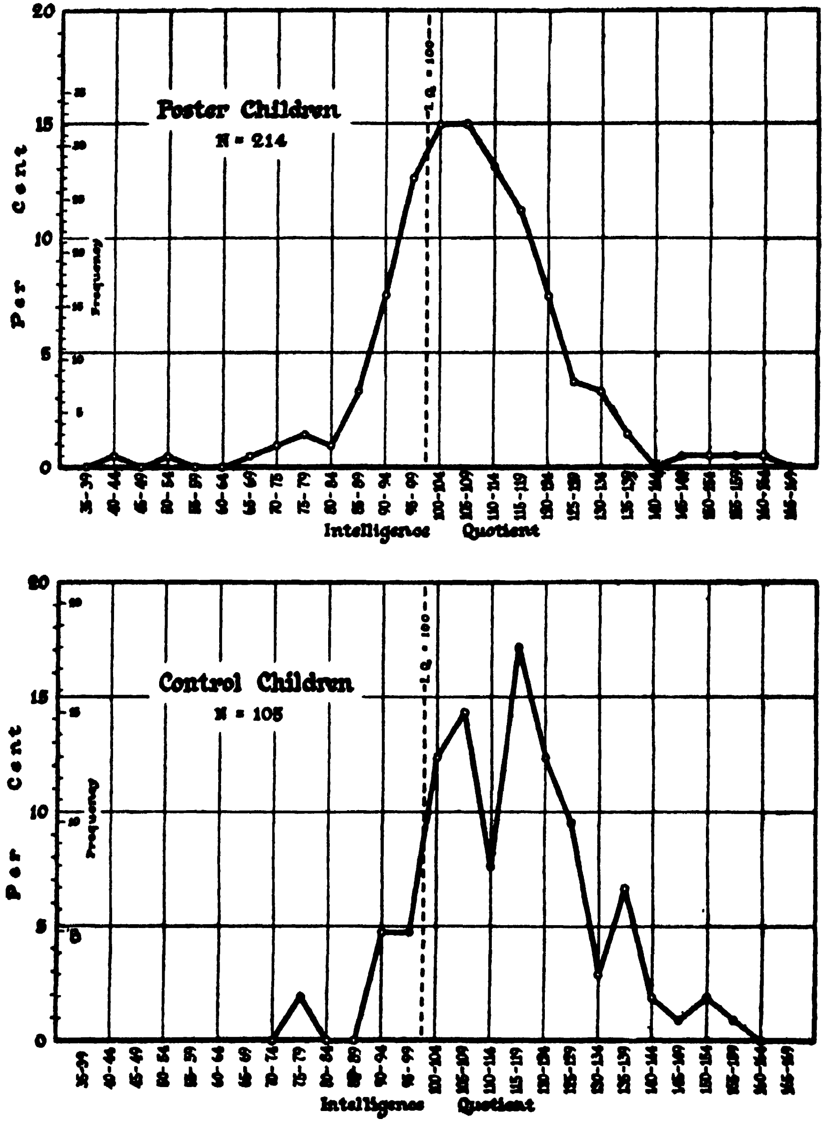


FIG. I.—I.Q. DISTRIBUTIONS OF FOSTER CHILDREN AND CONTROL CHILDREN PRESENTED IN TABLE XI.

children, if they had been reared in homes corresponding to the social level from which their true parents were drawn, would not have been far from 100, more probably a few points over than under 100.

5. Composition of the Groups

This very important phase of the investigation will be presented chiefly by the tabular comparisons of the Foster and Control Groups shown in Tables XI to XXVI.

TABLE XII.—AGE DISTRIBUTION OF CHILDREN BY SEXES

Age	Sex	Foster	Control
5	Boys	28	14
	Girls	36	16
6	Boys	9	8
	Girls	16	7
7	Boys	10	7
	Girls	9	4
8	Boys	15	7
	Girls	14	7
9	Boys	6	3
	Girls	17	4
10	Boys	7	1
	Girls	11	7
11	Boys	....	2
	Girls	12	5
12	Boys	7	2
	Girls	4	5
13	Boys	3	2
	Girls	5	3
14	Boys	2	....
	Girls	3	1
Mean		8.2 yrs.	8.2 yrs.
S. D.		2.6 yrs.	2.6 yrs.
N		214	105
Boys			
N		87	46
Percent		41	44
Girls			
N		127	59
Percent		59	56

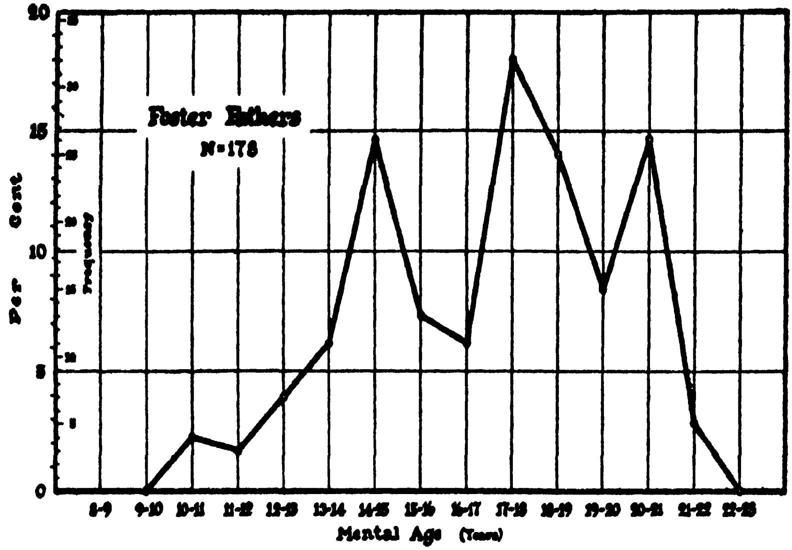
TABLE XIII.—INTELLIGENCE DISTRIBUTION OF PARENTS, IN M. A.

Mental Age	Foster		Control	
	Fathers	Mothers	Fathers	Mothers
21–0 to 21–6	5	5	9	3
20–0 to 20–11	26	19	12	8
19–0 to 19–11	15	5	8	10
18–0 to 18–11	25	20	8	8
17–0 to 17–11	31	35	16	25
16–0 to 16–11	12	23	6	6
15–0 to 15–11	13	15	10	10
14–0 to 14–11	26	21	15	15
13–0 to 13–11	11	21	6	6
12–0 to 12–11	7	16	6	10
11–0 to 11–11	3	14	2	5
10–0 to 10–11	4	8	2	....
9–0 to 9–11	....	2	....	1
Mean	17.0	15.8	16.9	16.3
S. D.	2.6	3.0	3.0	2.8
N	178	204	100	105

*See also Figures II and III.

TABLE XIV.—AGE DISTRIBUTION OF PARENTS

Chronological Age	Foster		Control	
	Fathers	Mothers	Fathers	Mothers
26–28	....	2	2	5
29–31	2	7	4	18
32–34	6	15	11	14
35–37	11	33	19	19
38–40	26	37	19	27
41–43	37	36	15	8
44–46	4	35	11	7
47–49	5	17	4	1
50–52	1	8	8	1
53–55	15	11	2	....
56–58	10	1	....	....
59–61	5	1	3	....
62–64	....	....	....	....
65–67	....	....	1	....
68–70	1	....	....	....
Mean	45.9	41.4	41.0	36.1
S. D.	7.1	6.1	7.2	5.3
N	206	203	99	100



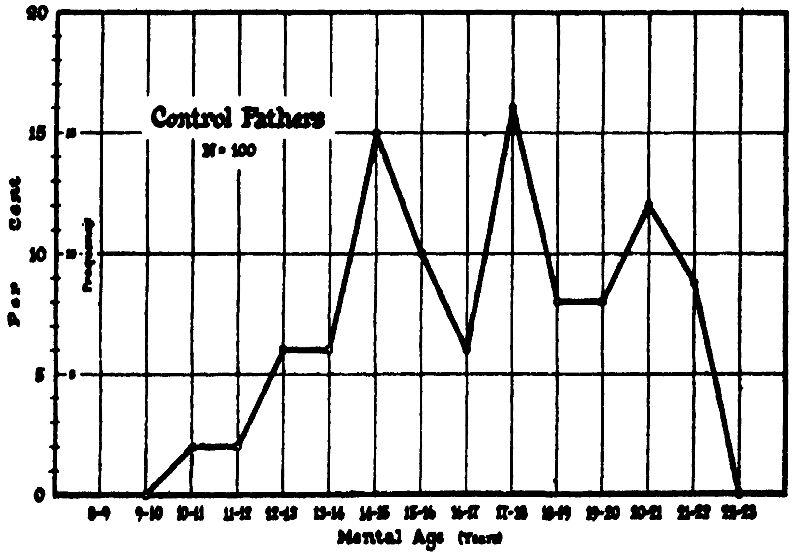


FIG. II.—MENTAL AGE DISTRIBUTIONS OF FOSTER FATHERS AND CONTROL FATHERS

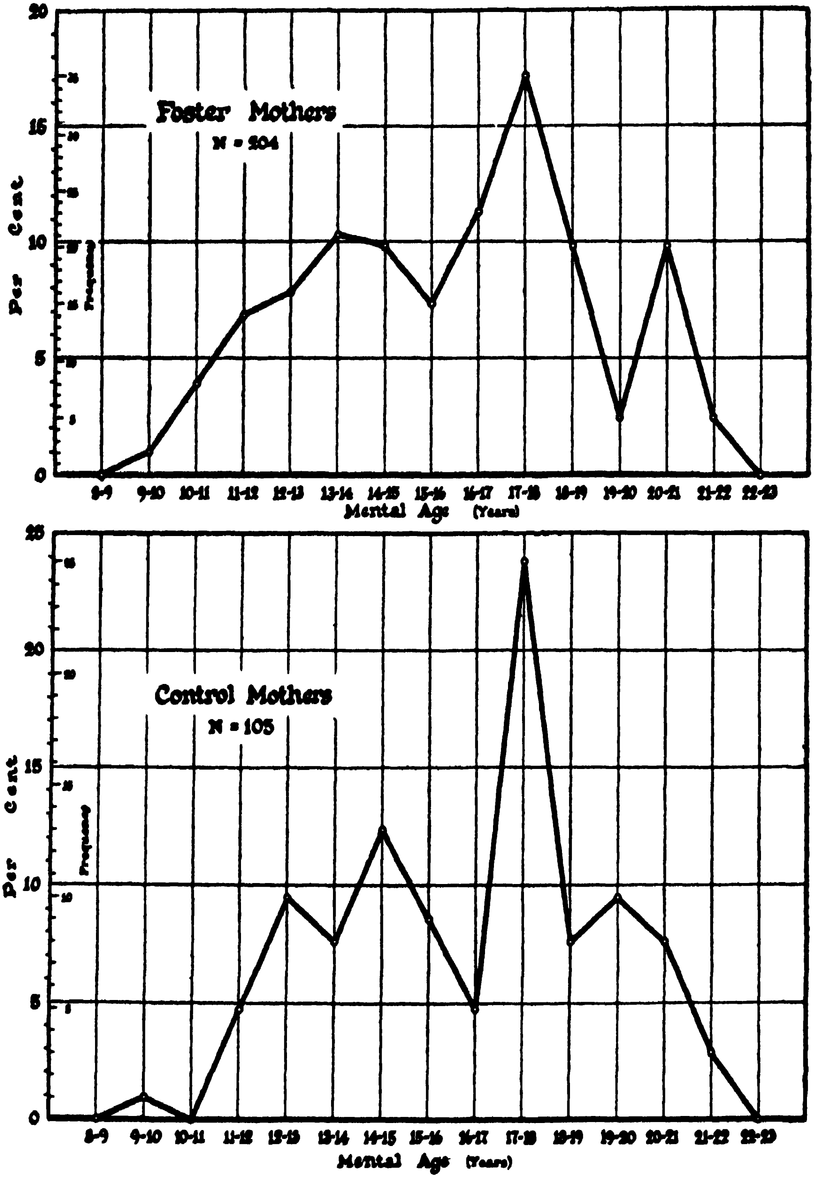


FIG. III.—MENTAL AGE DISTRIBUTIONS OF FOSTER MOTHERS AND CONTROL MOTHERS

TABLE XV.—OCCUPATIONAL CLASSIFICATION OF FATHERS

	Foster		Control	
	N	Percent	N	Percent
1. Professional (exclusive of teachers)	32	15.7	16	15.5
2. Teaching	5	2.5	5	4.9
3. Business owners and managers	80	39.2	33	32.0
4. Commercial employees	21	10.3	14	13.6
5. Salesmen	17	8.3	12	11.7
6. Ranchers	10	4.9	4	3.9
7. Skilled labor	31	15.2	11	10.7
8. Semi-skilled labor	2	1.0	5	4.9
9. Unskilled labor	4	2.0	2	1.9
10. Retired	2	1.0	1	1.0
Total	204	100.1	103	100.1

1. Professional includes architect, attorney, chiropractor, clergyman, commercial artist, dentist, engineer, musician, physician and surgeon, pharmacist.

2. Teaching includes professors and teachers.

3. Business owner and manager includes banker, business owner, (both large and small), contractor, labor organizer, manager, manufacturer, merchant, publisher, superintendent, undertaker.

4. Commercial employee includes auditor and accountant, bookkeeper, cashier, clerk, movie actor, postman, statistician, tax collector.

5. Salesman includes advertising agent, promoter, real estate agent, salesman.

6. Rancher includes both ranch owner and tenant farmer.

7. Skilled labor includes captain in fire department, carpenter, chef, lineman, locomotive engineer, mason, machinist, mechanic, pressman, ship caulker, traffic policeman.

8. Semi-skilled labor includes barber, conductor, delivery man, janitor, smelter.

9. Unskilled labor includes laborer, nightwatchman, switchman, teamster.

The classification of certain of these occupations in Table XV may be open to question, and the scheme of rubrics itself is not a conventional one. The scheme was devised merely to provide a basis for selecting control cases that should match the foster families in cultural status. As such a basis it was quite successful, and it can perhaps be justified on that account. In the Control Group the deviations in some of the rubrics from the fifty percent quotas which we intended to locate are due to insufficient coördination between the three field workers when carrying out the matching process. Fortunately, these deviations did not defeat the purpose of our occupational selection of controls, which was to insure comparability between the Foster and Control Groups in other respects, especially intelligence of parents.

TABLE XVI.—EDUCATION OF PARENTS (AVERAGE SCHOOL GRADE COMPLETED)

	Foster		Control	
	Fathers	Mothers	Fathers	Mothers
Mean	10.7	9.8	10.8	10.7
S. D.	3.9	3.2	4.0	2.9
N	173	194	102	103

TABLE XVII.—FAMILY INCOME (INCLUDING SALARIES AND OTHER SOURCES)

	Foster (N=183)	Control (N=99)
Mean	$6,200	$4,100
Median	3,600	3,000
S. D.	7,400	3,100

TABLE XVIII.—HOME OWNERSHIP AND VALUE OF HOME*

	Foster	Control
Number of homes owned	150	61
Number rented	30	42
Percent owned	83	58
Mean value of homes owned	$13,200	$9,500
S. D. of value	12,000	5,300

*Ranch homes and other homes that were part of income property were omitted from computations of home values. Hence, N for mean value and S. D. of value is only 165 in the Foster Group and 53 in the Control Group.

TABLE XIX.—RATINGS BY FIELD WORKERS OF THE TWO PARENTS IN SYMPATHY, KINDLINESS, AND TACT*

	Foster	Control
Average 'mid-parental rating'	3.1	2.9
S. D.	1.2	1.2
N	198	101

*Ratings are by a seven-point scale, on which 1 is highest rank, and 4 is average. (See Whittier scale, Item IV).

TABLE XX.—WHITTIER RATINGS OF HOMES*

	On Standard Scale				On Extended Scale			
	Foster (N = 206)		Control (N = 104)		Foster (N = 206)		Control (N = 104)	
	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.
Total Whittier index	23.3	1.9	23.0	2.3	23.9	2.4	23.5	2.6
Necessities	4.7	0.4	4.6	0.6	4.9	0.6	4.7	0.7
Neatness	4.5	0.6	4.4	0.6	4.5	0.6	4.4	0.6
Size	4.7	0.5	4.4	0.7	4.7	0.5	4.4	0.7
Parental conditions	4.8	0.4	4.8	0.3	4.9	0.6	5.0	0.5
Parental supervision	4.7	0.4	4.8	0.4	4.9	0.7	5.0	0.7

*For explanation of 'standard' and 'extended' scales, see page 231.

TABLE XXI.—CULTURAL RATINGS OF HOMES

	Foster			Control		
	Mean	S. D.	N	Mean	S. D.	N
Total culture index	16.9	4.2	186	16.3	4.3	101
Speech	3.5	0.9	199	3.4	0.8	103
Education	3.7	0.9	187	3.8	0.9	101
Interest	3.2	0.9	198	3.1	0.8	103
Library	3.1	1.1	201	3.0	0.9	103
Artistic taste	3.2	1.0	203	3.0	0.9	103

TABLE XXII.—HOME INSTRUCTION OF CHILDREN BY MEMBERS OF THE HOUSEHOLD IN HOURS WEEKLY*

(As reading or telling stories, number work, etc.)

Age of Child When Instructed	Foster			Control		
	Mean	S. D.	N	Mean	S. D.	N
2 and 3	3.7	4.4	179	5.0	5.2	101
4 and 5	4.8	5.1	179	5.2	5.2	101
6 and 7	3.8	4.7	127	2.1	3.2	71

*The information blanks of some mothers, we found, had not been filled out on the side providing space for the item on home instruction. When this particular item was not answered, it was entered in the computation as zero—providing the blanks met the criterion for inclusion in the tabulation. The criterion demanded that blanks should be filled out at least in part on the side containing this item.

TABLE XXIII.—PROPORTION OF CHILDREN WHO HAVE HAD PRIVATE TUTORING OUT OF SCHOOL
(Music, dancing, drawing, etc.)

Age	Foster Boys		Control Boys		Foster Girls		Control Girls	
	N	Percent	N	Percent	N	Percent	N	Percent
5	2	8	2	14	13	39	6	38
6	...	....	1	12	6	38	2	33
7	3	30	4	57	4	44	1	25
8	8	53	2	29	8	73	4	67
9	2	40	3	100	13	87	1	33
10	6	100	1	100	4	40	5	71
11	...	....	2	100	9	100	4	80
12	4	67	1	50	4	100	5	100
13	1	33	1	50	3	75	2	67
14	1	50	...	....	3	100	1	100
Total	27	34	17	37	67	58	31	55

TABLE XXIV.—HOUSEHOLD SIZE, NUMBER OF ADULTS, AND NUMBER OF CHILDREN IN HOUSEHOLD*

	Foster	Control
Mean number in household	3.9	4.7
Mean number of adults	2.4	2.3
Mean number of children	1.5	2.3

*Based upon 203 families in the Foster Group and 104 families in the Control Group.

In addition to the comparisons between the Foster Group and Control Group shown in Tables XI to XXVI, there are data pertaining only to the Foster Group which will be of assistance in the final interpretation of results. These are shown in Tables XXVII to XXX.

The conclusions to be drawn from this tabular material may be stated briefly:

1. The parents in the two groups appear to be remarkably similar in intelligence (Table XIII), occupational status (Table XV), kindliness or harmoniousness of temperament (Table XIX), education (Table XVI), early background (Table XXVI), and supervision given the children (Table XX). They are rather similar in attitude towards their children in so far as this is expressed in their ratings of the children upon ten traits (Table XXIV). However, there is a slight tendency for the foster parents

TABLE XXV.—TRAIT RATINGS OF CHILDREN BY PARENTS*

	By Foster Fathers			By Foster Mothers			By Control Fathers			By Control Mothers		
	M	S.D.	N	M	S.D.	N	M	S.D.	N	M	S.D.	N
1. Will power and perseverance	4.38	2.25	162	4.47	2.29	174	4.71	2.18	98	4.99	2.19	102
2. Cheerfulness and optimism	3.84	1.97	163	4.03	2.13	179	3.96	1.99	98	4.09	2.21	101
3. Musical appreciation	4.92	2.24	163	4.73	2.23	179	5.04	2.16	97	5.04	2.19	101
4. Sense of humor	4.22	2.04	158	4.39	2.09	175	4.00	1.78	98	4.81	2.13	102
5. Permanency of moods	5.66	2.68	159	5.78	2.70	174	5.72	2.84	97	5.79	2.58	101
6. Leadership	4.71	2.15	157	5.10	2.41	175	5.17	2.16	98	5.88	2.46	101
7. Sympathy and tenderness	3.85	2.30	163	3.86	2.27	180	4.31	2.25	98	4.03	2.20	101
8. Conscientiousness	5.10	2.25	163	5.17	2.24	176	4.93	2.15	97	4.90	2.13	101
9. Originality	4.72	2.16	160	5.01	2.24	175	5.22	2.09	97	5.63	2.20	100
10. General intelligence	4.39	2.01	164	4.93	1.91	181	4.87	1.71	98	5.29	1.54	101
Mean of ratings on all 10 traits	4.58			4.75			4.79			5.04		

*These ratings were made on a 13-point scale, in which 1 was highest, and 7 supposedly average. The mean rating by foster and control parents for all ten traits was computed only for what it might show regarding the attitude of the parents toward the children. Fathers seem to be slightly more lenient in their ratings than mothers and foster parents slightly more lenient than true parents.

TABLE XXVI.—BIRTHPLACE OF PARENTS

	Foster Fathers	Foster Mothers	Control Fathers	Control Mothers
Alabama	....	....	1	....
Arkansas	1	....	....	....
Arizona	1	1	....	....
California	42	57	30	35
Colorado	5	1	....	2
Florida	....	2	....	....
Idaho	....	....	1	....
Illinois	8	17	7	5
Indiana	7	3	4	2
Iowa	15	11	2	2
Kansas	4	11	3	1
Kentucky	1	1	....	1
Louisiana	....	....	1	....
Maine	1	1	....	....
Maryland	1	....	....	1
Massachusetts	4	3	2	1
Michigan	5	6	1	1
Minnesota	2	2	1	3
Missouri	6	6	7	5
Mississippi	1	....	....	....
Montana	....	....	....	1
Nebraska	3	5	4	2
Nevada	2	....	....	1
New Jersey	3	1	....	1
New York	12	14	6	3
North Carolina	1	....	....	....
North Dakota	....	2	1	....
Ohio	12	8	3	4
Oregon	3	3	....	2
Pennsylvania	7	8	3	5
South Dakota	....	2	2	1
Tennessee	....	....	3	1
Texas	2	3	1	1
Utah	3	....	2	4
Vermont	....	....	1	....
Washington	....	1	....	1
West Virginia	2	....	....	....
Wisconsin	5	5	1	2
Wyoming	1	....	....	....
Australia	1	....	1	2
Canada	2	1	4	5
Caroline Islands	....	....	1	....
Denmark	....	2	....	1
England	8	5	4	....
France	....	1	....	....

(Cont. p. 270)

TABLE XXVI—*Cont.*

BIRTHPLACE OF PARENTS

	Foster Fathers	Foster Mothers	Control Fathers	Control Mothers
Germany	4	1	....	1
Holland	....	1	....	....
Italy	....	....	....	1
Ireland	1	1	1	1
Jugo Slavia	....	....	1	....
Russia	....	....	1	....
Scotland	1	1	2	4
Sweden	....	....	....	1
Switzerland	1	....	....	....
Trinidad Islands	....	1	....	....
Total United States	160	174	87	88
Total foreign	19	15	15	16
Percent United States	89	92	85	85
Grand total	179	189	102	104

TABLE XXVII.—AGE OF PLACEMENT AND AGE OF ADOPTION OF 200 FOSTER CHILDREN

	Age When Placed	Age When Adopted
Mean	3 mos. 2 days	14.3 mos.
S. D.	3 mos. 1 day	9.1 mos.

TABLE XXVIII.—NUMBER AND PERCENT OF FOSTER CHILDREN WHO HAVE BEEN TOLD OF THEIR ADOPTION

Age	Number Told	Number Not Told	Percent Told
5	12	44	21
6	6	14	30
7	4	12	25
8	8	18	31
9	9	11	45
10	10	7	59
11	5	7	42
12	6	5	55
13	5	2	71
14	1	4	20
Total	66	124	35

TABLE XXIX.—LENGTH OF TIME FOSTER PARENTS HAVE BEEN MARRIED AT THE TIME THEY APPLY FOR FOSTER CHILDREN*

Years Married	Number of Couples	Years Married	Number of Couples
Less than 1	1	14	3
1	4	15	3
2	5	16	1
3	12	17	1
4	8	18	6
5	11	19	1
6	10	20	2
7	11	21	2
8	8	22	.....
9	5	23	1
10	9	24	.....
11	9	25	.....
12	6	26	1
13	4		

Mean	8.7 yrs.
S. D.	5.36 yrs.
N	124

*Data complete upon this point only for the Native Sons and Daughters cases. Figures above are based only upon these.

TABLE XXX.—INCOME AND BARR RATINGS OF FOSTER PARENTS AT TIME OF APPLYING FOR CHILD*

	Income	Barr Rating of Father
Mean	\$2,700	11.6
S. D.	2,000	2.6
N	146	195

*About 50 percent of the Barr ratings represent the combined judgment of Mr. Kurtz and myself.

to assign higher ratings. They are reasonably similar in amount of time spent in reading to the children, teaching them, etc., when the children are very young (Table XXII). The slight apparent excess of time spent by control parents over that spent by foster parents when the children were two and three, or four and five, may be due to the fact that the field workers took greater pains to get all blanks completely filled out for every case in the Control Group than in the Foster Group. In the Control Group there would consequently be less likelihood of this item being overlooked through carelessness, and of thus being entered as zero in our tabulation. Certainly, the foster parents showed fully as much affection for their foster children and devotion to their welfare as did the true parents encountered in the Control Group.

2. In age the foster parents show a significant, though not very large, excess over the control parents (Table XIV). This would certainly be expected, in as much as the foster parents seldom applied for foster children until it seemed probable that they could not have children of their own (see Table XXIX on length of time foster parents were married at time of application). It seems extremely unlikely that the four or five years average excess age of foster parents over control parents could introduce disturbing factors of any importance, because no significant change in the parent-child correlations occurred when the chronological age of the parents was partialled out, and no decided correlation was found between the chronological age of the parents and the I.Q.'s of the children.

3. As to the other specifically measured environmental influences the Whittier and culture ratings agree noticeably well in the two groups, item for item (Tables XX and XXI). However, the means and standard deviations of the family incomes correspond poorly in the two groups (Table XVII). But an investigation of the distribution of incomes in the Foster Group showed that the rather high average income can be accounted for chiefly by a few extremely high incomes. The median incomes for the two groups, it is seen, fall much closer together. Reference to Table XXV shows that at the time the foster parents applied for children, the mean income was *less* than the present mean in the Control Group. On an average, the foster parents applied for children eight or nine years before the time of this investigation. The foster fathers average five years older than the control fathers, and most of them are at a period of life at which the income for professional and business men often increases rather rapidly. As a consequence, the difference in income level between the two groups probably does not represent any real difference in the kind of surrounding in which the children have been reared, though it does reduce the significance of the correlations between family income and child's I.Q.

The percent of families owning their homes and the valuation of the homes are distinctly different in the two groups (Table XVIII). The excess of home valuation in the Foster Group over that in the Control Group can be explained in the same way as the similar discrepancy in average income. In so far as the environmental standards of the children are concerned, the difference in proportion of home owners found in the two groups is probably not significant, since homes were matched for locality and type of district, and the measures of material and cultural surroundings found in the two groups were practically indistinguishable. But the difference under discussion does render the biserial correlations which were computed between the child's I.Q. and family ownership of home unsusceptible of precise interpretation.

4. In proportion of children of different ages who have been given private tutoring (such as music, dancing, etc.) the Control Group shows an eighteen percent excess (Table XXIII). To the writer, the most probable explanation of this fact is that, other things being equal, brighter children are more likely to desire such tutoring. Eight points I.Q. difference in average level in favor of the Control Group may possibly be sufficient to account for the moderate excess of private tutoring received by that group.

5. The average size of the household is greater in the Control Group than in the Foster Group (Table XIV), and if association with siblings might conceivably influence the development of the I.Q., this difference would introduce disturbing factors. To test this hypothesis, the average I.Q's of foster children having older and younger foster sibs were computed, with the following results:

Status of Our Case	Average I.Q.	N	Percent
Is an only child....................	107.8	127	59
Has older foster sib or sibs.........	105.4	36	17
Has younger foster sib or sibs.......	106.4	38	18

Such slight differences as appear are obviously not significant.

6. Finally, certain facts regarding the placement and adoption of the foster children support the idea that the family relationship between foster parents and children is a perfectly normal one. Tables XXVII and XXVIII show that most of these children have lived in their foster homes from very early infancy, and that the majority of them have never been told that they are foster children.

7. As to the children themselves, we are not much concerned at this point about how closely those in the two groups correspond in mental development, since it is the effect of the factors already considered *upon* such mental development which this study seeks to evaluate. While the average mental levels of the two groups are disparate, it is important to mention that the variabilities of the two groups are nearly identical. In chronological age, the means and variabilities correspond exactly, and in proportions of children of the two sexes, the agreement appears to be sufficiently good.

From the foregoing evidence, the conclusion seems justified that, with the few relatively unimportant exceptions noted, the correlations reported in the next section for the two groups are fully comparable.

VI. Main Results of the Study

As this investigation was conceived, the chief emphasis was laid upon intelligence and the factors conditioning its development. It was decided to touch upon traits other than intelligence only to the extent that this was possible without undue expenditure of time. A consideration of the factors influencing vocabulary was possible, since a vocabulary test is one element of the Stanford-Binet Scale for mental levels beyond the seventh year. The Woodworth-Cady questionnaire was administered to children of ten or over in the hope that trends might appear that would suggest the factors underlying emotional instability. Unfortunately, the number of children eligible to answer this questionnaire was smaller than we had anticipated, and the results are consequently not very significant. Rough measures of several other mental and character traits of the children were secured through ratings and estimates by their parents (or foster parents), and a crude index of the children's school achievement was obtained by noting their grade placement. The results from our statistical treatment of all these aspects of the problem will be presented, but the data for other traits than intelligence can not be regarded as very reliable.

1. Factors Underlying Differences in Intelligence

Table XXXI presents corresponding correlation coefficients for the Foster and Control Groups between child's I.Q. and the environmental and hereditary factors for which we obtained measures.

TABLE XXXI.—CHILD'S I. Q. CORRELATED WITH ENVIRONMENTAL AND HEREDITARY FACTORS*

Factor	Type of r	Foster			Control		
		r	P.E.	N	r	P.E.	N
Father's M. A.	P.M.	.07	.05	178	.45	.05	100
Mother's M. A.	P.M.	.19	.05	204	.46	.05	105
Mid-parent M. A.	P.M.	.20	.05	174	.52	.05	100
Father's vocabulary	P.M.	.13	.05	181	.47	.05	101
Mother's vocabulary	P.M.	.23	.04	202	.43	.05	104
Whittier index	P.M.	.21	.04	206	.42	.05	104
Whittier index (using 5-yr.-olds only)	P.M.	.29	.08	63	....	....	....
Culture index	P.M.	.25	.05	186	.44	.05	101
Culture index (using 5-yr.-olds only)	P.M.	.23	.08	60	....	....	....
Grade reached by father	P.M.	.01	.05	173	.27	.06	102
Grade reached by mother	P.M.	.17	.05	194	.27	.06	103
Parental supervision rating 3 or 4 *vs.* 5 or 6	B.	.12	.05	206	.40	.09	104
Income	P.M., K.	.23	.05	181	.24	.06	99
No. of books in home library	P.M., K.	.16	.05	194	.34	.06	100
Owning or renting home	B.	.25	.07	149	.32	.10	100
....	....	....	....	....	....	....	....
No. of books in child's library	P.M., K.	.32	.04	191	.32	.06	101
Private tutoring (in music, dancing, etc.)	B.						
Boys		.06	.10	77	.43	.11	46
Girls		.31	.08	108	.52	.09	56
Five-year-girls only		.50	.12	31	....	....	....
Home instruction by members of household (hrs. weekly)	P.M.						
Ages 2 and 3		.34	.04	181	−.05	.07	101
Ages 4 and 5 (children over 5)		.15	.06	129	−.03	.08	71
Ages 6 and 7 (children over 7)		.03	.07	88	.24	.09	46
Ages 2 and 3 (5-yr.-olds only)		.18	.09	51	....	....	....
Ages 4 and 5 (5-yr.-olds only)		.13	.09	52	....	....	....
Father's rating of child's intelligence	P.M.	.49	.04	164	.32	.06	98
Mother's rating of child's intelligence	P.M.	.39	.04	181	.52	.05	101

*The following abbreviations are used in this table: M.A. for mental age. P.M. for product-moment correlation. B. for biserial correlation. K. for Professor Kelley's auxiliary score method.

See also the tables of correlation arrays for child's I. Q. with Father's M. A. and Mother's M.A., from which the corresponding r's in this table were computed. (Appendix, II).

The significance of the division of the table by the dotted line is explained in the text, p. 281.

In the field work it proved to be impossible to obtain full supplementary information upon all the cases tested, but correlations are given which utilize all the information we have with respect to each item. The number of cases entering into each correlation consequently varies somewhat.

Kelley's auxiliary score method, described in full in his text on *Statistical Method* (9), pp. 185 ff., is a device for straightening curvilinear regression lines empirically in such a way as not to capitalize chance. It was employed in the correlations with income and books in library.

Sheppard's correction was applied to all standard deviations used in the study—those published in tabular form as well as those entering correlation computations.

A word should be inserted at this point regarding the probable errors of the correlations reported. They have all been computed by standard formulae. Now, it can be shown that if, in two correlated series, some of the items in one variable enter the correlation array in more than one pair of measures, the effective N to use in computing the probable error of the coefficient is less than the total number of pairs. The effective N lies at some value intermediate between the total number of pairs and the total number minus the number of items entering the correlation more than once. We have such a situation in the Foster Group, for it contains 21 pairs of *double cases, i.e.,* foster siblings being reared in the same home. Correlations of various factors with measures of the children consequently have about 21 items which enter the correlations twice. We have ascertained, however, that within the limits that the effective N must lie in our Foster Group, no change occurs in the first or second decimal place of the P.E.'s in most of the correlations, nor greater than 1 point in the second decimal place in the remaining correlations. This difficulty can therefore be neglected, and was mentioned only to avert possible criticism of the P.E.'s that are published. There is no such difficulty in the Control Group, of course, since only one child was considered from any one family.

In addition to the correlations of Table XXXI, a few coefficients were computed for the Foster Group only. These (Tables XXXII and XXXIII) were based upon specially selected groups of subjects.

As a matter of interest, a biserial correlation was computed between the child's I.Q. and his knowledge or lack of knowledge that he was an adopted child. With 189 cases for which we had data upon this point, the biserial correlation was .10 ± .06, to which we can attach no significance. Telling or not telling a child of his

TABLE XXXII.—CORRELATIONS BETWEEN I. Q.'S OF CHILDREN AND FACTORS INFLUENCING THESE I. Q.'S IN THE CASE OF CHILDREN WHO WERE LESS THAN ONE MONTH OLD WHEN TAKEN BY FOSTER PARENTS

Factor	r	P.E.	N
Foster father's M.A.	.02	.09	60
Foster mother's M.A.	.15	.08	66
Mid-foster parent M.A.	.08	.09	58

TABLE XXXIII.—CORRELATIONS BETWEEN THE I.Q.'S OF FOUNDLINGS* AND FACTORS INFLUENCING THESE I.Q.'S

Factor	r	P. E.	N
Foster father's M.A.	.15	.21	10
Foster mother's M.A.	.14	.18	13
Mid-foster parent M.A.	.24	.20	10

*"Foundling" defined as a child picked up on a doorstep, in an automobile, etc., without any means of identification.

adoption is related much more to the age of the child and the intelligence of his foster parents than to his I.Q., as is evident in the following correlations:

	Biserial r	P.E.	N
Foster father's M.A. and whether or not child was told	.21	.07	156
Foster mother's M.A. and whether or not child was told	.43	.06	156
Age of child and whether or not he was told	.25	.07	156

When age of child is partialled out from the correlations with father's and mother's mental age, the first two correlations above are .21 and .42, respectively.

Correlations between the I.Q's of the pairs of unrelated foster siblings which are encountered are also of interest. Pairs were correlated against one another using chance arrangement.

	r	P.E.	N
Two unrelated foster children reared in same home	.23	.14	21
Same cases plus 7 in which a foster child and a true child were reared in same home	.11	.13	28

The first (which is also the higher) of these correlations is probably the more valid, since the seven true children in the second correlation introduce a sample from a non-comparable population of children having a higher central tendency than that of the foster group.

Another point of interest lies in the possible effect of schooling upon intelligence. The five-year-olds provide the only group in which this point can be considered, since practically all the older children had attended school. In the Foster Group our records show that 30 five-year-olds had attended kindergarten or first grade, and that 30 five-year-olds had not. The mean I.Q. of those who had attended kindergarten or school was 111, and of those who had not attended, 107. The difference of 4 points in favor of the first group is quite possibly due to schooling, but since the probable error of the difference is 2.8 points, it cannot be considered as reliably established. Moreover, even if the difference were a reliable one, nothing in our data could show that part of the difference was not due to a tendency on the part of parents to enter bright children in school at an earlier age than dull ones.

The following correlations were found between the mental ages of fathers and mothers in the two groups:

	r	P.E.	N
Foster	.42	.04	174
Control	.55	.05	100

Let us now return to the data of Table XXXI. A dotted line was there inserted to separate from the more important coefficients certain coefficients that are ambiguous because they represent relationships between variables that might conceivably have reciprocal effects upon one another. For example, do the books in a child's library stimulate the growth of his I.Q. or does the child of high intelligence tend to collect more books around him? Does reading the Burgess bedtime stories to a two-year-old enhance his mental potentiality or does the child with high mental potentiality clamor loudest for the bedtime stories? Such considerations relegate these correlations to the realm of speculation; they are presented only for what interest they possess.

However, the first correlations of the table offer a clearer picture. The variables listed there could scarcely be thought of as influenced to any appreciable degree by the intelligence of the children in a home (at least by the intelligence of children as young as ours). Consequently, these correlations, when significantly greater than their own probable errors, can be taken as actual measures of the effect of environment in the Foster Group, and as measures of the combined effects of heredity and environment in the Control Group. The point is emphasized in this section—where results are first presented which are to serve as a basis for subsequent statistical treatment and final conclusions—that *the differences between corresponding correlation coefficients in the Foster Group and Control Group are striking and consistent.*

2. Corrections for Attenuation

While the raw correlations of Table XXXI show the drift of the evidence, they do not tell the complete story. In Spearman's nomenclature they are "attenuated," owing to the unreliability of the measures upon which they are based. Spearman's formula of correction for attenuation was applied to some of the most important correlations to yield the best available estimate of what the relationships would have been if perfectly reliable, *i. e.*, 'true' measures, could have been used. Coefficients so computed represent more accurately than raw coefficients the actual contributions of various factors to variability in a criterion.

In computing coefficients corrected for attenuation, the problem of ascertaining sound reliabilities to be used in the Spearman formula was a perplexing one.

To find the reliability of the Stanford Binet Test for children and for adults, the Spearman-Brown formula[11] was applied to 'split halves.' This formula, which is based upon an assumption that the 'split halves' are fully comparable with respect to the function that both halves purport to measure, does not provide an entirely satisfactory measure of reliability for a battery as variegated as the Stanford Binet. Unless we may make the additional (as yet unproved and possibly untrue) assumption that the function we call intelligence is due to a general factor plus no specific factors, the measures of Stanford Binet reliability which are reached by methods described below may probably be considered as too low. Possibly, the high correlations reported by Herring (6)

[11] The formula is:

$$r_{11} = \frac{2r_{\frac{1}{2}\,\mathrm{I/II}}}{1 + r_{\frac{1}{2}\,\mathrm{I/II}}}$$

between I.Q's measured on the Herring and on the Stanford revisions of the Binet Test (.97 to .99) provide a better estimate of the true reliability of the Binet Scale, although these reliabilities, too, seem open to question. The material in the two versions is so similar as possibly to capitalize chance skills, techniques or information that an individual has happened to acquire during the course of his life.[12]

Fortunately, the indeterminate error in the reliability coefficients for the Stanford Binet does not seriously affect the results based upon the reliabilities. This is because the total corrections (by the formula for correcting attenuation) amount to only a few points in the second decimal place when reliabilities are even approximately as high as those found and when the raw correlations to which the corrections are applied are not high themselves.

The specific procedure employed to determine the reliabilities was as follows:

a. Children. From the test files of the Stanford Psychology Department, a distribution of fifty complete tests was built up. The subjects represented by the tests matched the Foster Group in age and I.Q. The reliability of half the test was computed by correlating I.Q.'s based upon halves split by the alternate item method, and then the Spearman-Brown formula was applied to find the reliability of the test as a whole.

Next the correlation between the complete form and the 'lopped' form (see explanation of 'lopped' form p. 230) was computed for this group. The reliability of the lopped form was then inferred by a formula derived for this purpose by Professor Kelley,

$$r_{2\,\mathrm{II}} = r^2_{12}\, r_{1\,\mathrm{I}} \quad \text{where}$$

$r_{2\,\mathrm{II}}$ is the reliability of the lopped form
$r_{1\,\mathrm{I}}$ is the reliability or the complete form
r_{12} is the correlation between the composite and the lopped form.

The following coefficients were found in using the successive steps:

$r_{\frac{1}{2}\mathrm{I}/\mathrm{II}} = .79$
$r_{1\,\mathrm{I}} = .88$ (correction by Spearman-Brown formula)
$r_{12} = .97$
$r_{2\,\mathrm{II}} = .83$, the value of reliability used in this study.

The same reliability coefficient was used for the tests of the control children as for the foster children, since the variabilities in I.Q. of the two groups were exactly the same.

---

[12] In this connection, see Professor Kelley's *Note on the Reliability of a Test* (8) in which the conditions tending to raise or lower reliabilities spuriously are set forth.

By way of interest it may be noted that the above values for $r_{1\,\mathrm{I}}$ and r_{12} agree quite well with similar values found by James DeVoss (2) and by Floyd Ruch (14), respectively. The former found I.Q. reliabilities clustering around .92 for single age groups; the latter found a correlation of .98 between the complete and lopped Binet forms, using an adult population.

b. Adults. One third (59) of the tests of foster fathers were selected in such a way that the numbers from each mental age level would be proportional to the corresponding numbers in the entire group. These were split by the alternate item method; the halves were correlated, and the correlation corrected by the Spearman-Brown formula, yielding .86. But this 'reliability,' based upon lopped tests, is spuriously high, since the lopped test assumes perfect performance on all tests below the lowest level given, and consistent failure on all tests above the highest level given. Now in the built-up distribution of tests of children, the complete tests had a Spearman-Brown corrected reliability of .88, and a similar procedure yielded .90 for the lopped version—a value only .02 points higher. The empirical assumption was made that the 'reliability' .86 found for the lopped adult tests was .02 higher than a real reliability based upon complete tests would have been. Letting

$$r_{1\,\mathrm{I}} = .84$$
$$r_{12} = .97$$

then, by Kelley's formula, $r_{2\,\mathrm{II}} = .79$, the value of reliability used in this study.

The reliabilities of adult and children's vocabularies were easily determined by correlating one list of the vocabulary test against the other list, and inferring the reliability for the complete test by the Spearman-Brown formula. This was done for fathers, mothers, and children in the Foster Group, and the reliabilities thus determined were used also in the Control Group.

Reliabilities for the Whittier Scale and the Culture Scale were taken to be the correlations between independent ratings by Mrs. Jensen and myself upon all the Whittier and Culture blanks in the Control Group. Strictly speaking, these correlations are not reliabilities, for they are almost certainly somewhat higher than two series of ratings based upon data gathered twice from the same homes with a month or a year intervening would have been. They represent rather the upper limit of reliability.

The reliability of family income was assumed to be unity, which, of course, is too high, although our impression was that the parents attempted to give us accurate information upon this point.

While fully realizing that the determination of reliabilities of these variables is not without flaws, I think it probable that the correlations forthwith presented represent a closer approximation to the truth than the raw correlations. It is possible for reliabilities to vary several points in the second decimal from their proper value without seriously distorting the correlations corrected for attenuation. Moreover, it is probable that some reliabilities estimated too high are compensated for by others estimated too low; so that in the multiple correlations which have been computed very reasonable values may be obtained.

Summarizing this discussion, the following reliabilities were used:

Stanford Binet, children's I.Q.	.83
Stanford Binet, adult mental age	.79
Vocabulary, fathers	.96
Vocabulary, mothers	.96
Vocabulary, children (age 8-14)	.93
Whittier index	.92
Culture index	.95
Income	1.00

TABLE XXXIV.—CHILD'S I. Q. CORRELATED WITH ENVIRONMENTAL AND HEREDITARY FACTORS AND CORRECTED FOR ATTENUATION*

	Foster		Control	
	r	N	r	N
Father's M.A.	.09	178	.55	100
Mother's M.A.	.23	204	.57	105
Father's vocabulary	.14	181	.52	101
Mother's vocabulary	.25	202	.48	104
Whittier index	.24	206	.48	104
Culture index	.29	186	.49	101
Income	.26	181	.26	99

*The P.E.'s are all in the neighborhood of .06.

It is obvious that the fairly high correlations between I.Q. and environmental factors in the Control Group are due to a large extent to the high association between parental intelligence and environmental factors. Such association can be shown by the correlations of Table XXXV.

It is perhaps surprising that income correlates somewhat less with parents' mental level than do the other environmental measures. Nevertheless, the income correlations probably approximate their true values, in as much as the correlation of foster father's mental age and income, .31, agrees with the corresponding value in the Control Group fairly well.

In interpreting all the foregoing correlation comparisons between the Foster Group and the Control Group, it should be borne in mind that the *squares* of the correlations, rather than the correlations themselves, represent the portion of the variance of children's I.Q's that can be accounted for by reference to the respective variables [I.e $\sigma_{1.2} = \sigma_1^2 \, (1 - r_{12}^2)$]. This consideration emphasizes the differences in strength of relationship found in the two groups.

TABLE XXXV.—PARENTAL CORRELATIONS IN CONTROL GROUP

Factor	Father's Mental Age		Mother's Mental Age	
	Raw	Corr. for atten.	Raw	Corr. for atten.
Whittier index	.60	.70	.60	.70
Culture index	.67	.77	.71	.82
Income (by aux. score meth.)	.. 8	.43	.40	.51
Father's education	.46	....	....	....
Mother's education	....	....	.62	....

3. Multiple Correlations

It should also be pointed out that the foregoing correlations do not provide an absolute basis for evaluating the relative influence of various environmental factors. The intercorrelations between these factors are so complex, and the status of the factors as possible causes and effects of one another is so uncertain, that their unique contributions to the variance of the children's I.Q's are impossible to extricate. Because of the difficulties just mentioned, partial correlation technique is obviously unadapted to the problem (1). It is possible, however, to arrive at an estimate of the *total* effect of our measured environmental factors through multiple correlation technique.

Accordingly, we determined the multiple correlation of the following factors with child's I.Q. in the foster and control groups:

Father's mental age	Father's education
Mother's mental age	Mother's education
Father's vocabulary	Whittier index
Mother's vocabulary	Culture index
	Family income

To have gone through the operation of computing multiple correlations that utilized all nine of the variables in question would have been enormously time-consuming. To save labor, certain variables were eliminated, after first demonstrating, through multiples using three of four variables, that they contributed practically nothing to an estimate of the child's I.Q. not already contributed by variables retained for the final multiple. For example, in the foster multiple, income was retained, but Whittier and Culture indices were dropped out, because the multiple of I.Q. with all three together (.34) was only .01 higher than the correlation (.33) between I.Q. and income alone; again, mother's vocabulary was retained, but mother's mental age and mother's education were dropped out because the multiple of I.Q. with all three together (.254) was only .005 higher than the correlation (.249) between I.Q. and mother's vocabulary alone. Similarly, in the Control Group, certain variables were not used. The variables finally employed no doubt yield values for the multiple correlations that attain, within one or two points in the second decimal, to what the values would have been had we used all nine variables.

The factors retained for the foster multiple were: father's mental age, father's vocabulary, mother's vocabulary, income. Those retained for the control multiple were: father's mental age, father's vocabulary, mother's mental age, Whittier index. The unexpected predictive prepotency of parental vocabulary over parental mental age is probably an adventitious fact, due to the higher reliability of the vocabulary test.

The multiple correlations of Table XXXVI summarize the chief statistical results of the study in two clear-cut comparisons. They

TABLE XXXVI.—MULTIPLE CORRELATIONS OF HEREDITARY AND ENVIRONMENTAL FACTORS WITH CHILD'S I. Q.

	Foster			Control		
	r	P.E.	N	r	P.E.	N
Raw multiple	.35	.05	164	.53	.05	95
Multiple using r's corrected for attenuation*	.42	...	164	.61	...	95

*The P.E.'s of the multiples using correlations corrected for attenuation are not subject to calculation by any methods at present available, but they are, of course, somewhat higher than the P.E.'s for the raw multiples.

show more distinctly than do any of the results from previous sections the significant differences between the outcomes for the Foster Group and those for the Control Group. They will comprise the basis for the fundamental evaluation and interpretation of results which is offered in the eighth, or concluding section.

VII. Minor or Collateral Results

1. Factors Underlying Differences in Other Traits

Only a small proportion of the total space apportioned to this study can be devoted to a discussion of traits other than intelligence. With the exception of vocabulary, the measures or ratings used for the special traits other than intelligence are not sufficiently reliable to make the results of notable significance; and in many situations the possibility of interaction between the trait measured and the influence we seek to evaluate is so markedly present that the relationships uncovered can be interpreted only in a tentative way. Hence, the results reported in the next few pages are presented with little discussion, and only for what interest or suggestive value they may have.

2. Vocabulary

In dealing with vocabulary scores, we were confronted with a test for which no age norms in terms of means or sigmas were available. It seemed inadvisable to use raw scores in the correlations and then try to correct the results by partialling out the chronological age of the children, since tests such as the vocabulary lists tend to show a fan-shaped distribution when plotted against chronological age (owing to increasing standard deviation with age).

The treatment decided upon was to use raw vocabulary scores in the parent groups, and to derive scores in terms of age for the groups of children separately as follows:

(1) Plot the scores of children from eight to twelve in the two groups separately, and draw empirically a straight regression line as nearly as possible through the mean scores of the successive age groups. (This proved to be unexpectedly easy to do, as the line through the actual means in both groups was almost straight.)

(2) Take as each child's score his deviation from the mean for his age (interpolated for his exact age in years and months).

Correlations were then computed for the different age groups, and the results averaged for a measure of probable relationships.

The numbers are so small in each age group that the separate correlations fluctuate considerably, but the averages are probably about as stable as correlations would be based upon a number of cases equal to the total number entering the age-group correlations.

TABLE XXXVII.—CORRELATIONS WITH CHILDREN'S VOCABULARY

Age	8		9		10		11		12		Weighted average	
	r	N	r	N	r	N	r	N	r	N	r	N
Foster Group												
Father's vocabulary	.38	23	.25	16	.03	17	.45	8	.44	8	.28	72
Mother's vocabulary	.24	29	.53	19	.13	16	.49	10	.42	11	.34	85
Whittier index	.03	29	.24	20	.35	15	.39	12	.62	10	.25	86
Culture index	.17	25	.51	16	.17	15	.27	8	.64	10	.32	74
Books in home library	.08	28	.39	17	.05	15	.32	11	.41	11	.22	82
Books in child's library	.39	27	.42	16	.25	16	-.18	11	.41	10	.29	80
Hours of home instruction at ages 4 and 5	-.13	25	.59	17	.49	13	.34	9	.43	10	.28	74
Child's I. Q.	.82	29	.77	19	.75	17	.91	11	.95	11	.82	87
Control Group												
Father's vocabulary	.61	13	.28	7	.12	8	.77	7	.73	7	.51	42
Mother's vocabulary	.42	12	-.23	7	.30	8	.22	7	-.36	7	.12	41
Whittier index	.24	13	.62	7	-.14	8	.22	7	.46	7	.26	42
Culture index	.40	12	.84	7	.20	8	.23	7	.04	7	.34	41
Books in home library	.37	13	.01	7	.17	7	.14	7	.01	7	.17	41
Books in child's library	.31	13	.51	7	.49	8	.23	7	-.02	7	.31	42
Hours of home instruction at ages 4 and 5	-.01	12	.34	6	-.69	8	-.61	7	-.27	7	-.24	40

The figures of Table XXXVII suggest, at least, that vocabulary is more susceptible to the influence of environment than is intelligence (measured by the Stanford Binet as a whole). It should be pointed out that in the age columns for each group there is a significant correlation of *errors*. For example, in the twelve-year column of the Control Group, nearly all the correlations are low or negative, owing to one low-testing child who happens to be in a very superior environment, and one high-testing child situated in a less favorable environment.

To secure reliability coefficients for vocabulary so that some of these correlations might be corrected for attenuation, the following procedure was employed. In the case of the children, derived scores for one list of the vocabulary test were correlated against derived scores for the other list, separately by ages for both the

Foster and Control Groups; the results were averaged for all ages within the two groups and, since the two determinations differed by only a few hundredths, the final results for the two groups were averaged. In the case of the adults, one vocabulary list was correlated against the other list separately for fathers and mothers in the Foster Group, and as the results were the same to two decimal places, it was assumed that the same reliabilities would hold for the control parents, who matched the foster parents closely in intelligence. In both cases the Spearman-Brown correction was used, yielding final reliabilities of .93 for the children and .96 for the adults.

TABLE XXXVIII.—CORRELATIONS WITH CHILDREN'S VOCABULARY, CORRECTED FOR ATTENUATION

	Foster	Control
Father's vocabulary	.30	.54
Mother's vocabulary	.35	.13
Whittier index	.27	.29
Culture index	.34	.36
Child's I.Q.	.93	....

The raw correlations upon which Table XXXVIII is based are the weighted averages given in Table XXXVII; consequently, the probable errors are indeterminate.

3. Home Reading Done by Child

The following correlations are among the hardest to interpret in the entire study. They are consistent only in showing a significant relationship between the child's I.Q. and the amount of reading he does. The source of the measures used was the mother's information blank.

4. Moral and Personality Traits

With the exception of the results from the Woodworth-Cady questionnaire, which was administered only to the older children, the results reported here are based upon subjective judgments of the parents and of the field workers. Limitations upon time and funds made a thorough study of moral development out of the question.

TABLE XXXIX.—CORRELATIONS WITH HOME READING DONE BY CHILD

	Home reading done before nine by seven- and eight-year-old children					
	Foster			Control		
	r	P.E.	N	r	P.E.	N
I.Q. of child	.20	.11	38	.15	.14	22
Hours of home instruction at ages 6 and 7	.03	.11	38	−.28	.13	22
Number of books in home library	.06	.11	36	−.18	.14	22
Number of books in child's library	.24	.12	34	−.23	.14	21
	Home reading at nine and ten by children nine and ten					
	Foster			Control		
	r	P.E.	N	r	P.E.	N
I.Q. of child	.58	.13	12	.64	.14	8
Hours of home instruction at ages 6 and 7	.76	.08	12	−.27	.20	10
Number of books in home library	.16	.19	12	.83	.07	9
Number of books in child's library	.34	.18	11	.29	.20	10
	Home reading at nine to twelve by children eleven and twelve					
	Foster			Control		
	r	P.E.	N	r	P.E.	N
I.Q. of child	.41	.14	15	.39	.17	12
Hours of home instruction at ages 6 and 7	.52	.13	15	.14	.19	12
Number of books in home library	.65	.10	15	.13	.19	12
Number of books in child's library	.44	.14	15	−.08	.19	12

It will be recalled that in every family the two parents were asked to rate the child independently upon ten traits. To give an approximate idea of the reliability of such ratings, the correlations between independent ratings by fathers and by mothers are reported in Table XL. Ratings for which there was any reason to suspect collusion between the parents were thrown out of this set of correlations. As a rule, the parents filled the blanks out alone, or with the help of the field visitor, at the time they took the Stanford Binet Test. When it was necessary for the field visitor to give assistance in the use of the rating scales, care was taken not to suggest in any way the ratings which the child in question might probably merit.

TABLE XL.—CORRELATIONS BETWEEN PARENTAL TRAIT RATINGS OF CHILD (Father's rating vs. mother's rating)

	Foster			Control		
	r	P.E.	N	r	P.E.	N
1. Will power and perseverance	.25	.06	121	.23	.07	96
2. Cheerfulness and optimism	.34	.05	120	.30	.06	95
3. Musical appreciation	.55	.04	121	.52	.05	94
4. Sense of humor	.20	.06	121	.27	.06	96
5. Permanency of moods	.21	.06	121	.34	.06	94
6. Leadership	.55	.04	117	.39	.06	95
7. Sympathy and tenderness	.44	.05	122	.49	.05	95
8. Conscientiousness	.42	.05	118	.43	.06	94
9. Originality	.32	.06	121	.41	.06	93
10. General intelligence	.43	.05	121	.54	.05	95

These ratings cannot, of course, be taken as real reliability coefficients, since it is possible that the correlations are partly accounted for by 'collusion' over many years of child-rearing, even though collusion in actually making the ratings for this study was not present. 'Halo' effect is another possible influence that renders the interpretation of these ratings ambiguous.

The traits 'cheerfulness and optimism,' 'sympathy and tenderness,' and 'conscientiousness' showed a fair agreement for parental rating, and accordingly these three traits were chosen as variables on which to compute correlations with the mid-parental rating of the parents by the *field visitors* on the trait 'kindliness, sympathy, and tact.' Only those cases were used in which both parents had been rated by the field workers and in which collusion between parents in making their ratings was improbable. Table XLI summarizes these correlations and indicates results of no measurable significance.

The trait 'musical appreciation' was next considered. In the space provided for 'interests or hobbies' of the parents on the parents' information blanks, music is one of the hobbies suggested as an example. Music was often mentioned by parents as an interest or hobby, and occasionally such occupation as 'church organist' was noted in the space calling for 'positions of honor or trust' held by parents. Using information gained from the information blanks, biserial correlations were computed between the average rating of children by their parents in musical appreciation and the presence or absence of music as a hobby or interest of the parents.

TABLE XLI.—CORRELATIONS OF MID-PARENTAL 'KINDLINESS, SYMPATHY, AND TACT,' AS RATED BY FIELD WORKERS WITH CERTAIN TRAITS OF CHILD

Trait	Foster			Control		
	r	P.E.	N	r	P.E.	N
Cheerfulness and optimism	−.11	.06	122	.02	.07	94
Sympathy and tenderness	−.04	.06	122	.09	.07	94
Conscientiousness	.11	.06	125	.24	.07	93

Whether these slight correlations should be accounted for through environment or through bias in the parents' ratings is difficult to say.

Scores for the Woodworth-Cady questionnaire are available upon only 27 foster and 14 control children. In consequence, the

TABLE XLII.—CHILD'S MUSICAL APPRECIATION CORRELATED WITH MUSICAL INTERESTS OF PARENTS

(Biserial r's)

	Foster			Control		
	r	P.E.	N	r	P.E.	N
Either father *or* mother gives music as interest: blanks from both parents were secured, but neither mentions music	.19	.06	164	.17	.09	99
Father's blank mentions music: father's blank does not mention it	.24	.08	121	.12	.08	101
Mother's blank mentions music: mother's blank does not	.21	.07	156	.12	.08	101

correlations using Woodworth-Cady score as a variable are not very reliable, but it seems significant that in Table XLIII the coefficients for the Foster Group and the Control Group agree perfectly as to sign, and approximately as to magnitude.

As a low score on the Woodworth-Cady questionnaire signifies high emotional stability and a high score signifies emotional instability, negative correlations in this table imply a favorable influence of the factors in question, and positive correlations imply an unfavorable influence. The correlations of Table XLIII suggest, at least, that from the point of view of a child's emotional stability, it is desirable to have kindly, sympathetic parents who give him little supervision.

TABLE XLIII.—WOODWORTH-CADY SCORES (IN TERMS OF AGE NORMS) CORRELATED WITH VARIOUS HEREDITARY AND ENVIRONMENTAL FACTORS
(Children ten or over)

Factor	Foster r	Foster P.E.	Foster N	Control r	Control P.E.	Control N
I. Q. of child	−.12	.13	27	−.17	.18	14
Whittier index	.15	.13	26	.46	.14	14
Culture index	.11	.14	24	.37	.16	14
Mid-rating of parents by field visitors on 'kindliness, sympathy, and tact'	−.29	.12	26	−.30	.16	14
Parental supervision (biserial using ratings below 5 *vs.* ratings of 5 or 6)	.34	.16	26	.24	.23	14
Telling child of his adoption (biserial)	.18	.16	27	....	....	....

It may also be of interest to compare the mean scores of the foster and control children to the Woodworth-Cady norms, based upon tests of unselected children of 10 to 14 years.[13] Only 4 of the Foster Group of 27 have scores (in standard units) which deviate positively from the mean for their age and sex (*i. e.*, only 4 show more than average instability on this test). The mean standard score for the foster children is − .48. Only one of the Control Group of 14 has a positive standard score, and the mean of their standard scores is − .68. May these results possibly mean that the superior mental level and the unusually kindly and sympathetic parents of these groups have more than offset the unstabilizing effect which a maximal amount of care and supervision might otherwise have induced?

These figures complete the statistical treatment that has been given to the special traits which have been discussed in this sec-

[13] (See *Genetic Studies of Genius,* I, p. 511. (Ref. 15).

tion. However, since most of the questions here taken up have been based upon subjective material, and since this section offers suggestions rather than conclusions, a few additional comments may be pertinent as well as interesting.

I was much impressed by the fine spirit of rapport between parents and children which was nearly always to be found in both foster and control families, and apparently to about the same degree in both. The foster parents were perhaps more demonstrative toward their children. Time after time, they told the field visitors that taking a child to rear had transformed their lives and that they were perfectly sure that they could have felt no greater love for children of their own.

Hence, here we have two groups of parents, both of which exceed, on the average, the parents of unselected offspring in intelligence, education, culture, and character. One group of parents is rearing children conceived of its own germplasm; the other group is rearing the children chiefly of parents who, in their parenthood itself, showed a lack of inhibition or disregard of consequences. Yet the foster children, quite as noticeably as the control children, are strikingly free from serious behavior problems. Only two of sixty-five foster children visited and tested by the writer appeared to present such problems. It may be, of course, that the 'company manners' of many of the families tended to cover up such problems if they existed, but the internal evidence of several hundred hours of interviews with foster mothers and fathers made me morally positive that in the great majority of cases the family ties, marital and parental, were rather ideal.

The two cases mentioned were both five-year-old girls who were disobedient and subject to tantrums. One of these children was shy, entirely unresponsive to strangers, and easily moved to tears at kindergarten as well as at home. She was apparently somewhat neurotic, and it seems questionable whether environment will ever be able to compensate for an apparently congenital instability. The other was sunny, friendly in disposition, and seemed thoroughly normal in her behavior toward strangers. A few months after testing her, I learned that she had apparently begun to overcome her temperamental difficulties, probably as the result of some gentle, but effective, aids to self-control that her parents were using with her at the time.

In addition to these two cases there were two others who seemed to lack an entirely harmonious relation with their foster mothers. They did not present specific behavior problems, but the bond between mother and child fell far short of perfection. In both instances the fault seemed to rest in part with the mother and in part with the child. One child was a boy of seven whose foster mother found him "exasperating." Probably most boys of seven do have exasperating traits, and this child perhaps had more than most, though he was normal with respect to intelligence (I.Q. 109). But this mother had allowed the child to "get on her nerves." The situation may have been partly accounted for by the fact that she was harassed by the duties connected with several boarders whom she provided for in her small home to help eke out the family income. The other child was a bright girl of fourteen (I.Q. 113) who possibly tended a little more toward selfishness and 'headiness' than the average girl of her age, but the situation was undoubtedly aggravated by a foster mother who utterly failed to understand her. Small wonder, for the mental age of this foster mother was less than eleven. The mental age of the daughter actually exceeded that of the mother by more than five years![14]

On the whole, the paucity of maladjustments in the Foster Group seems significant, especially when it is remembered that there is only slightly more tendency than chance alone for the foster parents and children to resemble one another in mental quality. It indicates that there are many points of happy human contact other than those based upon congeniality of intellect.

Perhaps it is only natural that, given a home in which harmony and affection prevail, and given a child who is not suffering from some form of actual mental pathology, a fine and sound growth of the parent-child affection should result.

The comments of the last few paragraphs apply, obviously, only to a group of children under fifteen years of age. As the children reach adolescence and adulthood, problems not appearing in the younger group may arise. What will be the attitude of parents of unusual education and culture who find that the foster child whom they have adored since babyhood is unable to take a college education? Will rising young professional men and women remain loyal to foster parents whose mental level is many degrees below their own—to good, honest, unimaginative foster parents who can admire and applaud, perhaps, but never fully appreciate the ambitions and achievements of the young people they have reared? On these questions, our investigation throws no light. The possibility of such questions arising suggests that in child-placement work of the

[14] This case is reported in more detail in the appendix of case studies.

future every aid that science can offer should be brought to bear on each placement.

5. School Progress

All figures upon age-grade placement were computed for mid-term,[15] *i. e.*, in every case the age and mental age of any given child were projected forward or backward from the date of his test so as to give the age and mental age he would have at mid-term, and thus insure comparability in the data on grade placement of each child. The ages of children tested in the summer were projected ahead to the mid-term of the fall semester. In the figures upon acceleration and retardation, Ayres-Strayer norms were used. By these norms a child is considered at age for grade if he is six or seven in the first grade, seven or eight in the second grade, eight or nine in the third grade, etc.

TABLE XLIV.—SCHOOL ACCELERATION AND RETARDATION IN FOSTER AND CONTROL GROUPS*

Group	Comparison	Accelerated		Correctly placed		Retarded	
		N	Percent	N	Percent	N	Percent
Foster. .	With respect to C.A.	23	17	109	80	5	4
	With respect to M.A.	16	12	93	68	28	20
Control	With respect to C.A.	18	26	51	73	1	1
	With respect to M.A.	5	7	34	49	31	44

*Ayres-Strayer norms of grade placement used and children aged 6 to 14, inclusive.

The higher percentage of acceleration by chronological age in the Control Group, and the smaller percentage by mental age, are undoubtedly accounted for by the superior mental level of the control (115 I.Q. as against 107 I.Q. in the Foster Group). The correlation between school grade and chronological age (in half-year intervals) is .95 in the Foster Group and .96 in the Control Group, with the P.E. in both cases .01.

We also tried to find out whether or not any environmental factors appeared to have affected the grade placement of the foster children. Allowing only a six months span as 'normal' for each half grade and assigning such 'normal' placement a score of zero, the amount of acceleration and retardation in terms of half grades for foster children six or over was correlated with the mental age

[15] "Term" as here used refers to half a school year.

of foster fathers and of foster mothers. The correlations turned out to be virtually zero; accordingly, the conclusion seemed justified that this phase of environment has little to do with speed of progress through the grades.

VIII. Interpretation and Conclusions

The way has now been cleared to answer, if possible, the questions regarding the relative contributions to intelligence of nature and nurture which were raised in the beginning of the study. The interpretation of results to be presented will embrace the following aspects of the problem of factors conditioning children's intelligence:

1. Proportional contribution of total home environment to variance.
2. Unique contribution of parental intelligence to variance.
3. Estimate of total contribution of heredity to variance.
4. Numerical estimate of the potency of home environment to raise or depress the I.Q.

1. Proportional Contribution of Total Home Environment

Considering the correlations which have been reported, we have logical ground for believing that the multiple correlation corrected for attenuation (.42) is a measure in the Foster Group of the effect of home environment upon differences in children's intelligence. More precisely, the *square* of this multiple (.17) represents the portion of the variance of children in ordinary communities that is due to home environment.[16]

[16] In discussing the portion of the *variance* of the children due to this factor and that one, I follow Fisher (3). The justifications for dealing with contribution to variance (*i. e.*, squares of the S. D.) rather than with contributions to the first power of the standard deviation or to any other power are: (a) that such contributions to variance combine additively to give the total variance of the criterion, but contributions to any other power of the S. D. do not; and (b) contributions to variance, but not to other measures of variability, can readily be interpreted by a concept of the proportional number of *common factors* underlying the influences and the criterion. For example, a criterion composed of four equally variable factors, a, b, c, and d, correlates $1/\sqrt{4}$, or ½ with any of the 'influences'—a, b, c, or d. The square of the correlation ½, or ¼, gives the contribution of each factor to the variance of the criterion, and expresses the proportion of factors in the criterion contributed by each factor.

In this connection may be cited a paper by Pearson, "On Certain Errors with Regard to Multiple Correlation Occasionally Made by Those Who Have Not Adequately Studied this Subject" (11). In this article Pearson demonstrates that nearly the maximal predictivity, with respect to a criterion, of a large group of variables all showing considerable correlation among themselves is attained when only a few of such variables are used in a multiple correlation. It follows that the square of our multiple probably represents nearly the maximal effect of home environment, especially since various factors of home environment that were not used in the final multiple could be legitimately dropped out because they were found to contribute to I.Q. variance practically nothing in addition to the contribution of the variables retained.

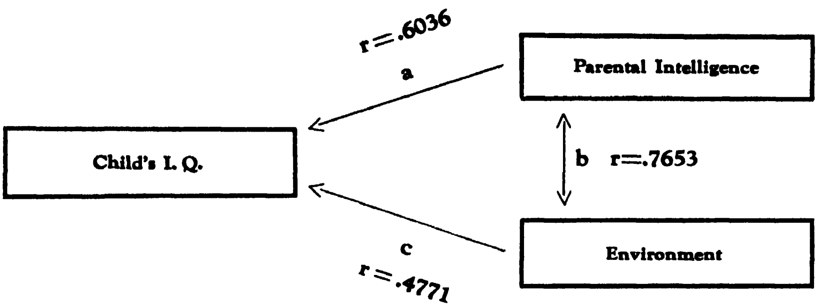
2. Unique Contribution of Parental Intelligence

In the Control Group, the square (.37) of the multiple correlation corrected for attenuation (.61) represents the *combined* effect of home environment and parental mental level upon the variance of children's intelligence. Neglecting the variable 'father's vocabulary,' which contributes only an insignificant amount in addition to father's mental age, it is extremely interesting to apply the Wright path coefficient technique to the correlations for the Control Group, to find out how much of the children's I.Q. variance can be accounted for by reference to parental intelligence alone.[17]

This situation is a particularly favorable one for using the Wright technique, for the assumptions regarding casual relation-

[17] The path coefficient method, to quote Sewall Wright (20) "depends on the combination of knowledge of the degrees of correlation among variables in a system with such knowledge as may be possessed of the causal relations." The method is limited by the rarity with which we have actual knowledge of causal relations; but it provides a tool of the nicest precision in such situations as do offer an adequate basis for postulating causation. It cannot, itself, uncover what is cause and what is effect, though in the absence of definite knowledge regarding causal relationships between variables, the method "can be used to find out the logical consequences of any particular hypothesis in regard to them." Conservatively stated, in any situation in which we feel justified in drawing conclusions regarding the effects of certain phenomena upon others, the Wright method provides a numerical expression of such conclusions. For a detailed explanation of its application, the reader is referred to two articles by Wright (20) (21).

ships are here at a minimum. It is only necessary to assume that parental intelligence and home environment affect the child's I.Q., but that the child's I.Q. does not contribute to these. It is not necessary to make any assumption at all regarding a possible casual or interacting relationship between parental intelligence and environment; merely the known correlation between the two is sufficient. The relation between the variables is represented in the 'set-up' shown herewith, in which



r (parental intelligence) (I.Q.) is the multiple, corrected for attenuation, between the I.Q. of the child and the mental ages of the two parents;

r (parental intelligence) (environment) is the multiple, corrected for attenuation, between the Whittier index and the mental ages of the two parents; and

r (I.Q.) (environment) is the correlation, corrected for attenuation, between the child's I.Q. and the Whittier index.

a represents the direct path of influence between parental intelligence and child's I.Q. and a^2 the percentage of I.Q. variance attributable to parental intelligence. The coefficients c and c^2 represent corresponding coefficients of environmental influence other than that reflected in a and a^2. The coefficient b represents the known correlation between parental intelligence and environment.

The directions of the arrows indicate the relationship of the variables with respect to cause, effect, and possible reciprocal action.

By Wright's formulas:

$$r_{\text{(I.Q.) (parental intelligence)}} = .6036 = a + bc$$

$$r_{\text{(I.Q.) (environment)}} = .4771 = c + ab$$

$$r_{\text{(parental intelligence) (environment)}} = .7653 = b$$

Solving these three equations for the two unknowns,

$$a = .5757 \qquad a^2 = .3314$$

$$c = .0367 \qquad c^2 = .0013$$

In addition to a and b, there is an effect upon the child's I.Q. due to the *combined* working of the two correlated variables, parental intelligence and environment. This effect is equal to 2abc, or .0322, and expresses the minute increment of variance, over and above what each variable contributes by itself, that results from the fact that the two variables are correlated and reënforce one another to some extent.

$a^2 = .3314$	(parental contribution)
$c^2 = .0013$	(contribution of environment *other* than parental intelligence)
$2abc = .0322$	(joint parental and environmental contribution over and above separate contribution of each)
$.3649$	

The sum of a^2, c^2, and 2abc is equal to the square of the multiple correlation of I.Q. with parental intelligence and environment.

A question of great interest concerns the difference between the contribution (.17) of total environment in the Foster Group and the contribution (.0013) of environmental influence (other than the direct influence of parental intelligence) in the Control Group. This difference is probably due to several facts, viz:

(1) The environmental as well as the hereditary contribution of parental intelligence is contained in a^2, and is consequently lacking in the value, .0013, of c^2. We should not expect this *environmental* contribution of parental intelligence to be over four or five percent, however, because the correlations (even when corrected for attenuation) between child's I.Q. and foster parents' M.A. are so very low (see Table XXXIV). The correlation squared is .0081 with foster father, and .0529 with foster mother; and both these values represent more than the unique foster parent contri-

butions, because they are increased by the relationship of parental M.A. to other influences of environment.

(2) Part of the joint parental and environmental contribution (.0322) should be properly attributed to environment when a comparison is made of environmental influences in our two groups.

(3) The probable errors of our determinations of degrees of influence could well account for the remaining discrepancy. It is not known exactly what the magnitude of these probable errors is; however, we do know that the P.E. of the multiple correlation (corrected) of child's I.Q. with environment in the Foster Group is greater than .05; and that the P.E. of path coefficients based upon corrected correlations of the size of the ones entering our calculations is fully as large as that.

3. Estimate of Total Contribution of Heredity

As has been noted above, a^2, or 33 percent, represents the proportion of I.Q. variance that is attributable to parental intelligence alone. Now, 37 percent is the proportion of I.Q. variance that we have already found attributable to parental intelligence and environment, alone and in combination. It follows that, if we could, without at the same time narrowing the range of parental intelligence, level all other aspects of home environment to a standard or average, the variance of children's intelligence would be reduced by 37 minus 33, or 4 percent. The contribution of parental intelligence to variance would then be equal to .33/.96, or 34 percent.

Such a contribution corresponds to a multiple correlation of $\sqrt{.34}$, or .58,—the multiple correlation (corrected for attenuation) which would be found between child's I.Q. and parental intelligence if the home environment of all families were made constant, but *parental intelligence continued to vary as much as before.* In this latter respect the coefficient differs radically in theory from the partial correlation coefficient, which in comparable situations has sometimes been interpreted erroneously (1). The partial correlation of I.Q. and parental intelligence with environment constant is here only .42, as contrasted with .58.

The value .58 probably represents fairly closely the actual degree of resemblance between children and their two parents based upon heredity alone. The undoubted fact that a small amount of parent-child resemblance due to environment, but not measured by the Whittier Scale, is still concealed in the coefficient probably enhances its value slightly. But the fact that parents were themselves molded in part by environment, and in consequence vary somewhat from their congenital mental level, and the further likelihood of slight random *environmental* effects (such as those from pre- and post-natal nutrition), suggest that the intrinsic genetic resemblance between parents and offspring is somewhat depressed thereby. The elevating and depressing effects undoubtedly cancel one another to some extent. The coefficient .58 can consequently be taken as a tentative approximation to the true genetic relation. Its probable error could in this case be computed similarly to the probable error of a regression coefficient if the coefficient .58 were not based upon correlations corrected for attenuation. It can only be observed that its probable error must be somewhat greater than the probable error (.06) of an ordinary regression coefficient based upon raw intercorrelations equivalent to the corrected ones used here.

We have now seen that the total contribution of systematic (or measurable) home environment is close to 17 percent, and that the contribution of home environment and parental intelligence together is represented by a multiple correlation coefficient (corrected) of .61, or by a percentage of .37. If not more than 35 or 40 percent of the variance of children's I.Q.'s is accounted for by reference to these factors, what contributes the other 60 or 65 percent?

Possibly a portion of this residual variance is due to the "random somatic effects of environment," to quote Fisher (3). But it seems reasonable to suppose that not a great deal is due to this effect, since numerous studies have shown a marked tendency for the I.Q. to remain constant over a period of years, while other studies have shown that identical twins correlate in intelligence about as closely as the reliability of the tests employed will permit (10). Probably the major share of the residual variance is due to congenital endowment, since in known modes of hereditary transmission the influence of heredity is *always* far stronger than parental correlations alone would indicate. This is necessarily the case because only half the chromosomes of each parent are passed on to the offspring. Hence, the parental deviation for any trait in question is determined by a number of factors other than the ones transmitted to the child. In hereditary traits such as stature, which are known to be influenced relatively little by ordinary differences in environment, the multiple correlation of child with parents is .64, but the contribution of heredity to variance approaches 100 percent. The closeness of our estimated value of the "genetic" multiple correlation for intelligence to this value of the multiple correlation for stature is striking. *Probably, then, close to 75 or 80 percent of I.Q. variance is due to innate and heritable causes.*

This estimate makes allowance for the 17 percent which the data of this study show is due to measurable home environment, plus an additional 5 or 10 percent due to the possible "random somatic effects of environment." In the opinion of the writer, the estimate is the most reasonable one that can be made from available data with available methods. But a determination of the total contribution of heredity can probably never be made beyond cavil until the genetic mechanics of mental heredity are first established by methods analogous to those used by Fisher in the study of physical traits (3).

4. A Numerical Estimate of the Potency of Home Environment to Raise or Depress the I.Q.

One further angle of interpretation will be especially pertinent to the general problem of the possibilities and limitations of training. From a practical outlook the point to be raised is undoubtedly of even greater significance than the more general problem of the proportional contributions of nature and nurture to mental variability. It is concerned with the question: "How far, in terms of measurable I.Q., is environment potent to increase or inhibit the development of innate intelligence?"

Let us turn to the data of Section V. It was there seen that empirical considerations, based upon facts given, strongly suggested that the 'congenital mental level' of the foster children was not more than two or three points above 100 I.Q. But the average I.Q. level actually found in this group was 107. Can this discrepancy be accounted for through superior environmental advantages?

Probably it can be. The average mental age level of the foster fathers is 16 years, 11 months, and of the foster mothers is 16 years, 3 months. The average *mid-parent* level, 16-7, is about one standard deviation above that of parents in general.

The army intelligence data (22) strongly imply that the average adult mental level of Americans is closer to 14 years than to the 16-year level which had been tentatively established previously. But the Army Alpha group test was different in many respects from individual tests, and psychologists have hesitated to assume without further evidence that the same outcome would necessarily hold for tests of the Binet type. However, our control data rather bear out the army conclusions when treated in the following manner: Summary cards for the cases were arranged from lowest to highest in order of father's mental age. Starting with the first case, the children's I.Q.'s were added and averaged as each additional case was inserted. When a point was reached at which the children's I.Q.'s averaged as close to 100 as our limited number of cases permitted (within three points of 100), the fathers' and mothers' mental ages for those cases were averaged separately and together with the following result:

		N
Fathers' mental age	12.9	21
Mothers' mental age	14.5	21
Average	13.7	

The same procedure was repeated with cases in which mothers' mental age was arranged from lowest to highest, with the result:

		N
Fathers' mental age	14.6	20
Mothers' mental age	12.4	20
Average	13.5	

Finally, first with the fathers and then with the mothers, and starting with 13.5 as a median, paired cases in which parent scores showed equal positive and negative deviations from 13.5 were selected until all possible pairs had been used. The average of fathers and mothers was 13.8 in the first instance and 14.1 in the second instance. The corresponding average I.Q.'s of children were 105 and 104, respectively, suggesting that 14 years may be a little high to represent the average adult level. But it seems justifiable on the basis of the foregoing to use 14 for an approximation to the truth. As the standard deviation of our mid-parent mental age is close to two years, the average level of the control parents (and similarly of the foster parents) is about one standard deviation superior.

It is difficult to say just how high above the mean of the generality are the other environmental measures (culture index, Whittier index, income, etc.) because no satisfactory norms for unselected populations are available upon them. Since most of the correlations between the measures of environment and the measures of parental intelligence are quite high, a safe estimate would be that the total complex of environment (including parental intelligence) is between one half and one standard deviation above average.

The multiple correlation (corrected for attenuation), .42, can now be used as a regression coefficient for predicting the average standard score of the Foster Group. A positive increment of .42 times one standard deviation (or 15 I.Q. points) would equal 6 I.Q. points; or times one half a standard deviation would equal 3 I.Q. points. An increment of 3 to 6 I.Q. points would bring the I.Q. level of our foster children very close to that actually found (107), provided my judgment is correct that their average innate intelligence is about 102 or 103.

We may now go through some of the variables which were correlated with the I.Q's of the foster children and ascertain, when various factors of environment are, say, one standard unit above or below the mean of American communities, how much the I.Q's of the children have been shifted from their "congenital" value in consequence. The column in the following table headed "Measured" is based upon raw correlations, and the column headed "Actual" is based upon correlations corrected for attenuation. Correlations used for the computations are those reported in Tables XXXI, XXXIV, and XXXVI. The values of Table XLV are

TABLE XLV.—AVERAGE SHIFT, DUE TO ENVIRONMENT, IN POINTS OF I.Q., OF FOSTER CHILDREN, WHEN VARIOUS FACTORS ARE ONE S. D. ABOVE OR BELOW THE POPULATION MEAN

Factor	Measured	Actual
Foster father's mental age	1.0	1.4
Foster mother's mental age	2.9	3.5
Foster mid-parent mental age	3.0	....
Whittier rating of foster home	3.1	3.6
Culture rating of foster home	3.7	4.4
Total environment	5.3	6.3

found merely by multiplying the correlations of foster children's I.Q.'s with the factors in question by the S.D., 15, of the children's I.Q.'s

The implications of this table seem to the writer of more profound significance than those of any other part of the study. While the intercorrelations between these environmental factors are so complex that the relative influences of the separate factors are probably not represented linearly by the differences in the corresponding I.Q. "shifts," the *order* of their influence is probably so indicated. From this argument two outstanding conclusions emerge:

1. *The total effect of environmental factors one standard deviation up or down the scale is only about 6 points, or, allowing for a maximal oscillation in the corrected multiple correlation (.42) of as much as .20, the maximal effect almost certainly lies between 3 and 9 points.*

2. *Assuming the best possible environment to be three standard deviations above the mean of the population (which, if "environments" are distributed approximately according to the normal law, would only occur about once in a thousand cases,) the excess in such a situation of a child's I.Q. over his inherited level would lie between 9 and 27 points—or less if the relation of culture to I.Q. is curvilinear on the upper levels, as it well may be.*

An influence of this magnitude, although significant, is emphatically not sufficient to account for genius upon a theory of environment. Francis Galton, whose I.Q. in childhood Professor Terman has estimated to have been close to 200 (16), was reared in a home of exceptional cultural advantages. Yet even without the possible 9 to 27 points contributed by his environment, he would still have ranked as a genius such as occurs in unselected populations only once in many thousands of individuals. Whether or not he would have succeeded in using his gifts with such telling effect if he had not had the training, education, and inspiring associates that were his, is of course another question. While many men and women have surmounted unbelievable obstacles to achieve eminence, there is no telling how many others, of weaker stamina, have crumpled by the way.

It is of further interest to note that, while the environmental conditions of gifted men, women, and children indisputably show a somewhat superior tendency, they are not, as a rule, so exceptional as those to which the fortunate young Galton was born. The average Barr rating of fathers of the California gifted children studied by Professor Terman is 12.77—a value close to that of the foster fathers and of the control fathers. Thus, the superiority of the gifted group must be due preponderantly to endowment and, on an average, less than 10 points of I.Q. must be due to environment. *Home environment in the most favorable circumstances may suffice to bring a child just under the borderline of dullness up over the threshhold of normality, and to make a slightly superior child out of a normal one; but it cannot account for the enormous mental differences to be found among human beings.*

If environment cannot account for men, like Galton, who far and away outstrip the majority of their fellows coming even from such a favorable environment as theirs, still less can it account for an impressive number of eminent men whose early conditions of life have been of the kind that depress rather than enhance the I.Q.: men like Lincoln of the backwoods; Carlyle, whose simple peasant mother learned writing while he was at college so that she might correspond with him; Dickens, whose nursery was a London slum; or Canning, a neglected little boy who "longed for bread and butter" as he followed the ragged fortunes of a band of strolling players in eighteenth century England.

5. Summary of Conclusions

By methods which have permitted the effects of environment to be studied separately from those of heredity in conjunction with environment, this study has sought to evaluate the factors conditioning the intelligence of a group of white American school children living in ordinarily variable circumstances. The main conclusions thereby reached are as follows:

1. Home environment contributes about 17 percent of the variance in I.Q.: parental intelligence alone accounts for about 33 percent.

2. The total contribution of heredity (i. e., of innate and heritable factors) is probably not far from 75 or 80 percent.

3. Measurable environment one standard deviation above or below the mean of the population does not shift the I. Q. by more than 6 to 9 points above or below the value it would have had under normal environmental conditions. In other words, nearly 70 percent of school children have an actual I.Q. within 6 to 9 points of that represented by their innate intelligence.

4. The maximal contribution of the best home environment to intelligence is apparently about 20 I.Q. points, or less, and almost surely lies between 10 and 30 points. Conversely, the least cultured, least stimulating kind of American home environment may depress the I.Q. as much as 20 I.Q. points. But situations as extreme as either of these probably occur only once or twice in a thousand times in American communities.

5. With regard to character and personality traits, upon which the data presented are less reliable and less objective than those upon intelligence, the indications are that environment is at least as potent as in the case of intellectual traits—possibly much more potent.

A more comprehensive study of such traits, however, must await the future. Whatever clear contribution is made to the general nature-nurture problem by this investigation must rest only upon the data which deal with intelligence. On this point, it is believed that the study finds support for the conclusion reached by the first pioneer to study mental heredity by statistical methods—that heredity is a force in the determination of mental ability by the side of which all other forces are "dwarfed in comparison."

IX. References

(1) Burks, Barbara S. "On the inadequacy of the partial and multiple correlation technique." *Jour. Ed. Psych.*, 17: 1926, 522-540.

(2) DeVoss, James C. "*The Unevenness of the Abilities of Gifted Children in California.*" Ph.D. dissertation, Stanford University, 1924.

(3) Fisher, R. A. "Correlation between relatives on the supposition of Mendelian inheritance." *Trans. Royal. Soc. Edinburgh*, 52: 1918, 399-433.

(4) Gordon, Kate. "The influence of heredity in mental ability." Children's Dept., Calif. State Board of Control, 4th Bienn. Rept., 1918-1920.

(5) Gordon, Hugh. "Mental and scholastic tests among retarded children." Ed. Pamphlets, No. 44, Bd. of Ed., London, 1923.

(6) Herring, John P. "Reliability of the Stanford and the Herring revision of the Binet-Simon tests." *Jour. Ed. Psych.*, 15: 1924, 217-223.

(7) Hildreth, Gertrude H. *Resemblance of Siblings in Intelligence and Achievement.* New York, 1925, (Teachers College, Columbia, Contributions to Education, No. 186).

(8) Kelley, T. L. "Note on the reliability of a test." *Jour. Ed. Psych.*, 15: 1924, 193-204.

(9) Kelley, T. L. *Statistical Method.* New York, 1924.

(10) MERRIMAN, CURTIS. "The intellectual resemblance of twins." *Psych. Monographs*, 33: 1924, pp. 58.

(11) PEARSON, KARL. "On certain errors with regard to multiple correlation occasionally made by those who have not adequately studied this subject." *Biometrika*, 10: 1914, 181-187.

(12) POYER, G. *Problèmes Généreaux de l'Hérédité Psychologique.* Paris, 1921.

(13) RICHARDSON, L. F. "Measurement of mental 'nature' and the study of adopted children." *Eug. Rev.*, 4: 1912-13, 391-394.

(14) RUCH, FLOYD. *On the Defensibility of Certain Abridgments of the Stanford Revision of the Binet-Simon Tests.* M.A. thesis, University of Iowa, 1925.

(15) TERMAN, L. M. and others. *Genetic Studies of Genius*, Volume I. Stanford University, 1925.

(16) TERMAN, L. M. "The intelligence of Francis Galton in childhood." *Amer. Jour. Psych.*, 28: 1917, 209-215.

(17) THEIS, SOPHIE VAN S. *How Foster Children Turn Out.* State Charities Aid Ass'n. Publication No. 165, 1924.

(18) *Whittier Social Case History Manual.* Calif. Bureau of Juvenile Research, Bull. No. 10, 1921.

(19) WILLOUGHBY, R. R. *Family Similarities in Mental Test Abilities.* Ph.D. dissertation, Stanford University, 1926.

(20) WRIGHT, SEWALL. "Correlation and causation." *Jour. Agric. Research*, 20: 1921, 557-585.

(21) WRIGHT, SEWALL. "The theory of path coefficients." *Genetics*, 8: 1923, 238-255.

(22) YERKES, R. M. (editor). *Psychological Examining in the United States Army*, 1921. (*Mem. Nat. Acad. Sci.*, 15).

APPENDIX

I. Brief Case Studies

A few brief case studies from the Foster and Control Groups are presented to give the reader an impression of the kinds of data that were available for all the cases of our groups, and to illustrate some typical situations. The names used in these notes are fictitious.

1. Foster Cases

1. One of our prize double cases. Nancy, age five, has an I.Q. of 118; Teddy, her six-year-old foster sibling has an I.Q. of 113. Both children are exceptionally friendly, well mannered and responsive, and no case was found in which the spirit of the entire family was more harmonious and delightful. The foster father is a successful professional man, mental age 17-1, and is proud of his family and devoted to it in a rare degree. The mother can be described only as fairy-like. She is gay, winsome, sympathetic, and bright. Her mental age is 18-6, and her *maternal* age surpasses all known scales. "Do the children know that they are adopted children?" the writer asked her. "They

were told as early as they could talk," the mother answered. "Some foster mothers dread to tell their children of the adoption, but personally I wouldn't miss for anything in the world the joy of having my children run up to me sometimes when they are at play and tell me how glad they are that I turned out to be their mother!" There is nothing in the heredity of these children that would forecast superior intelligence. It seems probable that what superior intelligence they show can be attributed to their very exceptional environment.

2. This family is struggling along happily on the meager income of a laborer. The foster father's mental age is 10-2, and the foster mother's mental age is 9-9, but neither parent seems to feel any sense of deficiency, and apparently they are well-adapted to the simple life they lead. Their great common interest is their love for their eleven-year-old foster daughter, who was adopted, without the knowledge of their friends, to take the place of a baby who died at birth. There are no frills in the home of these plain people, but they are doing the best they can for their little girl, and are even, at real sacrifice, providing her with piano lessons. The girl, who was the illegitimate child of a laundress, is quite normal (I.Q. 97).

3. This case is perhaps the most picturesque one found by the writer. Elaine, nine and with an I.Q. of 123, is growing up in one of the drabbest little homes of one of the shabbiest little towns imaginable. Her foster father and foster mother have mental ages of 9-10 and 12-11, respectively. Neither parent had had any education beyond the grades, nor any cultural advantages, but it must be said to their credit that they are both tremendously fond and proud of Elaine, and will probably try to help her secure the education that her mental caliber warrants. They think she would make a good musician, but the child's own vocational ambition is to be a teacher. She is a charming, poised, and friendly little person, and is outstandingly the most popular girl in her class at school. Her true mother was an eighteen-year-old factory girl. Possibly it is her true father, whose walk of life is not known, who accounts for her rare mental and temperamental endowment.

4. Billy, aged five, is feeble-minded (I.Q. 53). His foster parents, hard-working people of average intelligence and education do not yet realize the extent of the boy's deficiency, but are sure to have a bitter disappointment when they do. Nothing that is known of the boy's heredity augurs such a poor prognosis. His true father was an 'educated man,' and his true mother an 'attractive' girl who was divorced from her husband and got into an intrigue with the father of her child. The child himself is vigorous and healthy and shows no outward stigmata. Such a case as this would seem to be a strong argument for developing the use of infant mental scales as soon as it can possibly be done.

5. Mildred, age fourteen, with an I.Q. of 113, and her foster mother with a mental age of 10-11, are not very well adjusted. This is partly

because Mildred is selfish and headstrong, and partly because the mother, a nervous, complaining 'tired' woman, tries to 'run' her daughter without being able to understand her. "Mother never understands my jokes," the little girl told the writer. On the material and cultural side, the home conditions are about average. The father, a mechanic, declined to take the test.

2. Control Cases

1. The highest testing child of the Control Group is an eight-year-old girl with an I.Q. of 157. Her father is a chiropractor with mental age 19-3, and has an "exceptionally pleasing personality," according to the field visitor. The mother has a mental age of 17-1, and is also a chiropractor. The cultural conditions of the home are somewhat above average, though not exceptionally so.

2. The lowest testing child of the Control Group is a six-year-old boy with an I.Q. of 75. His father, with a mental age of 14-1, is a prosperous real estate salesman. The mother has a mental age of 11-11, but is able to add to the family income by selling lots herself. The combined income of the family is about $5000. The cultural level of the home is somewhat below average, and the only book in the family library is the Bible.

3. High intelligence in the parents does not always augur exceptional intelligence in the children. In this case a university professor and his wife exceed the limit of the Stanford Binet Scale, but their little five-year-old daughter has an I.Q. of only 102. As might be imagined, the cultural conditions in this home are decidedly superior.

II. Typical Correlation Arrays

In order to exhibit to the reader characteristic samples of the material on which the correlation coefficients discussed at length in the preceding pages have been based, and in order also to supply, as it were, a pictorial presentation of some of this material, four scatter diagrams are presented (pp. 314-317). These diagrams show the nature of the correspondence (1) between the I.Q. of the foster child and the M.A. of the foster father (2) between the I.Q. of the true child and the M.A. of its father (3) between the I.Q. of the foster child and the M.A. of the foster mother, and (4) between the I.Q. of the true child and the M.A. of its mother.

TABLE XLVI.—CORRELATION ARRAY: I.Q. OF FOSTER CHILD VS. M.A. OF FOSTER FATHER

M.A. of Foster Father	I.Q. of Foster Child 50–54	55	60	65	70	75	80	85	90	95	100	105	110	115	120	125	130	135	140	145	150	155	160
21–6										1	1		1	1		1							
21–0																							
20–6								1	4	2	2	4		4	3								
20–0										1	2	2											
19–6																							
19–0										1		2	3	4		1	2	1		1			
18–6					1	1	1	1		3	3	3	3	2	3	1							
18–0						1					1												
17–6										2	2	2	2	4	2								
17–0										4	3	3	3	1	2		1						
16–6	1						1			1	1											1	
16–0									2			1	3										
15–6								1	1		3			1									
15–0	.									1		1		1				1					
14–6								1		2	1	1	1	1	1	1	1						
14–0								2	3	2		4	1	2			2						
13–6										1	1	1	3		2	1		1					
13–0												2				1	1						
12–6						1					1		1	1									
12–0										1	1			1							1		
11–6									3														
11–0																							
10–6										1	1	1											
10–0 to 10–5										1													

r = .07 ± .05
N = 178

TABLE XLVII.—CORRELATION ARRAY: I.Q. OF TRUE CHILD vs. M.A. OF FATHER

M.A. of Father	I.Q. of True Child																	
	75–79	80	85	90	95	100	105	110	115	120	125	130	135	140	145	150	155	160
21–6						1	2		2	2		1		1				
21–0																		
20–6					1		1		2	1	1		2	1				
20–0							2	1										
19–6																		
19–0						1					1		3		1	1	1	
18–6						1			1	1	4					1		
18–0																		
17–6						1		2	1	4			2					
17–0						1	2		1	1	1							
16–6																		
16–0						2	2		2									
15–6						1		1	1	1	1							
15–0	1			1			1	1			1							
14–6					1	2			3	1		1						
14–0	1			1			2		1	1		1						
13–6				1		1		1	1									
13–0							2											
12–6							1		1									
12–0				1	2				1									
11–6																		
11–0				1	1													
10–6						1		1										

$r = .45 \pm .05$
$N = 100$

TABLE XLVIII.—CORRELATION ARRAY: I Q. OF FOSTER CHILD VS. M. A. OF FOSTER MOTHER

M.A. of Foster Mother	I.Q. of Foster Child 40–44	45	50	55	60	65	70	75	80	85	90	95	100	105	110	115	120	125	130	135	140	145	150	155	160
21–6													1	1		1	1	1							
21–0																									
20–6						1					1	2		1		1							1		
20–0												2	4		1	1	4	1							
19–6																									
19–0											1			1	2							1			
18–6										1	1		2	3	5	6			2						
18–0																									
17–6											1	2	3	3	5	3									
17–0									1	1	2	2	2	2	3	4		1							
16–6												2	1	2			1								
16–0												3	2	4		2	2		2					1	1
15–6									1					1		1									
15–0							1	1				2	3	2		1	1	1							
14–6										1	1		1	1		1		2							
14–0										1		2	2	1	1	1		2	1	2					
13–6							1	1			1	1		2	2		1			1					
13–0										1	1		4	1	2		1		1						
12–6												4			1		2								
12–0			1								2	2	2	1			1								
11–6	1										2		2	3		1									
11–0										1			1		1	2									
10–6										1	2	1			1		1		1						
10–0											1														
9–6												1			1										

$r = .19 \pm .05$
$N = 204$

TABLE XLIX.—CORRELATION ARRAY: I.Q. OF TRUE CHILD VS. M.A. OF MOTHER

M.A. of Mother	I.Q. of True Child																	
	75–79	80	85	90	95	100	105	110	115	120	125	130	135	140	145	150	155	160
21–6							1						1			1		
21–0																		
20–6									2	1			1					
20–0								1		2				1				
19–6																		
19–0					1	2			4	2			1					
18–6						1			1	1		1	1		1			
18–0							1				1							
17–6						1	1	3	3	2	3		2					
17–0							3		1	1	2	1	1				1	
16–6											1							
16–0							1	2	1									
15–6					1				2	1								
15–0						1	2	1	1									
14–6							1				2			1				
14–0				2	1	1	1	1		2		1						
13–6				1	1	1	1			1						1		
13–0						2												
12–6						1			2									
12–0				1	1	2	2				1							
11–6	2			1		1			1									
11–0																		
10–6																		
10–0																		
9–6																		
9–0							1											

$r = .46 \pm .05$

CHAPTER XI

COMMENTS ON THE CHICAGO AND STANFORD STUDIES OF FOSTER CHILDREN

BARBARA S. BURKS
Stanford University, Palo Alto, California

At the suggestion of the Chairman of the Committee on the 1928 Yearbook, I have prepared a brief statement which considers the outcomes of the Chicago and the Stanford studies of foster children, with respect to points of agreement and points of difference.

The two studies, employing somewhat different methods and types of subjects, were undertaken to gather evidence upon almost identical problems. Both deal with the influence of home environment upon the mental development of children. The Stanford study, through its use of a control group of parents and their true children, offers in addition some evidence upon the influence of heredity; but we are not concerned with that phase of the problem in this discussion.

COMPARATIVE RESULTS

1. Increase in I.Q. Due to Environment

Chicago study. Average I.Q. increased from 91.2 to 93.7 during average foster-home residence of 4 years in 74 cases of the pre-test group. This gain became 7.5 when a correction for age was applied. Children placed in homes above the average of homes in this group gained 5.3 points (10.4 if corrected for age); and those placed in homes below the average gained .1 (5 points if corrected).

A newly committed group of 137 children not yet placed in homes had a mean C.A. of 9-3 and mean I.Q of 88.6. The 260 legitimate *foster* children, with mean C.A. of 12-2, had a mean I.Q. of 94.1. The difference of 5.5 points I.Q. can probably be ascribed to environment.

Stanford study. A group of 214 foster children, whose average inheritance was judged to be close to normal or slightly above, had an average I.Q. of 107. The average environment of their foster homes was markedly superior, and the conclusion was drawn that 5 or 6 points of the excess over 100 I.Q. could be explained by environment.

Comment. The investigations agree in attributing small, but significant increments of I.Q. to superior environment. There are no grounds, in the data as reported in the two studies, upon which to compare the environmental level of the California and Illinois families directly; but it seems reasonable to suppose that the two groups are not widely different in average cultural status. If this is the case, the increments of I.Q. due to environment should be about the same, as, indeed, they are found to be.

Another point to be considered is the possibility that the age corrections made in the Chicago study (5 points for the pre-test group under discussion) may be a little too liberal. These corrections were based upon the drop in I.Q. found with age in the original group of 905 children upon whom the Stanford Binet was standardized. Recent data collected by Merriman[1] suggest that corrections of the magnitude of those used in the Chicago study were somewhat too large. Possibly, therefore, the Chicago and Stanford results with reference to the increase of I.Q. due to environment agree even better than they appear to.

2. Correlation Between I.Q. and Environment

a. Home ratings and foster I.Q.

Chicago study

	r	P.E.
Pre-Test Group		
Home rating and I.Q. (74 cases at time of placement)	.34	.07
Home rating and I.Q. (74 cases after 4 years)	.52	.06
Home rating and I.Q. (51 cases, with negroes and children outside range 5 to 14 eliminated)	.36	.08
Home rating and I.Q. (the above cases after 4 years)	.57	.06
Home Group		
Home rating and I.Q. (304 cases who have been in homes at least 4 years)	.47	.03
Home rating and I.Q. (156 cases placed under 2 years)	.52	.04

[1] *Psych. Monographs,* 1924, 33.

Home rating and I.Q. (59 cases "for which there was practically no information on the parentage")	.51	.06
Home rating and I.Q. (273 cases with negroes and children outside range 5 to 14 eliminated)	.49	.03
Home rating and I.Q. (104 cases placed under 2, with negroes and children outside range 5 to 14 eliminated)	.50	.05

Stanford study

Foster Group

Culture index and I.Q.	.25	.05
Whittier index and I.Q.	.21	.04
Multiple r, environment and I.Q.	.35	.05

b. *Foster parent intelligence and I.Q.*

Chicago study

Home Group

Otis score of foster father and I.Q. (180 cases)	.37	.04
Otis score of foster mother and I.Q. (255 cases)	.28	.04
Mid-foster parent and I.Q. (169 cases)	.39	.04
Mid-foster parent and I.Q. (112 cases with negroes and children outside range 5 to 14 eliminated)	.47	.05
Mid-foster parent and I.Q. (132 cases placed under 2 years)	.39	.05
Mid-foster parent and I.Q. (104 cases of above with negroes and children outside range 5 to 14 eliminated)	.50	.05

Stanford study

Foster Group

M.A. of foster father and I.Q. (178 cases)	.07	.05
M.A. of foster mother and I.Q. (204 cases)	.19	.05
M.A. of mid-foster parent and I.Q. (174 cases)	.20	.05

c. *Home ratings, true parent intelligence, and true child's I.Q.*

Chicago study

Home rating and I.Q. (36 true children in foster-own group)	.47	.09
Mid-parent intelligence and I.Q. (28 true children in foster-own group)	.35	.11

Stanford study

Culture index and I.Q. (101 true children in Control Group)	.44	.05
Whittier index and I.Q. (104 true children in Control Group)	.42	.05
Mid-parent intelligence and I.Q. (100 true children in Control Group)	.52	.05

Comment. At first glance the coefficients of correlation appear to indicate a stronger influence of environment upon foster children in the Chicago group than in the California group (though no discrepancy was apparent in the former treatment dealing with I.Q. *increases* through environment).

Further consideration leads me to believe that the discrepancies can be partly, if not wholly, explained by the existence of a greater degree of selective placement in the Chicago group than in the California group. It was thought at first that the presence (about 8 percent of the total group) of negro children, who, of course, were placed in negro homes, might be sufficient to account in all the groups for selective placement as high as that in the pre-test group (.34 between I.Q. and home rating). But when, at the suggestion of Professor Terman, a number of coefficients were re-computed, with negroes eliminated, the coefficients remained as high as before, including the coefficient of selective placement in the pre-test group. This would indicate that whatever selective placement was present was independent of race, and probably also of nationality, since the Chicago authors state that there were only three south-European children in the entire group.

However, selective placement may still prevail to a significant degree in the Chicago group. This possibility is supported by the following facts:

(1) There is no decrease in correlation of I. Q. with home rating when negro cases are deleted, though we are *sure* that negro children are placed selectively (*i. e.*, in negro homes). It seems probable, therefore, that selective placement is about as high in the white group as in the white and negro groups combined.

(2) Selective placement is as high in the pre-test group with negroes excluded as with negroes included. Effective selection thus occurs independent of race.

(3) The pre-test group, in which the selective placement is known to be as high as .34, shows at the end of several years residence "environmental" correlations no higher than those in the groups for which the degree of selective placement is unknown. This suggests that selective influences may be operative in *all* the groups about as effectively as in the pre-test group.

One thing is *unfavorable* to the possibility supported by the above three points. This is that the correlation is as high between child's I. Q. and home rating in a group of children placed under 2 years, and in a group "for which there was practically no information on the parentage," as it is in the total group. This result makes one speculate as to what the nature of the selective influence could have been.

The hypothesis seems untenable that the methods used in the Stanford study were not as well adapted to revealing relationships and resemblances as those used in the Chicago study, because the correlations between blood relatives, and between home ratings and I.Q. of *true* children are about equal in the two groups.

If the selective placement in all the Chicago comparisons should be about as high as it is in the pre-test group, the outcomes of the two studies could perhaps be reconciled as follows:

The contribution of one variable to the variance of another variable is represented by a proportion equal to the square of the correlation coefficient between the two variables. The square of .34, or .12, of the variance of the pre-test group could be accounted for by reference to home rating when the children were placed. After they had lived in foster homes for several years, the square of .52, or .27, of the variance could be accounted for by reference to home rating. The difference between .27 and .12, or .15, could then be thought of as the change in proportion of the variance that could be accounted for by reference to home rating. This proportion, .15, corresponds to an r of .39, which would probably have been the final correlation between home rating and children's I. Q. if there had been no selective placement to begin with. This value is not very far away from the values found between children's I. Q. and certain combinations of environment in the Stanford study.

One further comment of interest concerns the correlation between the I. Q's of Chicago siblings reared apart. In the limited sibling group, 125 pairs of siblings separated in early childhood yielded a correlation (oldest versus youngest) of $.34 \pm .05$. When negro children and children under 5 or 14 were eliminated, this correlation became, for 63 cases, $.44 \pm .07$. The latter value is not far from correlations ordinarily found for siblings reared in their own homes. The rise from .34 to .44 can probably be attributed to the higher validity of I. Q. technique for age ranges between 5 and 14, and hence the larger opportunity for actual resemblance to be revealed in the measurements.

PREFATORY NOTE TO CHAPTER XII

An opportunity to determine in some measure the effect of nurture upon nature is afforded in the case of children who are transferred from what is undeniably a distinctly unfavorable environment to what is undeniably a distinctly favorable environment. In the following chapter there is reported a study of two groups of children who were thus transferred, not to a distinctive *home* environment, but to exceedingly good and well-managed institutions.

The method consists in giving these pupils the Stanford Revision of the Binet Intelligence Test when they were entering the institutions and then a second time after a year or more of residence there. The amount of change discovered must, of course, be compared with the amount of change expected in re-testing without distinct change in the environment. The final results are clearly negative; that is, no appreciable alteration appears in the intelligence quotient as the result of this period of living in the much improved environment—this despite the fact that improvement in educational achievement was reported by the teachers.

CHAPTER XII

THE EFFECT ON THE INTELLIGENCE QUOTIENT OF CHANGE FROM A POOR TO A GOOD ENVIRONMENT

AGNES L. ROGERS, DOROTHY DURLING, and KATHARINE MCBRIDE
Bryn Mawr College

The purpose of this study was to determine the effect on the intelligence quotient of change from a poor to a good environment. The Stanford Revision of the Binet-Simon Scale was the instrument used.

SELECTION OF TEST AND CONTROL GROUPS

The subjects of the study were sixty-four girls of native American stock, who came originally from extremely poor social and educational conditions and who now are in superior environments. A control group was selected from pupils now in public schools in Bryn Mawr and Wayne, Pennsylvania. The schools from which the control group was drawn are superior in character as regards physical and educational conditions. Data on the control group are omitted in this presentation of the study.

The subjects were examined on first and second occasions by the resident psychologists of the institutions in which they were living.

Pupils from two different institutions are included. We shall refer to these as School A and School B. Accurate detailed information obtained by trained social workers about the conditions of the homes in which pupils in School A had lived before admission to the new environment was available for most of the children. Of every girl in School A, it could be said that she had come from surroundings that would undoubtedly be characterized as 'unfavorable,' when the term is used as it is by Terman.[1] For the meaning of a good or poor environment we accepted the descriptive definitions which are given there as in Table I.

[1] Terman L. M. *Genetic Studies of Genius.* Vol. I, p. 77. (Stanford University Press, 1925.)

TABLE I.—SCHOOL REPORTS ON HOME ENVIRONMENT

A. Probably Favorable Circumstances
  1. Systematic Home Instruction
  2. Well-Educated Parents
  3. Travel
  4. Excellent Companions

B. Probably Unfavorable Circumstances
  1. Excessive Indulgence
  2. One or Both Parents Dead
  3. Parents Divorced
  4. Imperfect Parental Control
  5. Unsuitable Companions
  6. Undue Severity
  7. Child Obliged to Care for Home
  8. Child Has to Look after Self
  9. Child Left to Care of Nurse
  10. Family Below Average Mentally

C. Unclassified
  1. Average Home
  2. Not Enough Information to Say

There were in School A some pupils whose fathers and mothers had been dead for over a year and who had been placed in a number of different homes in succession. These were in some instances the homes of relatives; in others, boarding homes. It was impossible to rate these homes because the data on them were insufficient. Moreover, it seemed unimportant to do so, because the child had spent such a short time in each. It is certain that the circumstances of all those belonging to this group before coming to school A would be granted to be 'unfavorable.' Each child in the group had lost both parents and had been left unprovided for entirely. They were rapidly transferred from home to home in a way that shows clearly that adjustment was a difficult problem.

The other children in School A came directly from their own homes, and there was sufficient evidence in records by trained social workers to rate these homes as regards economic circumstances and the supervision and care given the child. No child came from average or superior economic conditions. Their homes could be classified naturally under three heads. Five girls came from 'fairly comfortable' financial circumstances, eleven from 'poor' circumstances, and seven from 'extremely poor' circumstances. Only seven of the children in the first group described had had fairly competent care. Ten had had poor supervision and six had had extremely poor supervision.

A typical example of 'competent care,' as it is applied here, is illustrated by two pupils in School A who came from the home of their grandfather and step-grandmother. Their attitude to the children was kind. They lived in a good house in a good neighborhood. They had had little education but the cleanliness and comfort of the house indicated a standard of living of a fairly good level.

A characteristic instance of 'poor supervision,' as it is applied here, is that of two sisters in School A who lived with an aunt and grandmother who were unimaginative and exacting, who received the children merely from a sense of duty, and who showed no affection whatever. The home was in a crowded neighborhood and had no yard.

'Very poor supervision,' as we use it here, is illustrated in the case of pupils who had come from the home of their mother's mother who kept a house of ill repute. The father, a laborer, was mentally weak and failed to support his children. The mother was morally weak.

Atmosphere of the Two Schools

School A aims at providing, as far as possible, the kind of social experience a child receives in an excellent family. The girls are placed in cottages, with a house-mother in charge. Eight to thirteen girls of various ages are in each cottage. They live there in "village homes next door to village neighbors." The village is of a superior sort with a Community Playground and Community Library. The health of the children is carefully attended to. The girls are given a large measure of freedom. "Experience with real money, real neighbors, real friends, and real community life is given and not just in preparation for normal life later on, but to provide normal life now."[2]

The formal education at School A is likewise superior to that found in most schools. The best type of modern schooling is given. Educating, rather than training, children is the keynote of the instruction. There is an enriched curriculum. The class work in the physical sciences is particularly good. When pupils show the desire and capacity to study vocal and instrumental music, the opportunity is given. Domestic arts and sciences are likewise available. The spirit of the place, the caliber of the teaching and execu-

[2] Prospectus of the institution.

tive staff, and the social organization of School A justify the statement that it is seldom a greater change of environment could be contrived than the one these girls have had. All of them came from homes of a low economic level and most of them came from surroundings where the supervision and care provided varied from poor to definitely bad.

School B is an institution for delinquent girls. All its pupils come from exceedingly bad home surroundings. This school, like the former, is of a very superior kind and provides admirable physical and educational conditions. It, similarly, is run on the cottage plan, with competent house-mothers in charge. The teachers are good, and the scientific attitude towards mal-adjusted children is indicated by the presence of a trained psychologist and a vocational director on the staff. As in the former instance, a change from a definitely poor to a definitely good environment is unquestionably provided for these girls.

Ages and Intelligence of the Pupils on First Examination

Table II presents the distribution of chronological ages of the pupils at the time of the first examination.

Table II.—Distribution of Chronological Ages at First Examination

C. A.	School A	School B
4–0 to 4–11	1	0
5–0 to 5–11	2	0
6–0 to 6–11	6	0
7–0 to 7–11	6	0
8–0 to 8–11	8	1
9–0 to 9–11	8	1
10–0 to 10–11	6	2
11–0 to 11–11	4	5
12–0 to 12–11	3	7
13–0 to 13–11	..	4
Median C. A.	8 yrs. 9 mos.	12 yrs. 2 mos.
Number	44	20

The mental maturity at the first testing is shown in Table III.

TABLE III.—DISTRIBUTION OF MENTAL AGES AT FIRST TESTING

M. A.	School A	School B
3 to 3–11	..	..
4 to 4–11	3	..
5 to 5–11	6	..
6 to 6–11	4	..
7 to 7–11	7	3
8 to 8–11	8	5
9 to 9–11	11	6
10 to 10–11	3	4
11 to 11–11	2	1
12 to 12–11	..	1
Median M. A.	8–6	9–6
Number	44	20

The standing in intelligence on the first application of the Stanford Revision of the Binet-Simon Scale is shown in Table IV.

TABLE IV.—DISTRIBUTION OF INTELLIGENCE QUOTIENTS ON FIRST EXAMINATION

I. Q.	School A	School B
55– 59	...	1
60– 64	1	...
65– 69	1	3
70– 74	1	3
75– 79	2	4
80– 84	5	4
85– 89	12	2
90– 94	8	2
95– 99	4	1
100–104	5	...
105–109	4	...
110–114	...	...
115–119	1	...
120–124	...	...
125–129	...	...
130–134	...	...
135–139	...	...
140–144	...	...
Median	89.5	77.5
25th Percentile	85	72
75th Percentile	97.5	85
Number	44	20

Testing Procedure

In Table V is indicated the intervals between admission to the institution and the first application of the Stanford Revision of the Binet-Simon Intelligence Examination.

Table V.—Interval Between Admission to New Environment and First Intelligence Examination

Interval in Months	School A	School B
Before entrance		
6 months	1	...
5 "	2	...
4 "	2	...
3 "	6	...
2 "	3	...
1 "	11	...
0 "	5	...
After entrance		
1 month	13	5
2 months	...	6
3 "	...	5
4 "	1	1
Less than 4 months after entrance, but time undetermined	3	...
Median	1 mo. before	2 mos. after
Number	44	20

All the cases included in the investigation were thus examined for the first time not later than four months after admission to the institution. Examination prior to entrance at a longer interval of time obviously could not interfere with the object of the study.

The Retesting

The *interval between the two applications of the Stanford Revision* varied. Table VI shows how the groups studied compared in this respect.

Our subjects were not retested by the same examiners. This condition doubtless has affected the results we have obtained. All the examiners, however, were persons whose ability to test had been established with considerable care. Further, the conditions of testing were excellent, and it can fairly be said that the conditions of

TABLE VI.—INTERVAL BEWEEN FIRST AND SECOND APPLICATION OF THE STANFORD REVISION

Interval in Years	School A	School B
0.5	4	2
1.0	19	7
1.5	7	3
2.0	4	3
2.5	3	5
3.0	1	..
3.5	3	..
4.0	2	..
5.5	1	..
Median	1 yr. 4 mos.	1 yr. 6 mos.
Number	44	20

testing and retesting are as reliable as those that ordinarily prevail in schools.

The coefficients of correlation between the intelligence quotients on the two examinations for the two groups were .78 ± .04 for School A, and .89 ± .05 for School B. Table VII, taken from Freeman[5] summarizes the results so far published.

For the 1154 cases reported by Freeman, the coefficients lie between .84 and .95.

TABLE VII.—MEASURES OF THE VARIATIONS IN THE I.Q. ON RETESTING AS FOUND IN SEVERAL TYPICAL STUDIES

Author	No. Cases	Percentage Differing 10 Points or More	Limits of Middle 50 Percent	Average Change	Coefficient of Correlation Between 2 Tests
Terman[1]	435	.15	−3.3 to +5.7	4.5	.93
Rugg and Colloton[2]	137	.12	−2.3 to +5.6	4.7	.84
Garrison[3]	468	.085	−2 to +4 −3 to +4 −3 to +5	5.4	.88
Rugg, L. S.[4]	114	....	−1.2 to +1.9	3.1	.95

[1] *The Intelligence of School Children*, Ch. ix.
[2] "Constancy of the Stanford-Binet I.Q. as shown by retest," *Jour. Educ. Psych.* 12:1921, 315–322.
[3] "Additional retests by means of the Stanford Revision of the Binet-Simon Tests," *Jour. Educ. Psych.*, 13:1922, 307-312.
[4] "Retests and the constancy of the I.Q.," *Jour. Educ. Psych.*, 16:1925, 341-343.

[5] Freeman, F. N. *Mental Tests*, p. 345.

Our results upon retests are grouped in Table VIII so that the following facts stand out. There is a difference of more than ten points between the first and second tests in 18.1 percent of School A, and in 20.0 percent of School B. The combined group when re-examined had 81 percent within ten points of their first I.Q. The corresponding figure for the 1154 cases reported by Freeman was 85 to 91 percent. The middle fifty percent of change ranged from 3 points increase to 6 points decrease in our combined group. The corresponding figures obtained by Terman[4] range from 5.7 increase to 3.3 points decrease.

TABLE VIII.—DISTRIBUTION OF CHANGES IN I.Q. ON RETESTING

Amount of I.Q. Change	School A	School B
+16 to +20	..	1
+11 to +15	3	1
+ 6 to +10	6	1
+ 1 to + 5	5	4
0	5	1
− 5 to − 1	13	6
−10 to − 6	7	4
−15 to −11	2	2
−20 to −16	2	..
−25 to −21	1	..

Both Schools	
25th Percentile	+3
Median	−1
75th Percentile	−6
Q	4.5

CONCLUSIONS

These results show no appreciable effect of environmental change upon the intelligence quotient. Such differences as are found may well be due to accidental factors. Teachers have reported improvement in the pupils in school achievements. These changes in the pupils, however, are not reflected in the intelligence quotient.

The validity of our results depends on the reliability of the examiners. While re-examination by the same psychological examiner

[4] Terman, L. M. *The Intelligence of School Children*, p. 142.

would have increased the value of the study, we are warranted in saying that, under superior testing conditions, the intelligence quotient of these two groups of 64 girls shows no change beyond what could occur by chance, in spite of an extreme modification of environment.

PREFATORY NOTE TO CHAPTER XIII

The significance of this study lies in its use of a measuring instrument (the Rational Learning Test) which apparently depends very little upon facts or skills accumulated through schooling, or through the social *milieu*.

Differences between white and negro children upon this test are found to be comparable to differences repeatedly found in previous studies which have used tests of the highly verbal type. This result carries the implication (though it does not prove) that the differences between white and negro subjects which have appeared so persistently wherever tests have been given are due to a more elemental difference than can be produced by mere differences in educational opportunities. However, the evidence is necessarily indirect, since the possibility cannot be denied, at least from any data reported in this chapter, that the capacity to attack and solve a unique learning problem (as in the Rational Learning Test) may be itself influenced by environment.

CHAPTER XIII

COMPARISON OF WHITE AND NEGRO CHILDREN IN THE RATIONAL LEARNING TEST[1]

JOSEPH PETERSON
George Peabody College for Teachers,
Nashville, Tennessee

In mental testing it is desirable to have complex learning situations to present to subjects, situations which require for their solution as little as possible of such information and skills as may be derived from different cultural communities and yet which will tax the thinking activities of the subject to the extreme without discouragement of effort. Tests of information are not suitable for comparisons of racial and other groups which have different cultural environment or educational opportunities, and tests of any kind involving a considerable amount of verbal skill are likewise objectionable.

A NEW LEARNING TEST

It is the purpose of this article to give a very brief description of a learning test first worked out in 1916, which presents a problem that is different from that which can be solved by the usual associative learning, one that can be solved only by a series of choices—pure guesses at first—the range of which may be greatly limited through a rational organization of the situation. Each subject completes the learning at a single setting, which varies in length inversely with the subject's ability. The test has been standardized for adults,[2] and in a simplified form has been used by the

[1] The writer is indebted to his former students, Miss Pantha V. Harrelson and Dr. Lyle H. Lanier, the latter of whom served as his National Research Council research assistant, for aid in securing the data here reported. Dr. Lanier personally gave most of the individual tests in the later experiments. The earlier ones were given by Miss Harrelson. He is also indebted to Dr. K. C. Garrison for certain data on validity and reliability, and to Miss Gladys Dunkle and Mr. Howard Easley for giving some of the more recent tests.

[2] Peterson, Jos. "Experiments in rational learning." *Psych. Rev.*, 25: 1918, 443-467; also "Tentative norms in the Rational Learning Test." *Jour. Applied Psych.*, 4:1920, 250-257.

writer in race studies.[3] The method and results of these studies are given here very briefly. They will be reported more fully elsewhere.[4]

Rather extensive experience in testing white and negro children has demonstrated to us the shortcomings in several respects of group tests. Such tests presuppose that each person is interested in doing as well as he can; that because of individual competition he will be spurred to maximal effort. They stress speed very much. Persons who have come up through our modern schools know what it feels like to work in tests at approximately maximal speed. To the negro, speed has little significance; it may even have unfavorable associations, as it has for union men. Such effects of past environment, in different ethnic groups and in different classes of people in the same country, as degrees of readiness to work at one's utmost speed or to react with the greatest possible discrimination and speed throughout an exercise or test, must not be overlooked in mental testing; and they merit more study than they have received. Recognizing the weakness of group tests, even of the non-linguistic type, in race comparisons, we have sought to supplement our use of such tests with individual learning exercises and tests of such nature as to promote constant stimulation of each child in a natural way by the tester. The Rational Learning Test has the merit of stimulating the subject constantly throughout the test, but in a manner that follows a definite plan alike for all subjects. Moreover, while it gives opportunity for the play of the higher mental functions in the organization of a complex situation toward a practical and definite solution, it keeps well within the range of the experience of all subjects, however meager this experience may have been, so far as its demands on knowledge and skills are concerned. It has the additional advantage of leaving a definite record of all the relevant responses of the subject in the order of their occurrence. The subject does not have to write, and he is left

---

[3] Peterson, Jos. "Comparative abilities of white and negro children." *Comparative Psychol. Monog.*, 1:1923, No. 5; also Peterson, Jos., Lanier, L. H., and Walker, H. M. "Comparisons of white and negro children in certain ingenuity and speed tests." *Jour. Comp. Psychol.*, 5:1925, 271-283.

[4] These studies constitute part of a research program carried out under the auspices of the National Research Council Committee on Scientific Problems of Human Migration.

free to take his natural time for response. At the same time, his average speed can easily be determined later from the test records.

1. Nature of the Test

In the more recent form, with a fore-exercise, the subject, after responding to some questions as to age, name, etc., is instructed as follows:

> The letters A, B, and C are numbered 1, 2, and 3, but A may not be 1, B may not be 2, and C may not be 3. They are numbered just by chance. Now, I am going to call out A and you must guess until you find the number that goes with A. I shall tell you when you get it right, and shall then call out B. When you guess the number for B, I shall call out C. All you have to do is to remember the number that goes with each letter so that you can give all the numbers without making a mistake. Remember that each letter has a different number.

The time is kept from the calling out of the first letter until the subject has given the last number of two correct series of responses, when the fore-exercise is completed; and every number given by the subject is recorded in order. During the first trial the tester says "Right" whenever the right number is given. Here is a sample record:

	A	B	C	
Numbers assigned	1	3	2	
Responses in 1st trial	3	2	(4)	Tester says: "No; there are only three letters."
	2	3	2	
	1			
Responses in 2nd trial	1	2	2	
		3		
Responses in 3d trial	1	3	2	
Responses in 4th trial	1	3	2	

Summary: Time, 43 seconds; errors, 5; repetitions, 4.

The fore-exercise serves to acquaint the subject with the nature of the problem. When it is completed, the tester says to the subject: "Let us now take all the letters from A to E in the same way. The numbers will then go from 1 to 5. We shall do just as we did before." The test is then carried out on the same principle as illustrated in the fore-exercise, yielding a complete, objective record of all the subject's responses. To avoid influences of coaching, we have used in a consecutive circular order with successive subjects these different forms, or systems of numbering.

	A	B	C	D	E
First	4	1	5	2	3
Second	2	1	4	5	3
Third	5	3	1	2	4

In the first use of this test in our race studies we had no fore-exercise, but made the instructions correspondingly more full. Later experience shows that the understanding of what is to be done and of the nature of the test can best be effected by means of the fore-exercise.

2. The Reliability of the Test

While we do not have reliability data for children, those at hand derived from separate applications of the test to college students (when eight-letter forms were used) indicate that the reliability is rather high for a learning test. With 49 psychology students who took the test twice (different forms), the correlation coefficients between similar elements in the test were: time, .75; repetitions, .60; errors, .70. These factors, put in the form of percentiles and equally weighted, give a reliability coefficient for the whole test (correlation between first and second percentile scores) of .67 ± .05. Heron[5] found reliability coefficients of .58 for time, of .48 for repetition, and of .57 for errors. He used one ten- and one fifteen-letter form on 60 college students (without fore-exercise), thus greatly increasing the difficulty of the second test. Correlating half-test data by the odd-even trials method on 100 college students (mostly sophomores) and finding reliability by means of Brown's formula, we obtained a coefficient, for error scores, of .92. The inter-correlations of the different factors—time, repetitions, and errors—are usually .60 or above.

3. The Validity of the Test

The percentile ranks in Rational Learning correlate nearly as highly with psychology scores (the average in eight or ten competitive examinations) as do scores in Alpha or Otis, running somewhere between .30 and .55, as a rule, with different degrees of difficulty (or numbers of letters). With 107 twelve-year-old white children,

[5] Heron, W. T. "Individual differences in ability versus chance in the learning of the stylus maze." *Comp. Psych. Monog.*, 2: 1924, No. 8, p. 53.

selected irrespective of school classification, we obtained a correlation of .37 between the best weighting of time, repetition, and error scores in rational learning and a group form of Stanford Binet tests; and between the Myers Mental Measure and the Rational Learning scores, a correlation of .53. These correlations were all obtained from rather homogeneous groups. Miss Margaret Shickel found a correlation of .63 between scores on an eight-letter Rational Learning Test and the teacher's estimate of the intelligence of 33 sixth-grade children in Peabody Demonstration School. The teacher's estimate was secured by the paired comparisons method, requiring that each student be compared severally with every other one. Applied to a group of 10 fifth-grade, 59 sixth-grade, 28 seventh-grade, and 9 eighth-grade children, all twelve-year-olds selected irrespective of grade classification, the Rational Learning Test gave a marked and consistent difference between the mean scores of each grade and that of the one next above it. Dividing these differences by their probable errors, we get the following quotients: $\frac{\text{Diff.}}{\text{P.E. diff.}} = 6.13$ for fifth and sixth grades, 1.45 for sixth and seventh grades, and 3.98 for seventh and eighth grades. If the numbers had been larger, there is little doubt that all the differences would have been statistically reliable.

These data all indicate a fair validity for Rational Learning as an intelligence test, so far as group comparisons are concerned. The test has also the advantage of not having been constructed to correlate highly with standard intelligence tests, which measure largely the same functions as are required for successful performance in conventional school work. It undoubtedly puts stress on valuable rational organization processes in learning, neglected in most tests, and is little affected by the cultural influences which count considerably in the conventional tests.

Results of Racial Comparisons

Since we have three kinds of data—time for completing the test (minutes), number of repetitions necessary to produce two repetitions without error, and total number of errors—the group comparisons may be made in several ways. Since minutes, repetitions, and errors correlate highly, it is obvious that a difference in each of these factors may mean an even larger difference if we can somehow weight all together; because a given difference in errors with time factors equal, for instance, is less than an equal difference if the poorer group also requires more time. That, however, cannot be considered in the short space available here. A good way to compare our groups is to show curves of errors in successive repetitions. Differences at the concluding end of such curves may be relatively very great, even though rough total-error comparisons show little difference. This is because absolutely the greatest number of errors come in the beginning of the learning where they are due to the necessity of guessing largely at first. Figure I brings out clearly this fact, and shows that, while the race groups of equal ages at first differ but little (and unreliably), they differ markedly and very significantly at the end of the eighteenth trial, at which point all the groups of white children had completed the learning. Note that these curves are of median scores, too, not of averages, which would be unduly influenced by extreme cases. These curves

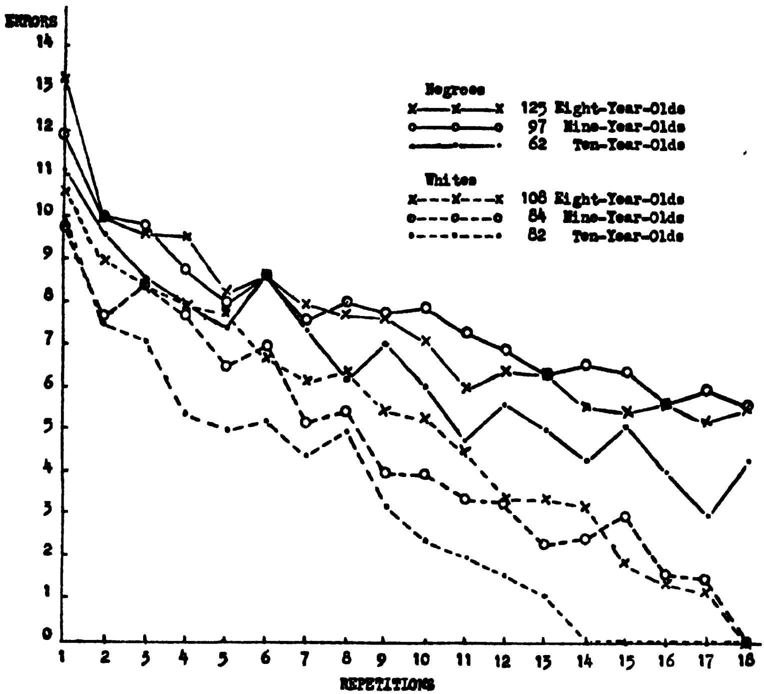


FIG. I.—LEARNING CURVES SHOWING THE RATE OF ELIMINATION OF ERRORS BY EACH OF THE AGE-GROUPS OF WHITE AND OF NEGRO CHILDREN are based on large numbers of children selected in representative Nashville Schools, irrespective of grades.[6] In this experiment time-differences (not represented in the curves) were greater than either error or repetition differences, averaging for the three age-groups 2.37 quartile deviation of the whites, as compared with 1.23 and 1.47 quartile units for repetitions and errors, respectively. The differences are all statistically reliable except those for the ten-year-olds in gross errors and in repetitions. The curves, however, show that these two cases of unreliability are due to the method of comparison of gross errors and that they are really marked and of a very significant nature when shown analytically, as in the curves.

These tests were given in the year 1920-21 to all the children aged 8, 9, and 10 years in three white and three negro schools in Nashville, and, in addition, to most of the children of these ages in a fourth white school. The schools were selected as fairly representative of the two races, with the exception that two of the negro schools were undoubtedly the best in the city. These children were all tested without fore-exercises, but any child who failed at first to get the idea clearly was given a three-letter test to insure that he did do so.

Three years later we applied the same test (this time giving the fore-exercise to all the children) to 69 white and 46 negro children, all twelve-year-olds, selected irrespective of grades. It was impossible, however, to get all the twelve-year-olds because of illness among them at the time; and in addition to this, 27 seventh-grade whites were left out because of wrong information given by one principal. All the white children in one school, with this exception, were tested, and all the negroes but 6 in the fourth grade, 1 in the fifth, and 8 in the sixth. The selection therefore definitely favored the negroes, since over a fourth of the whites in the upper grade (therefore of the brightest ones) were omitted. There were only 4 eighth-grade white, and 1 eighth-grade negro in the tested groups. In these groups the difference between the medians of the whites and of the negroes were, for time, 20.0 times the probable error of the difference; for repetitions, .61 times its probable error; and for errors, 3.0 times its probable error. The test with only five letters was evidently too easy for twelve-year-olds, as the median time for

[6] For a more complete statement and data see *Comp. Psych. Monog., 1,* No. 5, 93 ff.

the whites was only 3.96 minutes and for negroes 5.6 minutes. Only the time-difference is statistically reliable, but all differences favor the whites, in spite of the omission of the seventh-grade students mentioned above and of the fact that the tests were too easy. Stated in a common unit, for ease of comparison with measurements by other tests, these differences between the medians of the two races are as follows: for time to complete the test, 1.19 quartile units of the white distribution; for repetitions, .18 unit; and for errors, .92 unit. All differences favor the whites. As we have already pointed out, gross errors are a poor means of comparison, and error curves, as in the above case, would show the differences in a much more significant way. Data for such a comparison are not at hand.

More recently we have given the test in a more difficult form (seven letters), with the fore-exercises, to 119 Nashville white children, 17 Chicago whites, 87 Nashville and 40 Chicago negroes, all twelve-year-olds, securing as representative samples as possible. In this form of the test the median time for whites was 12.87 and 10.50 minutes for the Nashville and the Chicago groups, respectively; for the negroes the corresponding medians were 19.11 and 13.25 minutes. The number of Chicago whites is, of course, too small to be worth anything, but uncontrollable circumstances prevented our increasing the number as we had planned. Our data give quotients of differences divided by their probable errors as follows (the first quotient stands for the Nashville race-groups and the second for the Chicago groups): for time, 5.71 and 1.67; for repetitions, − .46 and 2.29; for errors, 2.06 and 2.79. The positive numbers favor the whites and negative number the negroes. Again, with but one exception, the differences all favor the whites, but only one difference is statistically reliable. The numbers, however, are very small in some cases. The almost uniform superiority of the whites, of course, helps establish the reliability of the race difference; for, while in one case the probability of superiority of one group over another by pure chance is 1 in 2, in two independent cases it is 1 in 4, and in three, only 1 in 8. In quartile units of the whites these differences are for the Nashville race groups 1.25 for time to complete the test, −.10 for repetitions, and .44 for number of errors made. That is, they excelled in terms of time and error criteria, but were slightly behind the negroes when judged by the repetitions criterion. The corresponding numbers for the Chicago groups are .62, .81, and 1.00 quartile units, all these differences favoring the whites.

This study is not complete yet, as we are extending it to tests of larger numbers of white and negro children in New York City; so it is impossible here to make comparisons of learning curves, as we have done with the first tests described in this paper.

It is interesting to note that both white and negro children in Chicago were superior to those of the same race in Nashville, but the Chicago groups, especially the white children, are too small to warrant anything but the most tentative conclusion. Moreover, the Chicago children were tested in the summer and may not be fair samplings.

It is important to mention here, in conclusion, that even such non-linguistic tests as the Myers Mental Measure and the International Group (Rotator) Tests (developed by Dr. Brigham and others at Princeton) show race differences on the same groups of Nashville children which are very much greater than those obtained with the Rational Learning Test, thus confirming in a measure what was said on an earlier page of this paper about the inadequacies of group tests as measures of race differences. These differences, stated in quartile units of the white distributions in each comparison, are 2.34 for the total score on the Myers Mental Measure, and 1.23 on the International Tests. The latter number is the average of the quartile unit differences obtained on the several tests used, since the scores in the five tests used—the first, fourth, fifth, sixth, and ninth—were not simply added. On a group form of the Stanford Binet which we devised by increasing the number of items in some of the tests suitable for group use (as will be explained in our forthcoming complete report) the difference between the Nashville whites and negroes was 2.83 quartile units of the whites' distribution. The difference divided by the probable error of the difference, in the case of these three tests yielded quotients of 13.60, 5.74, and 13.81, respectively, indicating a high degree of reliability, so far as the mere statistical calculations are concerned.

Even in the Rational Learning Test, speed differences between these two races are always marked, despite the fact that each child takes the speed natural to him. Is this an innate or a cultural difference? We don't know. Fortunately, we can make comparisons with this test which leave out of consideration entirely this questionable speed element.

PREFATORY NOTE TO CHAPTER XIV

It was found in this study of 149 seventh-grade Jewish girls that those in whose homes a large amount of English was spoken were not more likely to get high scores on a verbal than on a non-verbal intelligence test. The data also showed that the number of years the mothers of the pupils had lived in the United States correlated close to .28 with "foreign-language" measures, but zero with intelligence scores of the children.

From these facts are drawn the conclusions that verbal tests are valid measures of intelligence for such subjects as these (who had all lived in the United States for at least three years), and that inherited ability is reflected to some degree by the language used in the home.

These conclusions may be justified, for it would seem likely that any *large* effects of language handicap would have appeared if they had existed. However, there is one serious limitation in the data. This limitation results from using children of a single school grade as subjects at the same time that a test is employed (the International) on which there are as yet no age-norms. The authors have treated their results by using (a) raw test scores for all subjects; (b) raw scores for pupils of a single age range as well as grade range; and (c) raw scores for all subjects, with age partialled out of the correlations. All of these methods, of course, have the effect of under-valuating any actual relationships that exist; and, since the probable errors of the correlations reported are in the vicinity of .05, slight effects of language handicap might fail to come to light.

CHAPTER XIV

A STUDY OF THE RELATION BETWEEN INTELLIGENCE AND THE ACQUISITION OF ENGLISH[1]

Katharine Murdoch and Doris Maddow
Vocational Service for Juniors, New York City
and
Nettie Lesser Berg
Public School 91, New York City

The main object of this investigation was to determine what relation, if any, exists between innate intelligence and the amount of English spoken in the homes of a group of school children of foreign-born parentage. It was hoped to throw some light upon the question whether environmental factors solely or these plus inherited mental ability are reflected in the extent to which the family of an immigrant speaks the language of his adopted country.

The investigators were also incidentally concerned to learn how standard group intelligence tests, used in the schools, compare with a non-language test in fairly estimating the native intelligence of foreign-born children.

The fundamental difficulties encountered in contemplating such a study were: first, to secure a valid measure of mental ability, and second, to estimate accurately the amount of English spoken in the home. It was felt unwise to use the standard non-language tests because it is questionable whether they really measure abstract intelligence and because at best they are not sufficiently difficult to measure superior intelligence. On the other hand, the use of verbal tests for our primary problem was open to the objection, often made, that such test scores are lowered by a linguistic handicap and that probably even bilingualism affects test scores. It was important, therefore, to secure a test which, while possessing the recog-

[1] This investigation was conducted under the joint auspices of the Vocational Service for Juniors and Public School 91, Manhattan. The authors are indebted to Mary H. S. Hayes, the director of the Service, and Emma Sylvester, the principal of the school, for advice and coöperation. Acknowledgment is also made to Mr. Stuart C. Dodd, who supplied the International Test materials.

nized intelligence-measurement factors of the verbal tests, eliminated the language handicap entirely by permitting a non-linguistic response. The International Test, prepared by Stuart Dodd under the direction of Dr. Carl Brigham, seemed to meet these qualifications. Further to rule out language difficulties the test instructions were given in both English and the foreign language spoken in the home.

As no satisfactory measure of English spoken in the home then existed, the investigators constructed one for this purpose. A questionnaire[2] was drawn up and submitted to the children, supplemented by individual interviews in doubtful cases. The results were then summarized in two ways: first, by securing the proportion between the number of individuals in the family speaking mostly English and the total number in the home, and second, by classification on a language scale[3] evolved by direct reference to the information contained in the questionnaire.

The subjects were 149 school children, all girls, all Jewish (with Jewish as the foreign language spoken in the home) and all in Grade 7A. A more ideal group for our purpose would have been of one age unselected as to grade, but administratively this ideal group was not readily accessible. Of the 149 children studied, 112 were born in the United States. Of the 37 born in foreign countries, all had been here at least three years. In every case but one, both parents were foreign born.

Our data, in addition to the language measures mentioned, consist of scores on three tests—the Otis Intelligence Scale (Advanced Examination), Thorndike Word Knowledge, and the International—and information as to length of residence in the United States of father, mother, and child.

The Otis Test was used as typical of standard group-intelligence tests, and the Word Knowledge Test as a measure of a purely verbal function.

The International Test, as a new measure of intelligence and as the pivotal test in our study, requires some description. Its feature lies in the fact that it obviates the use of any specific language, yet retains the chief elements of the standard verbal tests. The material consists of picture, instead of word, problems, but, like the verbal

[2] See Supplement A.
[3] See Supplement B.

tests and unlike the standard non-verbal tests, it deals to a great extent with abstract situations. Ability to generalize, to analyze, and to understand analogies, for instance, is apparently required to the same degree as in verbal tests, but the thought processes may go on in any language. The answer to each problem is the marking of a picture or series of pictures, not of words. The test is of graded difficulty; the most difficult sections test superior adult intelligence as other non-verbal tests probably do not.

The reliability of the test has been proved by Mr. Dodd, the author, who found a split-half reliability coefficient of .96 in a college and in a seventh-grade group. He also reports the mean of the validity coefficients in these and other groups, including feeble-minded and Hindu children, to be .80.

In giving the test we made certain that no child misunderstood the directions because of a linguistic handicap. All instructions were given in English and then literally translated into Yiddish. The children were tested on three successive days. On the first day they worked in practice books. The results were then carefully examined to see what mistakes had been made by reason of a misunderstanding of directions. On the second and third days the previous directions were supplemented to insure greater clarity, translated into Yiddish, and further practice periods were given to insure complete understanding. Performance on the second and third days in the simpler sections of the test indicated that in practically every case the directions had been understood and that therefore intelligence was being tested without the complication of any linguistic factor.

The Otis and word knowledge tests were, of course, given in English only, in accordance with the authors' directions.

As the first step in constructing the needed language scale, a language questionnaire was drawn up containing a series of questions relating to the language spoken in the home by the members of the family of the child. When the answers were returned by the children, discrepancies were found indicating a need for fuller information. Pertinent questions were then added and the questionnaires returned. In this way the final set of questions was so constructed that it elicited clear replies. From the number speaking English most of the time as compared with the total number in each family a numerical ratio for English spoken in the home was obtained.

A careful examination of the replies revealed that they grouped themselves into several definite classes. A language scale was drawn up ranging from 'all Jewish spoken in the home' to 'all English spoken.' Records which did not fit into the scale were set aside for further classification. As a result of tentative classification and re-examination as described, the language scale in its final form was evolved (Supplement B).

The imperfections of such a scale as a scientific measurement of language spoken are obvious. It is probable that the steps on the scale are not equal, but we have made that assumption for purposes of correlation, in the absence of any immediately available basis of measurement that was more accurate.

It occurs to us that some combination of the two methods of measuring language might be attempted by future investigators with a view to obtaining a more nearly perfect measuring scale.

We made use also of the method commonly employed in language classifications in previous studies—groupings on the basis of 'all English spoken,' 'all Jewish spoken' and 'both languages spoken,' by father, mother, subject, and older siblings. The data, however, are not published here because several of the resulting groups were too small to allow of reliable interpretation.

In our study we used the number of years the mother had been in the United States, instead of similar information for the father, because with foreign-born children the former figure corresponded in every case with the child's own residence in this country, whereas the father often preceded the mother and children to this country by several years.

Intercorrelations were secured for the following elements: chronological age, scores in Otis, International, and Word Knowledge tests, proportion of English spoken in the home, classifications on the language scale, and years of mother in the United States.

We wanted primarily to compare relative intelligence with the amount of English spoken in the home. Several methods of using the data suggested themselves, but the exclusive use of any single one seemed inadvisable for our purposes. The use of data for the entire group was complicated in that our subjects were not of one age group and our measures of intelligence in the various tests were of necessity scores, rather than some computed measure of brightness—age norms are not yet available for the International Test. In this use of raw scores no allowance is made for the fact that older pupils, though of the same intelligence, would be expected to make higher scores than younger ones.

Another possible method was the division of subjects into age groups and the treatment of each age group separately. This procedure had the disadvantage that the aggregates were small and the results therefore not as reliable as those obtained by a study of the entire group.

Finally, the use of the partial correlation method with the entire group to eliminate the effect of age from the essential coefficients was suggested, but we did not feel that its use to the exclusion of the other methods would serve our purpose. The partialling-out method, it seemed to us, did not compensate for the inadequacy of our data in not representing one unselected age group. In our group, age was a very important factor in revealing intelligence because school grade selection is such that the older children on the average make lower scores in the tests than the younger ones—directly contrary to the common assumption when the method of partialling out age is employed. The result, therefore, would be to reduce the correlations significant in comparing intelligence with amount of language spoken in the home, instead of increasing them as would be the case if we had unselected age groups.

Our final decision was to use all three methods; namely, separate studies of the three largest age groups (11, 12, and 13 years), raw correlations for the whole group, and correlations for the whole group with age partialled out. Tables I to V indicate the correlations obtained.

Within the age groups the coefficient of correlation between proportion of English and the International tests is for the 11-year-olds .38, between the language scale and the same test scores .32; for the 12-year-old group the similar coefficients of correlation are .19 and .35. The same correlations for the 13-year-old group are slightly negative. This is the smallest of the age groups considered separately, and an inspection of our data shows that five extreme cases altered the direction of the general trend. Were these omitted, correlations similar to those noted for the 11- and 12-year-old groups would have resulted.

The total group, which is probably the most reliable measure (except for the reduction in correlation due to unselected age groups, as already mentioned), shows a coefficient of correlation of .15 between language scale and the International Scale. This is

TABLE I.—INTERCORRELATIONS FOR 37 11-YEAR-OLD JEWISH GIRLS
(Figures in Parentheses are the Probable Errors of the Coefficients Which They Follow)

	Chron. Age	Otis	Inter-national	Thorndike Word Kn.	Prop. of English	Lang. Scale	Yrs. Mother in U. S.	M	σ
Chron. Age	——	−.12 (.11)	−.20 (.11)	.02 (.11)	−.01 (.11)	−.165 (.11)	.15 (.11)	11.66 yr.	2.57 mo.
Otis		——	.63 (.07)	.67 (.06)	.26 (.10)	.26 (.10)	.01 (.11)	93.96	23.37
International			——	.43 (.09)	.38 (.09)	.32 (.10)	.05 (.11)	173.65	28.49
Thorndike Word Kn				——	.14 (.11)	.07 (.11)	.13 (.11)	33.30	10.44
Prop. of English					——	.69 (.11)	.09 (.11)	66.69	22.53
Lang. Scale						——	.13 (.11)	3.84	1.97
Yrs. Mother in U. S.							——	19.84	5.76

TABLE II.—INTERCORRELATIONS FOR 53 12-YEAR-OLD JEWISH GIRLS

	Chron. Age	Otis	Inter-national	Thorndike Word Kn.	Prop. of English	Lang. Scale	Yrs. Mother in U. A.	Mean	σ
Chron. Age	——	−.21 (.09)	−.03 (.09)	−.16 (.09)	−.12 (.09)	.09 (.09)	−.07 (.09)	12.48 yr.	3.18 mo.
Otis		——	.525 (.07)	.56 (.06)	.24 (.09)	.38 (.08)	−.04 (.09)	85.02	19.56
International			——	.30 (.08)	.19 (.09)	.35 (.08)	−.17 (.09)	156.13	31.84
Thorndike Word Kn				——	.17 (.09)	.18 (.09)	−.02 (.09)	44.15	10.85
Prop. of English					——	.20 (.09)	.25 (.09)	65.42	14.29
Lang. Scale						——	−.09 (.09)	3.80	1.57
Yrs. Mother is U.S.							——	17.43	5.91

TABLE III.—INTERCORRELATIONS FOR 29 13-YEAR-OLD JEWISH GIRLS

	Chron. Age	Otis	International	Thorndike Word Kn.	Prop. of English	Lang. Scale	Yrs. Mother in U. S.	Mean	σ
Chron. Age	—	.17 (.12)	.41 (.10)	−.06 (.12)	−.11 (.12)	−.11 (.12)	−.18 (.12)	13.22 yrs.	3.47 mo.
Otis		—	.66 (.07)	.68 (.07)	−.18 (.12)	−.25 (.12)	−.03 (.13)	85.22	22.09
International			—	.50 (.09)	−.12 (.12)	−.04 (.13)	−.07 (.12)	157.76	32.79
Thorndike Word Kn.				—	.06 (.12)	−.06 (.12)	−.02 (.13)	43.45	13.48
Prop. of English					—	.72 (.06)	.25 (.12)	65.08	23.77
Lang. Scale						—	.35 (.11)	3.66	2.03
Yrs. Mother in U. S.							—	15.79	8.81

TABLE IV.—INTERCORRELATIONS FOR ENTIRE GROUP OF 149 JEWISH GIRLS, AGED 10 TO 16

	Chron. Age	Otis	International	Thorndike Word Kn.	Prop. of English	Lang. Scale	Yrs. Mother in U. S.	Mean	σ
Chron. Age	—	−.16 (.05)	−.14 (.05)	−.08 (.05)	−.16 (.05)	−.20 (.05)	−.41 (.05)	12.88 yr.	14.79 mo.
Otis		—	.62 (.03)	.62 (.03)	.13 (.05)	.17 (.05)	.05 (.06)	86.75	22.13
International			—	.39 (.05)	.09 (.05)	.15 (.05)	−.06 (.06)	160.10	33.08
Thorndike Word Kn.				—	.10 (.05)	.10 (.05)	.015 (.06)	43.70	11.58
Prop. of English					—	.58 (.04)	.29 (.05)	63.67	20.23
Lang. Scale						—	.27 (.05)	3.59	1.79
Yrs. Mother in U. S.							—	16.77	7.46

TABLE V.—CORRELATIONS AFTER CHRONOLOGICAL AGE IS PARTIALLED OUT

	Otis	International	Thorndike Word Kn.	Prop. of English	Lang. Scale
Otis	——	.61	.61	.11	.14
International		——	.38	.07	.13
Thorndike Word Kn.			——	.09	.08
Prop. of English				——	.57
Lang. Scale					——

three times the probable error of the coefficient. The coefficient of correlation between the proportion of English spoken (the other language measure used) and the International Scale is .09 with a P.E. of .05.

It will be seen from the tables that, on the whole, a positive correlation exists between the International Test scores and the two language measures—that is, in general, higher intelligence, as measured by the test, was connected with a larger amount of English spoken in the home. It must be remembered also that the real correlation between language spoken in the home and the International scores is probably lowered by the low reliability of the two methods of measuring language. We have no actual measure of the reliability of our language scales, but their low correlation with one another (.58 for the entire group) would indicate that their reliability is low.

The factor of number of years of residence in the United States can be eliminated, we feel, from a consideration of the correlation between intelligence and amount of English spoken, because the correlation of years of residence of mother in the United States and the Otis Test, as well as the same data compared with the Word Knowledge Test, is practically zero.

Another point of interest to the investigators was the question of the relative size of the correlation between the two language measures and the International Test (in which the language factor is omitted) and the same language measures and the Otis and Word Knowledge scores (in which tests the use of English is demanded).

In Table IV the coefficient of correlation between proportion of English and the Otis Test score is .13, of language scale and Otis Test score is .17, as compared with .09 and .15, the corresponding coefficients when International, instead of Otis, scores are considered. The differences, it will be seen, are very small—in each case smaller than the probable error of the coefficients. A similar result is found when comparison is made between the International Test scores and the Word Knowledge scores in their relation to the language measures. The coefficients here are .10 between proportion of English and Word Knowledge test, and .10 between language scale and Word Knowledge, as compared with .09 and .15 when International scores are considered similarly. Again the differences shown are very small. The correlations are also undoubtedly lower than they would be in an unselected age group and with a more perfect language measuring instrument. This, however, does not invalidate the actual relationship shown between language measurements and the International on the one hand and the language measurements and the Otis and Word Knowledge tests on the other hand.

The interpretation of the data in the individual age groups is difficult because the total number studied is small, but here again the significant results are similar to those shown for the entire group.

Conclusions

To summarize, all the evidence tends to establish that a large amount of English spoken in the home is not any more connected with high scores on a verbal intelligence test like the Otis, or even on a straight English word test, than it is with high scores on an intelligence test like the International which obviates all necessity of understanding English. In other words, in the group studied, the Otis Test, representing the standard verbal intelligence test, would seem to be as fair a measure of mental ability as a similar test which does not involve knowledge of English.

We feel warranted in drawing the following conclusions from our investigations:

1. In the case of Jewish families living in New York City who have been in the United States at least three years, inherited mental

ability is to some degree at least reflected in the extent to which a family speaks English in the home.

2. Standard verbal tests when applied to Jewish children of foreign-born parentage who have reached Grade VIIA in school are valid measures of intelligence.

SUPPLEMENT A

Language Questionnaire

1. What language does your father speak when he talks to your mother?
2. What language does your father speak when he talks to your older brothers and sisters?
3. What language does your father speak when he talks to you?
4. What language does your father speak when he talks to your younger brothers and sisters?
5. What language does your mother speak when she talks to your father?
6. What language does your mother speak when she talks to your older brothers and sisters?
7. What language does your mother speak when she talks to you?
8. What language does your mother speak when she talks to your younger brothers and sisters?
9. What language do your older brothers and sisters speak to your mother?
10. What language do your older brothers and sisters speak to your father?
11. What language do your older brothers and sisters speak to you?
12. What language do you speak when you talk to your father?
13. What language do you speak when you talk to your mother?
14. What language do you speak when you talk to your older brothers and sisters?
15. What language do you speak when you talk to your younger brothers and sisters?
16. How many people in your home usually speak English?
17. How many people in your home usually speak some other language?
18. List by name and relationship the people living in your home: give their ages and the language they speak at home most of the time.

SUPPLEMENT B

Language Scale

1. All Jewish.
2. All members of family speak mostly Jewish (some speak some English).
3. Both parents speak Jewish to one another and to the children. Children speak Jewish to both parents and English to one another.

4. Parents speak Jewish to one another. One parent speaks English and one Jewish to the children. Children speak Jewish to one parent and English to one another and to the other parent.
5. Parents speak Jewish to one another. *Both parents* speak Jewish to the children. Children speak English to them and to one another.
6. Parents speak Jewish to one another. *One parent* speaks Jewish to the children, one English. Children speak English to them and to one another.
7. Parents speak Jewish to one another and English to the children. Children all speak English to them and to one another.
8. Mostly English; every member of family speaks English mostly. Some speak some Jewish.
9. All English.

PREFATORY NOTE TO CHAPTER XV

In a comparison of 48 pupils entering first grade with at least four months of nursery school or kindergarten experience, and 41 pupils entering without such experience, an advantage of nearly six points of I.Q. in favor of the group with previous schooling is reported. The author believes that this difference is not due to selection, because the schooled and non-schooled groups were "practically equal" in age, economic and racial status, and sex proportion. Nevertheless, she points out that the difference is not statistically reliable, owing to the limited number of the cases; nor are three points *improvement* of the non-schooled group during 18 months of schooling, and a failure of the previously schooled group to register any average increase, reliably established. It would be highly profitable to amplify the study at hand with more data treated in exactly similar fashion. However, the general outcome is what might have been predicted by those who regard the I.Q. as primarily determined by heredity, because the advantage in performance gained by preliminary schooling tends to disappear as soon as the two groups compared have both had a year and a half of subsequent schooling. In other words, it is a fair presumption that the increase (if established) produced by nursery school and kindergarten training upon the I.Q. is only temporary, and, one might say, even artificial.

CHAPTER XV

THE EFFECT OF SCHOOL ENVIRONMENT UPON STANFORD BINET TESTS OF YOUNG CHILDREN

GERTRUDE HILDRETH
The Lincoln School of Teachers College, Columbia University,
New York City

The Stanford Binet Test is given to each child accepted for admission to the first grade of the Lincoln School, either shortly before his entrance or almost immediately afterward. Most of the children who remain in school for several years are subsequently retested. Whenever children are known to have been tested in other institutions before entering the first grade of the Lincoln School, an effort is made to obtain the test results.

TWO GROUPS OF ENTRANTS

All the entering pupils may be divided into two groups on the basis of the amount of school experience each child has had prior to the first Binet test. Some of the new pupils have attended nursery schools or kindergartens before entering the first grade. Others have had no previous school experience. The test records of these two groups of pupils afford data with which to study the effect of school environment upon test results.

Records were available for 48 pupils who had had two or more Binet tests and had not attended school before taking the first test. A group of 41 pupils had had two or more Binet tests and had attended nursery schools or kindergartens for a period of four months or longer before the first Binet test. The time interval between the first and second tests was approximately 18 months, on the average, for each group. The test records of these two groups constitute the data for this study. Throughout the report, the first of these groups, consisting of the pupils who had had 'no schooling' before the first test, will be referred to as Group NS. The pupils who had had 'schooling' before the first test will be called Group S.

The average chronological age of each group is six years. The two groups are practically equal in economic and social status, and sex proportion, and in both groups the racial representation is equally random. Most of the children of both groups come from superior homes and constitute a highly selected group of the general population. The majority of them are above the average of the general population in intelligence. Consequently, there is not in these groups of children a tendency for the relatively brighter ones to be sent to the kindergarten and for the duller ones to be kept at home, as is usually the case in the general population. In fact, many of the more intelligent children are supervised at home by nurses and governesses instead of being sent to the kindergarten. We would therefore expect to find, aside from chance differences due to sampling, practically equal distributions and central tendencies of intelligence in both groups of children.

Comparisons of the Two Groups in Test Results

Tabulations were made and means computed for the intelligence quotients for both groups on all tests. These results are shown in Table I. Several children had had more than four tests, but the

Table I.—Distributions of Intelligence Quotients

I.Q.	Group NS				Group S			
	Test				Test			
	1	2	3	4	1	2	3	4
160–164....	..	..	..	..	..	..	1	..
155–159....	..	..	..	..	1	..	..	..
150–154....	2	..	..	..	..	1	..	..
145–149....	..	2	2	..	1	..	..	2
140–144....	..	1	..	1	3	3	2	..
135–139....	..	2	3	2	4	3	..	1
130–134....	5	4	1	..	4	4	2	..
125–129....	4	8	2	..	7	5	4	..
120–124....	5	6	5	2	4	8	3	1
115–119....	7	7	4	..	1	4	7	2
110–114....	10	8	1	1	6	7	2	1
105–109....	9	6	5	..	3	..	2	..
100–104....	3	1	..	1	3	2	..	..
95– 99....	1	1	..	1	4	2	..	..
90– 94....	1	2	..	..	..	1	..	..
85– 89....	1	..	..	..	..	1	..	..
N.........	48	48	23	8	41	41	23	7
Mean......	113.95	117.25	120	119.3	119.635	118.91	121.5	128.4
S.D.of Mean	12.74	12.515	...	.....	15.2	14.24	.....	.....

number of such cases is too small to justify inclusion in the table. The mean I.Q. for Group NS at the first test is 113.95; for Group S, 119.64, 5.69 points higher than for the non-school group. At the second test, the mean I.Q. of Group NS is 117.25 and of Group S, 118.91, a difference of 1.66 points. Approximately half of the children had had three tests and a third of these, four tests. The means of the I.Q.'s of both groups for the third and fourth successive tests vary in small amounts from the means of the I.Q.'s of the first and second tests.

For Group NS, the difference in points of I.Q. on the average between the first and second tests is 3.3 points gain. Group S shows a loss of .73 points of I.Q. when the means of the first and second test are compared. The significance of these differences was determined by computing the P.E.'s of the differences between the mean I.Q.'s of the first and second test, for each group. For Group NS, the P.E. of the difference between the means of the first and second test is 1.74 points of I.Q. The difference found between any two measures cannot be considered a significant difference unless it exceeds at least 4 times the P.E. of the difference. Since the actual difference found between the means of the first and second tests for this group is a little less than twice as large as the P.E. of the difference, the chances are large that the difference is not a significant one. The P.E. of the difference between the means of Group S for the first and second tests is 2.198. Since the actual difference found was .73 points, it must be considered insignificant.

All the I.Q.'s of each child were paired in all possible ways, and the differences between the I.Q.'s of each pair were computed. These differences are shown in Table II. This table also includes a distribution of the differences between pairs of I.Q.'s of Lincoln School pupils of all ages in all grades, 1112 pairs of tests of 441 pupils, each of whom had been tested two or more times. The median change in I.Q. is an increase of .96 points, Q_3 is an increase of 5.71 points and Q_1 is a decrease of 3.50 points. Group NS of the young children shows a median increase in I.Q. of 3.66 points. For this group Q_3 is 9 points increase and Q_1 is 2 points decrease. These results indicate the greater probability of positive changes in the I.Q.'s of young children first tested before school entrance, than in the case of Lincoln School pupils in general. The amount

TABLE II.—DIFFERENCES IN POINTS OF I.Q. BETWEEN PAIRS OF TESTS

Points of I.Q. Change	Group NS		Group S		Unselected Lincoln School Group All Pairs
	First Test with Second Test	First Test with Third and Later Tests	First Test with Second Test	First Test with Third and Later Tests	
35 and over	..	..	..	..	2
30– 34	..	..	..	1	4
25– 29	1	..	..	..	5
20– 24	1	..	1	..	17
15– 19	3	1	1	1	33
10– 14	6	3	4	5	91
5– 9	10	8	7	3	170
0– 4	12	5	9	8	313
− 1–− 5	8	5	7	3	260
− 6–−10	4	2	10	1	116
−11–−15	2	2	..	3	62
−16–−20	..	..	..	7	29
−21–−25	1	1	2	1	5
−26–−30	..	..	..	..	3
−31 and below	..	..	..	..	2
N	48	27	41	33	1112
Lowest	−23	−24	−25	−23	−32
Q_1	− 2.0	− 2.75	− 6.25	−14.75	− 3.50
Median	3.66	3.25	.075	1.13	.96
Q_3	9.0	9.1	6.37	6.75	5.71
Highest	25	15	23	30	51
Q	5.5	5.92	6.31	10.75	4.61

of gain is, however, too slight to be considered very significant. The median change in points of I.Q. of Group S is .075 points, a slightly smaller positive change as compared with .96 points, the median change for the general group. The chances are about equal that the I.Q. of a child in Group S will increase or decrease in successive tests. For the general Lincoln School population there is more probability of increase than of decrease. The difference between the two groups is probably due to sampling.

CONCLUSIONS

No real difference is found in the intelligence quotients of a group of 48 school children when the results of testing before school experience and after a period of schooling are compared.

Similarly, with a control group of 41 children, very similar in all respects to the first group, (except that the first Binet test had in all cases been given after a period of schooling, four months or more in length), no significant difference between the average I.Q. of the first test and successive tests is revealed.

No generalization which would apply to an unselected group of young children can be derived from the results of this study based on a highly selected group of children with superior social and economic advantages.

PREFATORY NOTE TO CHAPTER XVI

For a concise statement of the methods and conclusions of this investigator the reader is referred to the excellent summary by the author appearing at the end of the chapter. It is only necessary here to note that no effects of nursery-school training upon the I.Q. were reliably established.

It would be desirable in the future to supplement this work by data on more cases (only 28 nursery-school children and 28 control children were studied).

It is interesting to compare this study with one by Woolley (noted in the summary of nature-nurture literature appearing near the end of Part II of the Yearbook). Dr. Woolley's evidence indicated a possible favorable effect of a year's attendance at a nursery school as high as eight points of I.Q. The discrepancy between the two studies may conceivably be due to any of the following points:

1. Woolley used the Stanford Binet; Goodenough used the Kuhlmann Binet.

2. The social environment of Woolley's and Goodenough's subjects may have had important differences, so that nursery-school stimulation may have had a better chance to boost the I.Q.'s of one group than those of the other.

3. The number of cases was not sufficient in either study to establish the results conclusively.

CHAPTER XVI

A PRELIMINARY REPORT ON THE EFFECT OF NURSERY-SCHOOL TRAINING UPON THE INTELLIGENCE TEST SCORES OF YOUNG CHILDREN

FLORENCE L. GOODENOUGH
Institute of Child Welfare, University of Minnesota
Minneapolis, Minnesota

INTRODUCTORY

This report is based upon a comparison of the intelligence quotients earned by 28 children before and after a year's attendance at the nursery school conducted by the University of Minnesota Institute of Child Welfare with the corresponding mental ratings of an equal number of paired controls. The study constitutes one phase of a rather extensive experimental investigation into the reliability of the Binet Test at the pre-school age level. Since the results obtained from the nursery-school children can be properly interpreted only in the light of certain results from the main study, a brief description of the experiment and a partial summary of the results will be given at this point.

THE MAIN EXPERIMENTAL STUDY[1]

The Kuhlmann 1922 Revision of the Binet Scale was given to three groups of 100 children each, aged two, three, and four years, respectively. These children were selected to constitute a representative sampling of the population of Minneapolis by matching the distribution of paternal occupations in each age group with the corresponding distribution for adult males in the entire city as reported by the 1920 census. Exactly 50 boys and 50 girls were included in each age group. All cases were re-examined after an average interval of six weeks.

In so far as they have a bearing on the study presently to be described, the results may be summarized as follows:

[1] See *The Kuhlman Binet Tests for Children of Pre-School Age: A Critical Study and Evaluation*, by Florence L. Goodenough. University of Minnesota Institute of Child Welfare Monograph Series, No. 2, 1927.

1. The mean I.Q. earned on the first test by the two-year-old group was 105.1, with standard deviation of 13.0; by the three-year-olds, 104.4, with standard deviation of 18.2; by the four-year-olds, 109.4, with standard deviation of 16.6. On the second test, the mean rating for the two-year-olds was 108.1, with standard deviation of 15.5; for three-year-olds, 107.6, standard deviation of 21.7; for the four-year-olds, 116.0, with standard deviation of 15.3. It is seen that the norms obtained by the original standardization are too lenient for this group at each of the three ages considered, and that the discrepancy is greatest for the four-year group.

2. The exceptionally high ratings of the four-year-olds are shown to be in large measure attributable to the fact that, for children of this age, the tests standardized at the six-year age level are almost exactly equal in difficulty with those at the five-year level. Since these tests are given equal credit with others in the scale, children below the age of five have more than the usual opportunity to earn credit through scattering success in the upward direction, without being correspondingly penalized by unusually difficult tests in the opposite direction. This factor is also largely responsible for the exceptionally high variability of the three-year-old group, since the brighter children of this age usually earn some credits at the five- and six-year levels and thus gain in rating by virtue of the irregular standardization, while the ratings of the more backward children who drop out at a lower point in the scale are unaffected by this inequality.

3. A high correlation was found between social status, as indicated by paternal occupation, and intelligence. The difference between the mean I.Q. earned on the first test by children whose fathers belong to the professional classes and that earned by children of day laborers amounted to approximately one and one-fourth standard deviations of the total distribution at each of the three ages considered. Intellectual differences between social classes are apparently well defined by the age of two years.

4. The ratings on the second test are in general higher than those on the first test. This tendency toward gain in rating is not distributed at random, but is greatest for the children whose fathers belong to the upper occupational groups and least for the children of day laborers. An attempt was made to account for this apparent

differential gain. No correlation was found between gain in rating and I.Q. on the first test, a very low correlation (+.28) between gain and mental age on the first test, and no relationship between gain and presence or absence of mother during the first test. It is shown, however, that the second test exceeds the first in reliability and validity. This gain in accuracy involves a change in the slope of the regression line, and, consequently, a greater difference between the means of the arrays at the two extremes when the test results are plotted against another related variable. The apparent differential gain appears to be largely, if not entirely, a function of the comparative accuracy of the two tests.

5. The reliability of the total scale, as measured by the correlation between the I.Q.'s earned on the first and second examinations, was found to be $+.813 \pm .012$. The reliability, calculated on the basis of the correlation between half-scales, corrected by the Spearman-Brown formula, averaged $+.870 \pm .018$ for the three age groups on the first examination and $+.897 \pm .016$ on the second examination.

6. The complete range of changes in individual I.Q.'s from first to second test was from + 39 to − 21. The mean algebraic change for the two-year-old group was + 2.1 points; for the three-year-old group, + 4.8 points; for the four-year-old group, + 7.5 points. Of the total group of 300 cases, 8.9 percent gained 20 points or more; 4.9 percent lost 10 points or more.

The results just quoted show that, in spite of the decided tendency toward constancy in rating which is indicated by the obtained correlation of +.813 between the I.Q.'s earned on the first and second tests, nevertheless, individual fluctuations of considerable magnitude sometimes occur. Moreover, under certain conditions, constant tendencies toward gain or loss in the mean ratings of selected groups upon retests may be expected to occur, either as a result of inequality in standardization, owing to the small number of cases upon which the norms for the pre-school ages have been based, or to unequal accuracy of the two ratings as a result of greater operation of factors unrelated to the field of measurement upon one or the other occasions. It is, therefore, necessary to exercise extreme caution in the interpretation of such changes in rating and to make certain that spurious factors such as those just mentioned are not affecting the results.

I.Q. Changes in Nursery-School Children

The 28 nursery-school children upon whom this study is based are divided into two distinct age groups; a younger group, whose ages at the time of school entrance ranged from 2 years, 0 months, to 2 years, 6 months, and an older group whose ages at entrance ranged from 3 years, 3 months, to 4 years, 2 months. All these children were given the Kuhlmann Binet tests about 6 weeks before the opening of the nursery school and were retested within a week after entrance. A third test was given near the end of the school year. The mean I.Q. earned on the first test by the younger group was 110.9, on the second test 114.1, on the third test 122.4. The mean ratings of the older group were 111.6 on the first test, 119.5 on the second test, and 128.1 on the third test. Taking these figures at their face value, and averaging the results of the two earlier tests, it is seen that, at the end of a year's nursery school training, the younger group of children had made a gain in I.Q. of 9.9 points, and that the older group had gained 12.5 points. An uncritical consideration of these figures offers strong presumptive evidence that the stimulating environment and special training provided by the nursery school may exert a distinct influence upon mental growth during early childhood. Nevertheless, in view of the results obtained by the main experimental study, it was felt that, without carefully selected comparative data, even the modicum of confidence in the results which the small number of cases and the relatively brief duration of the experiment might ordinarily inspire would not, in this case, be warranted.

Each of the nursery-school children was therefore paired off against another child not in the nursery school, but who had been given two examinations during the fall in the course of the main study. The children were matched in regard to the following characteristics: sex, age, I.Q. on each of the first two tests, interval between the first and second test, paternal occupation, education of parents, and nativity of parents. The interval between the second and third tests was also kept constant for the two groups. In making the selection, the individual, rather than the group, was considered the unit; that is, the attempt was not simply to make such a selection that the group averages on the separate variables would be equal, but the particular combination of these variables existing for each of the nursery-school children was matched as closely as possible in the paired control. With so many factors to consider, precise equalization could not in all cases be achieved, either for the individual cases or for group means, but it is believed that for the most part the differences may be considered negligible. The extent of the discrepancies may be judged from Tables I and II. Table I gives the comparative data for the first five paired cases taken in alphabetical order, while Table II shows the means and standard deviations of the several characteristics for each of the four groups.

In Tables I and II the chronological ages are recorded as at the time of the first test, taken to the nearest half month. The interval between tests is taken to the nearest week. Each of the four groups includes exactly 7 boys and 7 girls.

Table III shows the comparative standing of the groups on the third test, which was given after a year of special training for the nursery-school children. As a result of the unequal standardization mentioned before, both groups now apparently rate distinctly higher than they did on the earlier tests. This is an artifact due to the imperfection of the measuring instrument, and does not indicate true acceleration in rate of mental growth. It illustrates, however, the misleading effect of accepting changes in intelligence quotients at their apparent face value without checking spurious factors which may have contributed to the results.

The nursery-school children show a slightly more pronounced tendency toward gain in I.Q. than do the children of the control group. The difference, however, is well within the limits of chance. The two greatest individual gains were made by control-group children, and the three greatest individual losses also occurred among the control-group cases. Of the 28 nursery-school children, 26 rate somewhat higher on the third test than on the average of the first and second tests, and 2 rate slightly lower. Of the 28 control cases, 22 rate higher, 2 the same, and 4 slightly lower on the third test than on the average of the two given in the fall. When the individual members of the pairs are compared with each other, we find

TABLE I.—SUMMARY OF COMPARATIVE DATA FOR FIVE NURSERY-SCHOOL CHILDREN[1] AND THEIR PAIRED CONTROLS

Pair	Child	Sex	CA	First I.Q.	Second I.Q.	Weeks Between Tests	Occupation of Father	Occupational Class[2]	Education of Father	Education of Mother	Father's Birthplace[3]
1	B. A.	M	4– 0.5	99	115	6	Sales promoter	III	4 yrs. H. S.	3 yrs. Coll.	Norway
	C. W.	M	4– 4	98	107	6	Fur dealer	III	2 yrs. Coll.	4 yrs. Coll.	U. S.
2	C. B.	F	3–11	118	130	6	State agt. fire insurance	III	4 yrs. H. S.	1 yr. Coll.	U. S.
	M. Q.	F	3–11.5	123	133	7	Owner of Q. Tire Co.	II	3 yrs. H. S.	3 yrs. H. S.	U. S.
3	R. C.	F	3– 7.5	141	140	6	Service mgr. in filling station	III	8th grade	3 yrs. Coll.	U. S.
	E. P.	F	3– 8.5	128	155	6	Foreman in factory	III	8th grade	8th grade	U. S.
4	M. C.	F	3– 7.5	86	89	8	Milkman	V	3 yrs. H. S.	8th grade	U. S.
	H. W.	F	3– 7	94	94	7	Watch repairer (in small jewelry shop)	III	4 yrs. H. S.	2 yrs. H. S.	Germany
5	R. D.	M	2– 3.5	115	111	5	Bridge carpenter	III	Not stated	1 yr. H. S.	France
	W. A.	M	2– 5.5	114	108	5	Millwright	III	8th grade	1 yr. Coll.	U. S.

[1] First five cases taken in alphabetical order. The first member of each pair is the nursery-school child.
[2] Described in monograph previously mentioned.
[3] All the mothers and all the children were born in the United States.

TABLE II.—MEANS AND STANDARD DEVIATIONS OF PAIRED CHARACTERISTICS FOR NURSERY AND CONTROL CHILDREN

Group	Chron. Age (in Months)		I.Q. on First Test		I.Q. on Second Test		Educ. of Father (in years)		Educ. of Mother (in years)		Interval in Weeks Between Tests 1 and 2		Percentage Born in U. S.		Paternal Occupation: Percentage in Groups	Paternal Occupation: Percentage in Groups	Paternal Occupation: Percentage in Groups
	M	S.D.	M	S.D.	M	S.D.	M	S.D.	M	S.D.	M	S.D.	Father	Mother	I and II	III and IV	V and VI
Younger Nursery	26.8	1.7	110.9	12.2	114.1	13.1	12.6	5.8	12.0	3.3	7.1	1.6	86	100	65	28	7
Younger Control	26.4	2.5	111.3	11.7	113.8	14.2	11.4	5.8	11.1	3.2	6.1	0.6	79	86	57	35	7
Older Nursery	44.1	3.3	111.6	16.3	119.5	17.5	12.9	4.9	13.1	4.0	6.4	0.8	86	93	43	50	7
Older Control	44.9	4.5	111.3	12.4	122.4	17.7	12.0	4.7	11.7	3.9	6.3	0.8	79	93	43	57	..

that in 13 cases the nursery-school child rated slightly higher than the control member of the pair when the first two tests are averaged; that in 11 cases, the control child rated higher; and that in 4 cases the two ratings were exactly equal. On the third test the nursery child was the higher in 16 cases; the control child was higher in 11 cases, and the ratings of the remaining pair were equal. These differences can hardly be considered significant.

If the I.Q. is appreciably affected by nursery-school training, it is reasonable to expect that a positive correlation would be found between number of days of attendance at the nursery school and

TABLE III.—MEANS AND STANDARD DEVIATIONS OF I.Q.'S EARNED ON THE THIRD TEST

Group	Tests 1 and 2	Test 3		Weeks Between Tests 2 and 3		Mean Change in I.Q.
	Mean I.Q.	Mean I.Q.	S.D.	Mean	S.D.	
Younger Nursery . .	112.5	122.4	17.1	25.6	3.5	+ 9.9
Younger Control . . .	112.6	120.6	21.3	26.8	2.3	+ 8.0
Older Nursery	115.6	128.1	12.2	27.1	1.2	+12.5
Older Control	116.8	125.6	15.2	26.7	1.7	+ 8.8

gain or loss in I.Q. These correlations were calculated for the two groups of nursery-school children. The total range of attendance for the younger group was from 27 to 111 days, with a mean of 73 days and standard deviation of 25.7 days. The changes in I.Q. found by subtracting the mean rating on the first two tests from that on the third test ranged from − 4 to + 24, with a mean of 9.9 and a standard deviation of 8.0 points. The correlation between gain in I.Q. and number of days attendance was +.279 ±.168. The attendance of the older children ranged from 50 to 109 days, with a mean of 87 days and a standard deviation of 14.8 days. The I.Q. changes for this group ranged from − 1 to + 27, with a mean of 12.5 and a standard deviation of 9.0 points. The correlation between I.Q. change and number of days attendance for this group was −.009. Hence, no relationship between gain in I.Q. and length of attendance is indicated.

It is, however, conceivable that, even though the gross results of the test show no difference between the groups which we are justified in regarding as significant, nevertheless, the marked change in environmental stimulus might bring about a 'qualitative' change in the order of development of specific types of ability. In order to check this hypothesis, the number of nursery-school children who passed each of the various sub-tests of the scale on each of the three occasions was tabulated and compared with the corresponding figures for the paired controls. The results failed to reveal a single instance of such qualitative difference in development. As a matter of fact, the performances of the two groups on the third test were, if anything, more consistently similar than on either of the two earlier tests. As a further check on this hypothesis, the entire scale was gone over with the nursery-school teachers, who were asked to rate each test according to the amount of training given on similar or related material. The ratings given ranged from "1," which was defined as "a considerable amount of direct training on closely similar material," to "4," which signified "no training which could be expected to affect results on this test." The nursery-school children showed no superiority over the control group in regard to the particular types of performance which had been most emphasized in the nursery-school curriculum.

Summary

1. A group of 28 children who had been given a year's training in a nursery school were compared as to changes in I.Q. with an equal number of children who had not had nursery-school training. The groups were similar as regards sex, age, paternal occupation, education of parents, nativity of parents, and intellectual status at the beginning of the experiment. Intervals between tests were also kept constant. The two groups were tested by the same examiners, under uniform conditions. The Kuhlmann Revision of the Binet was used for all tests.

2. An apparent increase in I.Q. was found for both groups. Comparison with the results of a more extensive study indicates that this change is largely, if not entirely, attributable to irregular standardization of the scale at the early ages, and consequently should not be regarded as an indication of an actual increase in intelligence.

3. The average I.Q. increase made by the nursery-school children was slightly greater than that made by the control group, but the difference is too small to be reliable. The two greatest individual gains were made by control-group children, and the three greatest individual losses also occurred among the control group.

4. Practically a zero correlation was found between gain in I.Q. and length of nursery-school attendance.

5. A point for point comparison of the performances of the two groups on the individual tests in the scale failed to reveal any differences in the order of development of different types of ability which could be attributed to the differences in training.

6. Each of the separate tests in the scale was rated by the nursery-school teachers as to the amount of training given in the nursery school on similar or related material. The nursery-school children showed no greater superiority over the control group on those types of performance which had been most emphasized in the nursery-school curriculum than on the scale as a whole.

PREFATORY NOTE TO CHAPTER XVII

A most difficult problem is to obtain data upon the relation between physical condition and mental growth that can give added light on the question of factors that influence the I.Q. or the E.Q. Merely to establish a correspondence between mental status and health is not enough for conclusions as to causal influence, since superiority or inferiority in either or both might be due simply to common environmental or congenital bases, or to selection (as, for example, the prevalence of hookworm among the poor mountain whites).

The only unambiguous way to deal with the problem would appear to be as the authors of this chapter have done—follow a group of subjects for a sufficient period of time (two years in this instance) and note whether or not any changes in mental capacities concur with measurable changes in physical constitution. The results of this investigation show very little tendency toward improvement either in the I.Q. or in the E.Q. following upon improvement in physical condition as measured by several criteria.

CHAPTER XVII

THE INFLUENCE OF IMPROVEMENT IN PHYSICAL CONDITION ON INTELLIGENCE AND EDUCATIONAL ACHIEVEMENT

CAROLYN HOEFER and MATTIE CRUMPTON HARDY
Elizabeth McCormick Memorial Fund, Chicago, Illinois

I. INTRODUCTION

With the coöperation of Superintendent H. Ambrose Perrin and the public schools of Joliet, Illinois, the Elizabeth McCormick Memorial Fund, under the directorship of Miss Mary E. Murphy, has been making a three-year study of the effect of intensive health instruction on various aspects of the physical and mental growth of a group of 450 children. The present chapter comprises a report by two of the representatives of the Fund upon one phase of this investigation, viz., the effect of improvement in physical condition upon the intelligence and the educational achievement of a group of 343 elementary-school children. That there may be no confusion between these two studies, only those data having a direct bearing upon the point at issue are included in this article, except where the technique of procedure is so interwoven that separation is impossible. The main investigation is soon to be published elsewhere.

II. THE SELECTION OF THE CHILDREN

The 343 children in this study, economically and socially representing a random sample, were selected in 1923-1924 from Grades 3B, 3A, and 4B from 12 different schools in the Joliet school system. In age they ranged from eight to eleven years. Further criteria of selection included a high degree of assurance of permanency of residence, the coöperation of the parents, sufficient understanding of the English language for a comprehension of the work, American-born children belonging to the white race, an intelligence quotient not less than 75, and an absence of such physical deformities as would prevent the taking of complete anthropometric measurements.

III. The Method of Securing the Data

A series of examinations was given to each child each year during the period of the study, allowing a range of only two weeks' variation from an exact calendar year. This consisted of a stripped physical examination, anthropometric measurements of several physical traits, a psychological examination, and educational achievement tests. The physical examination and physical measurements were taken annually on the same date, the psychological as near that date as expedient, and the school tests in May. Although the three series of various tests were not made at the same time, each item marked an exact year's growth in the trait it measured.

1. The Anthropometric Measurements and the Physical Examination

The anthropometric measurements were made under the direct supervision of Dr. Baldwin, according to the technique described by him[1] and later revised in his new book now in preparation. Only four of the measurements taken will be included in this report. They were made by the same person, except for the last five months when, owing to a change, it became necessary to train a second person. In order that the details of the method used by the two persons might be comparable, measurements on the same children were made by both for one month prior to the change. Every possible effort was, therefore, made to insure the reliability of the results.

The physical examinations were made with a very few exceptions by the same pediatrician, Dr. S. C. Henn,[2] in the presence of either the father or mother from whom data additional to the general data were obtained. To insure accuracy, the results were recorded by two persons on blanks devised for the purpose of the experiment.

2. The Psychological Examinations

The Stanford Revision of the Binet-Simon Intelligence Test was given by the same psychologist to each child in the school he was attending. The usual instructions and order of administering the

---

[1] Baldwin, B. T. *Physical Growth of Children from Birth to Maturity.* State University, Iowa City. Studies in Child Welfare.

[2] Less than one percent of the examinations were made by a second pediatrician, Dr. Stanley Gibson, who had been a teacher of the physician regularly employed for this study.

tests were followed with a slight modification, in that certain questions requiring verbal responses were asked during the 'lay-outs' of the Pintner-Patterson Performance Tests (which were included in the psychological series but not here reported). Such items as giving the date, making change, making sentences, and rhymes, seemed sufficiently disconnected to permit such variation. An attempt was made to reduce to the minimum variations in judgments by recording each year verbatim responses for many of the tests and evaluating these yearly responses in reference to each other, as well as to their norms. After the final examination, each item on the three intelligence tests was carefully checked by the psychologist and all mathematical operations rechecked by a statistician.

3. The Examinations of Educational Achievement

As a measure of school accomplishment, the Stanford Achievement Test was given. Comparisons were based on Form A, Primary and Advanced Examinations. They were given and scored by members of the Elizabeth McCormick Memorial Fund staff, after they had received careful instructions in the methods of procedure.

IV. The Criteria for Physical Improvement

Since the specific problem of this paper is to determine, in so far as possible, what effect an improved physical condition has upon intelligence and school achievement, the reliability of the comparative results and conclusions will depend to a large extent upon the interpretation given to the term 'physical improvement.' In this report the approach has been made from four different angles. Each criterion utilized defines improvement in terms of its own special procedure. These four criteria are:

1. The general physical condition, as based on the judgment of the physician.
2. The condition of the tonsils.
3. Three physical traits.
4. The habit of drinking coffee.

1. The General Physical Condition, as Based on the Judgment of the Physician

The physician's judgment of a child's physical condition was based on the degree to which specific defects had a detrimental effect on his general condition at the time of the examination. A threefold classification—'good,' 'fair,' and 'poor'—resulted, since there were no children rated as 'excellent' or 'very poor.' The children were divided into four groups on the basis of their condition at the time of the first and last examinations in order to show improvement or lack of it. These four groups were:

Group I. Good condition at the time of the first and last examinations. This group is designated "Good."

Group II. Fair condition at the first examination and good at the last, or poor condition at the first examination and fair or good at the last. This group is called "Improved."

Group III. Good condition at the first examination and fair or poor at the last, or fair condition at the first examination and poor at the last. This group is called "Poorer."

Group IV. Fair condition at the first examination and fair at the last, or poor condition at the first examination and poor at the last. This group is called "Fair or Poor."

2. The Condition of the Tonsils

The method just described may be open to criticism in being partly subjective. Further analysis of the results from the physical examinations based on the more objective description of a localized condition formed the second approach to the study of physical improvement. Among the defects most emphasized by workers familiar with the development of school children, those of the respiratory tract, especially the condition of the tonsils and adenoids, probably are considered of greatest importance. The 343 children were, therefore, regrouped according to the following five-fold classification of tonsillar condition: (1) tonsils removed prior to the beginning of the experiment, 1923; (2) tonsils removed during the experiment; (3) tonsils in normal condition; (4) tonsils showing definite improvement; and (5) diseased condition of tonsils—large and cryptic, associated with a history of frequent colds as evidence suggesting necessity for removal, with medium sized and cryptic tonsils designated as borderline.

3. Acceleration or Retardation of the Rate of Growth in Three Physical Traits as Indicative of Improvement

The third method of determining physical improvement included the improvement noted in three physical traits, as indicated by acceleration or retardation of the rate of normal growth. For this analysis the following measurements were selected: (1) strength of grip, as an index of general muscular condition; (2) breadth of shoulders, as indicative to some degree of a child's general physical robustness; and (3) an increase in weight in proportion to height according to age, as a rough index of nutritional status.

4. The Habit of Drinking Coffee

Since health habits have been considered in the past an important factor in the improvement of physical condition of children, it would seem that some recognition should be given here to the topic. Hence, as the fourth criterion of physical improvement the health habit concerning the use of coffee was analyzed. The children were classified into groups on the basis of coffee drinking during the period of the experiment: (1) rarely drinking coffee, (2) daily drinking coffee, (3) increasing the amount of coffee drinking, and (4) decreasing the amount of coffee drinking.

V. The Results and Their Interpretation

In school progress, as indicated in Table I by the age-grade distribution, and in intelligence, as shown by the I.Q. distribution, Table II, the group of 343 children included in this study may be considered as approximating normal.

The results will be presented according to the criteria previously explained for judging physical condition: (1) the physician's judgment upon the general physical condition, (2) the condition

Table I.—Initial Age-Grade Status of 343 Joliet Elementary-School Children

	Grade 3B			Grade 3A			Grade 4B			Total		
Age	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total
8	19	20	39	6	7	13	0	1	1	25	28	53
9	37	23	60	33	43	76	16	17	33	86	83	169
10	12	6	18	13	18	31	23	22	45	48	46	94
11	3	0	3	4	3	7	8	9	17	15	12	27
Total	71	49	120	56	71	127	47	49	96	174	169	343

TABLE II.—DISTRIBUTION OF THE INITIAL INTELLIGENCE QUOTIENTS OF 343 JOLIET ELEMENTARY-SCHOOL CHILDREN

Intelligence Quotient	Boys	Girls	Total
160–169	2	..	2
150–159	1	..	1
140–149	1	1	2
130–139	5	..	5
120–129	5	8	13
110–119	40	37	77
100–109	62	49	111
90– 99	40	50	90
80– 89	15	21	36
70– 79	3	3	6
Total	174	169	343
Average I.Q.	105	102	

of the tonsils, (3) acceleration or retardation of the rate of growth in three physical traits, and (4) the habit of coffee drinking.

1. General Physical Condition

As mentioned in the statement of the criteria for measuring physical improvement, four groups were evolved from the results of the physician's judgment of general physical condition. These groups were labelled (1) Good, (2) Improved, (3) Poorer, (4) Fair or Poor.

All the data from the tests were first analyzed by age and sex for each of these four physical conditions, but because of the small numbers in some of the year groupings, the discussion is confined to the comparisons of the averages of the totals for each group.

Comparisons presented in Table III show that at the beginning of this experiment the four groups were for practical purposes

TABLE III.—THE INITIAL MENTAL AND EDUCATIONAL STATUS OF THE CHILDREN GROUPED ACCORDING TO GENERAL PHYSICAL CONDITION

Group	Number	Intelligence Quotient		Reading Quotient		Arithmetic Quotient		Spelling Quotient	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Good	145	104	15.0	110	15.8	109	15.0	113	14.9
Improved	131	105	11.1	108	17.0	108	13.2	111	13.5
Poorer	27	99	8.9	105	14.0	102	17.3	109	13.9
Fair or Poor	40	101	6.2	103	11.2	105	10.0	107	11.5

equal in intelligence and educational achievement, since the range of differences for the four groups was small and within the so-called normal range, though there was a fairly consistent tendency for the children in the better physical conditions to have the higher average intelligence and subject-matter quotients.

2. The Influence of General Physical Condition on Intelligence

In Table IV, which gives the average initial I.Q.'s and the average amount of gains with their standard deviations for each of the four groups, it will be noted that the differences between the groups are small, with the larger gains being consistently found with the better physical conditions, although none of the differences in gains meets the statistical standard for reliability. In only one instance is the difference even as great as twice its standard deviation, *viz.*, that between the Good and the Fair groups.

TABLE IV.—COMPARISON OF THE AMOUNT OF GAIN OVER THE INITIAL INTELLIGENCE QUOTIENT OF THE CHILDREN GROUPED ACCORDING TO GENERAL PHYSICAL CONDITION

Groups	Number	Intelligence Quotient		S.D. of Gain
		Initial	Gain	
Good	145	104	4.79	7.56
Improved	131	105	3.72	7.59
Poorer	27	99	2.44	5.10
Fair or poor	40	101	1.78	6.78

These results may be further analyzed in terms of the average rate of change between the first and third mental ages. This rate is found by dividing the difference between the two mental ages by the time intervening between the tests, in this study 24 months. For example, a child with a mental age of 104 mental months at the first test and 130 months on the final test has an average monthly mental rate of $\frac{26}{24}$, or 1.08, during the period of observation.

Reference to Table V shows that the higher mental rates here again are always found with the children of the Good and Improved groups. Though most of the differences are small, there is a difference of .19 points monthly rate between those in good physical condition, Group I, and those in fair physical condition, Group IV.

TABLE V.—COMPARISON OF THE AVERAGE MONTHLY RATE IN MENTAL AND EDUCATIONAL GROWTH IN CHILDREN GROUPED ACCORDING TO GENERAL PHYSICAL CONDITION

Group	Mental		Reading		Arithmetic		Spelling	
	Rate	S.D.	Rate	S.D.	Rate	S.D.	Rate	S.D.
Good	1.30	.431	1.27	.519	1.24	.460	.94	.321
Improved	1.25	.430	1.23	.526	1.20	.489	1.00	.341
Poorer	1.12	.310	1.11	.578	1.27	.639	.74	.318
Fair or poor	1.11	.368	1.16	.482	1.25	.497	.88	.358

On the basis of the initial I.Q., the *expected* rate of Group I was 1.04 and that of Group IV was 1.01, which would give a difference of .03 points in favor of Group I. At the expiration of the study the difference is .19 points, showing a difference of .16 for Group I, although both groups exceeded their normal expectancy. In other words, within this period of three years, with all factors fairly constant, except that of physical condition, the difference in the rates between these two groups has increased five times. The validity of this difference in mental rate is indicated by the analysis of the standard deviation of the difference, that is, 2.80 standard errors. All the other comparisons with the groups fail to show reliable differences.

In so far as may be judged by these results, there is evidence that those children whose physical condition was "good" throughout this study had a more rapid mental growth than those whose condition was "fair" throughout. Viewed from the standpoint of improvement in physical condition, a similar result is indicated from the inspection of raw scores of the other groups, but these differences are not statistically reliable.

3. The Influence of General Physical Condition on Educational Achievement

As the Advanced Examination of the Stanford Achievement Test includes subjects not in the Primary Examination, comparisons have been based on the subjects common to both tests: reading, arithmetic, and dictation. The results in this study are given in terms of monthly rate of improvement for each of these subjects. The rate here is again the difference between the initial and final subject ages, divided by the number of months intervening between tests.

Inspection of Table V shows no consistent changes for the four groups, for, while Groups I, Good, and II, Improved, have the highest rate in reading and dictation, Groups III, Poorer, and IV, Fair, have the highest in arithmetic. Analyzing the various differences, however, there is little evidence for believing that the groups are different, for only in dictation are the differences significant. The differences here between Group I, Good, and Group III, Poorer, and between Group II, Improved, and III, Poorer, are more than three times their standard error. It is interesting to note that the only definite indication of a relationship to physical condition occurs in a school subject largely dependent upon mere neural retentiveness.

It appears, then, that the improved physical condition of these children may be an effective factor in progress in dictation (spelling), but there is no indication that it is influential in reading and arithmetical achievement.

4. The Influence of General Physical Condition on Intelligence and Educational Achievement as Analyzed by Pairs

In order to obviate the discrepancies in the sizes of the groups, 27 children from each of the four groups have been paired in respect to age, sex, and initial I.Q. An analysis of these results (see Fig. I) confirms the rank relationship previously found among the larger groups, but with no reliable differences. The value here lies in the indication of the superiority of Group I, Good, over the other three groups, in which this group consistently maintains first or second rank in intelligence and educational achievement.

5. A Specific Physical Defect, as Indicated by the Condition of the Tonsils

The condition of the tonsils, as revealed by a survey of the physical examinations, showed that the children could be classified as having: (1) tonsillectomies prior to the beginning of the experiment, 1923, (2) tonsillectomies after the beginning of the experiment, (3) normal tonsils, (4) improved condition of the tonsils, and (5) diseased tonsils.

The initial status in intelligence and educational achievement, presented in Table VI, shows a small range of differences among the five classes. The children having tonsillectomies before 1923 ranked highest and the class with diseased condition of the tonsils lowest in every instance.

The interpretation of the data in Table VII shows:

1. No definite tendency appears for any group to maintain a consistent ranking in all comparisons, although the children having

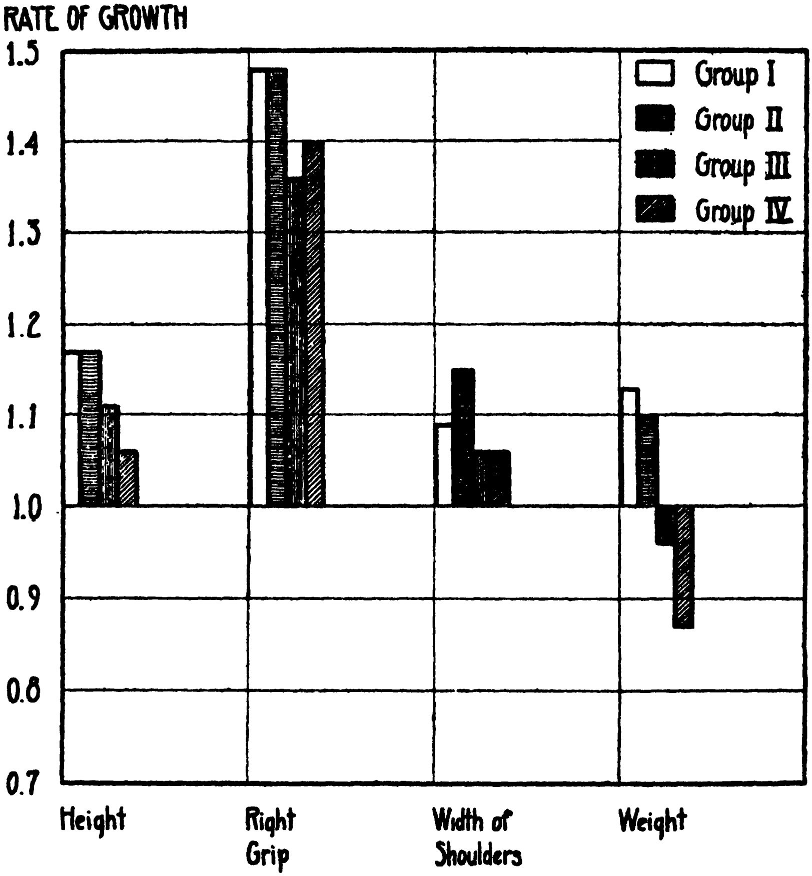


FIG. I.—THE INFLUENCE OF GENERAL PHYSICAL CONDITION ON THE RATE OF GROWTH IN PHYSICAL, MENTAL, AND EDUCATIONAL TRAITS FOR 27 CHILDREN FROM EACH OF THE FOUR GROUPS, PAIRED IN RESPECT TO AGE, SEX, AND INITIAL I.Q.

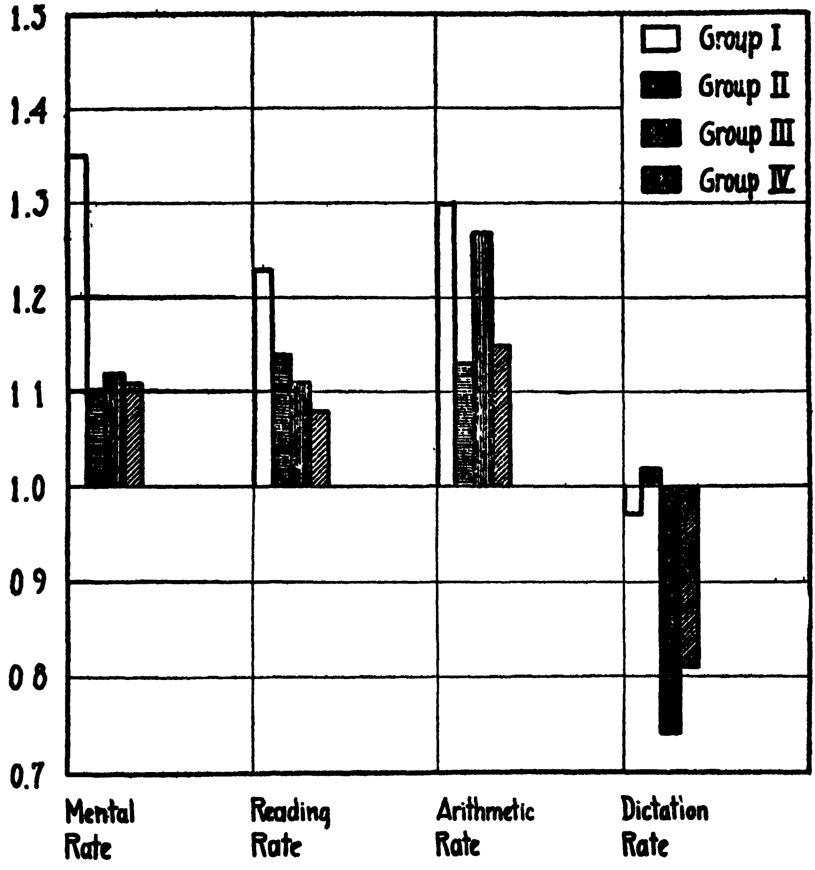


FIG. 1—CONTINUED

tonsillectomies before 1923 maintain their superiority over those with diseased tonsils in three out of the four comparisons.

2. No evidence is apparent that the good or bad condition of tonsils has any effect on intelligence.

3. When improvement in the condition of the tonsils is used as a basis for comparison, it is found that the children showing physical improvement have the highest rate in improvement in school achievement.

TABLE VI.—THE INITIAL MENTAL AND EDUCATIONAL STATUS OF THE CHILDREN GROUPED ACCORDING TO THE CONDITION OF THE TONSILS

Group	No.	Intelligence		Reading		Arithmetic		Spelling	
		Quotient	S.D.	Quotient	S.D.	Quotient	S.D.	Quotient	S.D.
Tonsillectomies before 1923	78	107	14.8	112	13.8	111	12.6	114	13.2
Tonsillectomies after 1923	64	104	11.7	107	16.4	109	11.1	110	15.0
Normal condition of tonsils	53	103	7.5	109	13.9	107	15.9	112	13.3
Improved condition of tonsils	30	102	9.8	107	14.9	106	14.1	111	18.2
Diseased condition of tonsils	118	102	13.6	106	16.7	106	13.2	109	15.8

4. The children who have had tonsillectomies before 1923 excel in reading rate of improvement the children classified as normal, *i.e.*, having an unimproved tonsillar condition.

5. Since the children with tonsillectomies before 1923 were superior at the beginning of the experiment to those in the diseased group, the final differences are more difficult to interpret. The difference in dictation is 3.26 times its standard deviation.

TABLE VII.—COMPARISON OF THE AVERAGE MONTHLY RATE IN MENTAL AND EDUCATIONAL GROWTH ACCORDING TO THE CONDITION OF THE TONSILS

Group	Mental		Reading		Arithmetic		Spelling	
	Rate	S.D.	Rate	S.D.	Rate	S.D.	Rate	S.D.
Tonsillectomies before 1923	1.28	.458	1.37	.589	1.23	.455	1.02	.345
Tonsillectomies after 1923	1.33	.510	1.25	.498	1.21	.494	.95	.373
Normal condition of tonsils	1.25	.376	1.13	.382	1.20	.529	.90	.292
Improved condition of tonsils	1.20	.341	1.21	.528	1.39	.425	1.05	.438
Diseased condition of tonsils	1.19	.374	1.18	.522	1.23	.508	.86	.316

6. Acceleration or Retardation of the Rate of Growth in Three Physical Traits

As stated previously, three physical traits were selected as a criterion for physical improvement. These are: right grip, indicative of general muscular condition; breadth of shoulders, indicative of robustness; and weight in relation to height, age, and build, indicative of nutritional status.

The children were classified, according to their rate of improvement, as "accelerated" or "retarded" with respect to each physical trait.

The rate of improvement for right grip and for breadth of shoulders was the ratio of each child's actual gain during the period of three years to his expected gain. The actual gain was the difference between the initial and final measurements, and the expected gain was the normal rate for his age and sex, derived from standard tables.

The rate of improvement in weight was based on Baldwin's standards, in metric units, for weight, height, age, and sex of children measured nude. The ratio of the actual rate of gain to the expected rate involves several computations, since the norm of expectancy in weight is based on a child's age, sex, and height. For illustration, a boy 8 years old and 125.6 centimeters tall should weigh 24.9 kilograms. Two years later, at the age of 10, his height is 137.0 and at this height he should weigh 30.2 kilograms. The difference between these two weights represents what a boy of his age and his increase in height should be expected to gain; thus, in this case 5.3 kilograms may be called his normal expectancy. However, his weight at the first examination was 25.4 kilograms and at the last examination was 33.1; thus, he made a gain of 7.7 kilograms. His rate of improvement, therefore, is $\frac{7.7}{5.3}$, or 1.45. Since there were only a few children having a rate of exactly 1.00, these cases were assigned alternately to the accelerated or retarded group.

By this method of measuring the rate of improvement the factors of age and sex have already been considered and may, therefore, be eliminated in the final tabulation of the results.

From Tables VIII and IX the following conclusions may be drawn:

1. The groups classed as "accelerated" and as "retarded" in each of the physical traits are very similar to each other in average intelligence and educational achievement.

2. In right grip and breadth of shoulders the mental rates for the accelerated and retarded groups are practically identical.

TABLE VIII.—THE INITIAL MENTAL AND EDUCATIONAL STATUS OF CHILDREN GROUPED ACCORDING TO THREE PHYSICAL TRAITS

Group	No.	Intelligence		Reading		Arithmetic		Spelling	
		Quotient	S.D.	Quotient	S.D.	Quotient	S.D.	Quotient	S.D.
Right Grip									
Accelerated.......	262	105	11.2	109	16.3	108	14.4	112	13.5
Retarded.........	81	100	13.4	105	13.5	107	11.3	109	14.5
Breadth of Shoulders									
Accelerated.......	164	104	15.0	108	13.5	107	15.3	110	16.4
Retarded.........	179	103	11.6	108	16.3	108	14.0	112	13.0
Weight									
Accelerated.......	153	104	12.7	108	15.2	108	14.7	111	15.6
Retarded.........	190	103	14.2	108	16.2	107	15.4	111	14.5

TABLE IX.—COMPARISON OF THE AVERAGE MONTHLY RATE IN MENTAL AND EDUCATIONAL GROWTH OF CHILDREN GROUPED ACCORDING TO THREE PHYSICAL TRAITS

Group	Mental		Reading		Arithmetic		Spelling	
	Rate	S.D.	Rate	S.D.	Rate	S.D.	Rate	S.D.
Right Grip								
Accelerated..............	1.25	.427	1.26	.526	1.27	.504	.97	.331
Retarded................	1.24	.398	1.14	.497	1.11	.452	.85	.347
Breadth of Shoulders								
Accelerated..............	1.24	.398	1.23	.569	1.27	.503	.94	.370
Retarded................	1.25	.449	1.23	.475	1.20	.483	.94	.311
Weight								
Accelerated..............	1.30	.454	1.25	.586	1.24	.445	.93	.348
Retarded................	1.20	.375	1.21	.472	1.23	.531	.94	.346

3. In weight the difference in mental rate favors the children who are exceeding the normal rate of growth, though this difference may be due to chance.

4. Breadth of shoulders and weight apparently have no influence on educational improvement for the accelerated and retarded groups.

5. Those classed as "accelerated" in right grip show a consistent tendency in all three school subjects to surpass those classed as "retarded." The differences in arithmetic and dictation are significant. This consistency might possibly be interpreted to mean

that those in better muscular condition may have an advantage in withstanding the effect of the fatigue element and persistence of effort due to long school hours.

It would seem, therefore, that there is very little evidence of a relationship between an accelerated or retarded rate of growth in weight and breadth of shoulders and rate in mental and educational growth, but there is some indication of a relationship between rate of improvement in right grip and school achievement. The attention of the reader is called to the fact that rate of physical growth as used here, refers to the *rate* over a stated period only and should not be confused with absolute physical status.

7. The Habit of Drinking Coffee

As a fourth method for determining improvement in physical condition, the children were classified according to those reporting: (1) rarely drinking coffee, (2) daily drinking coffee, (3) an increased amount of coffee drinking the last year, and (4) a decreased amount of coffee drinking during the period of the experiment. The results from the fourth class are presented in the tables, but are eliminated from the conclusions, since parents may have reported less coffee drinking at the last examination in order to create a favorable impression.

A study of the data given in Table X shows that in initial status the 'non-coffee' group ranked highest in every instance. The figures for intelligence are such as to warrant the belief that the non-coffee drinkers are really superior.

The data on the rate of mental growth (shown in Table XI) confirm this evidence of the superiority of the 'non-coffee' group. This general tendency raises the interesting question as to the detri-

Table X.—The Initial Mental and Educational Status of Children Grouped According to Their Use of Coffee

Group	No.	Intelligence		Reading		Arithmetic		Spelling	
		Quotient	S.D.	Quotient	S.D.	Quotient	S.D.	Quotient	S.D.
Non-coffee	171	107	9.9	110	17.4	110	13.2	113	15.1
Coffee	104	101	11.1	106	16.1	104	15.8	109	13.6
Increased Amount	30	99	7.6	105	11.2	108	10.5	109	15.2
Decreased Amount	38	101	10.1	105	16.7	107	12.8	110	12.5

Table XI.—Comparison of the Average Monthly Rate in Mental and Educational Growth of Children Grouped According to Their Use of Coffee

Group	No.	Mental		Reading		Arithmetic		Spelling	
		Rate	S.D.	Rate	S.D.	Rate	S.D.	Rate	S.D.
Non-coffee	171	1.33	.442	1.31	.531	1.28	.491	.97	.331
Coffee	104	1.17	.386	1.14	.517	1.28	.521	.90	.333
Increased Amount	30	1.14	.329	1.06	.441	1.00	.397	.99	.361
Decreased Amount	38	1.20	.383	1.25	.496	1.09	.412	.87	.358

mental effect of coffee on mental growth during childhood.

There is some indication that the progress in reading is also affected by coffee, but no observable tendency is noted in the other two subjects analysed.

Of course, improvement or a lack of it may not be due to coffee drinking *per se*, but to the entire health control and cultural status in the home, of which coffee drinking is merely one sign.

VI. Summary

Briefly summarizing the results of a three-year experiment on the influence of physical improvement on intelligence and educational achievement of 343 Joliet elementary-school children, we may say that:

1. On dividing the children into various groups on the basis of general physical condition and rate of growth in three physical traits, there appeared to be no initial difference between the groups in mental status, as measured by the Stanford Revision of the Binet Scale, and in educational status as measured by the subject scores in reading, arithmetic, and dictation of the Stanford Achievement Tests. There was, however, a consistent tendency, as shown by the averages of these two scales, for the children in the better physical condition to have higher ratings in intelligence and educational achievement. When the children are grouped on the basis of the condition of the tonsils and of coffee drinking, significant differences among the groups were found.

2. In the comparison of the average amount of gain of the final I.Q. over the initial I.Q. the differences between the groups when classified according to general physical condition were small; none

of them met the statistical standard for reliability, but the larger gains were consistently found with the better physical conditions.

3. When compared on the basis of normal expectancy in mental growth, based on the initial I.Q., with all factors fairly constant except that of physical condition, the initial superiority of those in good general physical condition over those in fair general condition as shown by the differences in mental rate increased five times.

4. In so far as may be judged by these results, there was evidence that the children whose physical condition was "Good" throughout the study had a more rapid mental growth than those whose condition was "Fair" throughout. Viewed from the standpoint of improvement in physical condition, a similar result was indicated from the inspection of the raw scores of the other groups, but these differences were not statistically reliable.

5. The improved general physical condition of these children may have been an effective factor in progress in dictation (spelling), but there is no indication that it was influential in achievement in reading and arithmetic.

6. No evidence is apparent that the good or bad condition of the tonsils had any effect on intelligence, but those children who showed improvement in condition of tonsils had the highest rate in school achievement.

7. There was very little evidence of a relationship between an accelerated or retarded rate of growth in breadth of shoulders and in weight and growth in intelligence and educational achievement, but there was some indication of a relationship between rate of improvement for right grip and educational achievement.

8. There was definite superiority in the intelligence of non-coffee-drinkers over the coffee-drinkers and also over those who reported an increased amount of coffee-drinking. There was an indication of an injurious effect of the habit of coffee drinking on the rate of improvement in reading.

PREFATORY NOTE TO CHAPTER XVIII

Summarizing the published literature upon puberty praecox with special reference to the mental status of cases, the authors conclude: "The rate of mental development tends to be normal or subnormal; it is seldom, if ever, really accelerated. Specific trends of mental development are probably not closely correlated with specific types of pathological development or functioning of the endocrine glands underlying the disorder."

The significance of this study is in the virtual independence found between mental and sexual development. While the summary has had to depend upon rough data (almost no mental test material being available for the cases investigated), marked absence of any unusual tendency toward mental acceleration despite extreme precocity of sexual development can be considered as well established. This is in line with the results of several recent investigations on the relation between intelligence and such anatomical or physiological indices as ossification of the wrist-bones, dentition, etc. Here, too, practically a zero correlation has been found.

CHAPTER XVIII

NOTES ON THE MENTAL DEVELOPMENT OF CHILDREN EXHIBITING THE SOMATIC SIGNS OF PUBERTY PRAECOX[1]

CALVIN P. STONE and LOIS DOE-KULMANN
Department of Psychology, Stanford University, California

For the most part, physicians examining cases of puberty praecox have stressed the somatic phases of this disease and only incidentally touched upon its psychological aspects. Hence, our knowledge of the mental development associated with puberty praecox is still in an equivocal state, despite the fact that this very striking disorder has received its due share of attention throughout the history of clinical medicine.

Our interest in the psychological aspects of the subject led us to make an extensive survey of original case records for the purpose of assembling all data on mental development embodied therein. We hoped to find in some histories casual remarks on school achievement, intellectual performances at specific ages, etc., from which rough estimates of the individuals' intellectual level might be made, even though specific opinions on mentality were not rendered by the medical examiners. This survey yielded a total of 190 cases presenting definite signs of puberty praecox. Of these, only 62 contained any reference whatever to mental development, school achievement, or intellectual performance.

ILLUSTRATIVE CASE

(Tierney, 1922. Case 1.) Boy of American parentage. Chronological age when examined, 4 years, 4 months. Weight at birth, 12.5 lbs. After the first year development was rapid, as is indicated by the following figures. Weight at end of first year, 21 lbs.; second year, 42 lbs.; third year, 52 lbs.; at the time of examination, 4.3 years, 81 lbs. (Childhood diseases or other pertinent data on early development not given).

[1] For a more extensive report on this subject see the article by Kulmann and Stone, *Jour. of Abnormal and Social Psych.* October, 1927.

Physical development: Height at the age of 4.3 years: symphysis to vertex, 26.5 inches; symphysis to soles of feet, 23.5 inches. (Normal standing height for a boy of this age ranges between 35 and 43 inches, with an average falling at approximately 39 inches. *This boy's height closely approximates the average height of 10-year-old American boys.*)

Weight is 81 lbs. (Average weight for a boy whose height is 53 inches is approximately 68 lbs. For a boy 4.3 years of age who is also of average height, the weight is approximately 35 lbs. *With respect to weight, this boy equals that of the average boy of 11 or 12 years.)*

Ossification of metacarpal and carpal bones is far in advance of the normal development for his age; equals that of a 12-year-old boy.

Dentition: No permanent teeth present.

Hirsute: There was an increased growth of hair about his legs and forearms, but none about his chest. The growth of pubic hair was extensive.

Genitalia: Both penis and testes equal in development to those of a normal adult (penis length, when flaccid, 3.75 inches).

Mental development: Retardation of mental development apparent. (Evidence on which this statement was based not reported by examiners.) Talked at the age of 12 months.

In the report of this case the author presents a photograph contrasting the development of this boy with the development of a normal boy of 10 and a normal boy of 4.5 years. A second photograph contrasts his genitals with those of the normal boy of 4.5 years. Two others illustrate ossification of the wrists of the patient and those of the normal boy of 4.5 years.

Etiology of Puberty Praecox

For detailed information regarding the etiology of puberty praecox, the following articles afford excellent introductions to the subject: Krabbe (1923), Tierney (1922), Gordon (1921), McCord (1914), and Lord (1899). According to the opinions of these men, this disorder results primarily from pathological growth or pathological functioning of the gonads, suprarenal glands, or pineal body.

Mental Development in Puberty Praecox

So far as we could ascertain, data from standardized tests of mental development have been reported for only six cases of puberty praecox. They are the cases of Morse (1913), Tierney (1922), Thoms and Hershman (1923), Wheelon (1925), and Gesell (1926). As already stated, about one third of the 190 cases reviewed contain some remarks on mental development. But these, for the most part, consist of such equivocal statements as: the child "was bright," "was apt at school," "was dull," "was subnormal," etc. Contrary to our expectation, the examining physician rarely offered specific comments on achievements at stated ages which might have served as a crude basis for roughly estimating the child's intellectual level. In view of these facts, it is obvious that no scientific generalizations can be made concerning the grades of intelligence normally associated with the phenomena of puberty praecox. At best we can only summarize the subjective estimates of physicians living in different countries, working with individual criteria of intelligence and ofttimes separated by many decades in point of time.

Disregarding the factors of etiology, sex, physical development, etc., and considering all of the cases together, one gets a somewhat better picture of the estimated intelligence of puberty praecox than when one takes them in small groups according to the various bases of classification. Of the 62 case studies containing remarks on mental development, intelligence is reported in 21.3 percent as above average, in 37.7 percent as average, and in 41.0 percent as below average.

In other words, it is the consensus of opinion among examining physicians that the mental development of cases of puberty praecox tends to be somewhat below the average of normal children. This conclusion coincides with statements of other writers who have reviewed the literature of this subject. Strauch (1918) says: "In striking contrast to the premature development and excessive growth of the body in precocity, is the persistence of the psychic functions in their infantile stage in most cases. The mind corresponds, as a rule, to the real age; their behavior is childish, but the premature awakening of the *vita sexualis* lends it a peculiar color. Rarely does the psychic progress take place *pari passu*. In other instances there exists a delayed mental development, a debility, imbecility or even idiocy. In the case of pre-existing weak-mindedness, neither in normal nor in premature puberty could any better development of the psychic functions be expected from these biological processes." Roger Williams (1902), who reviews 104 cases of puberty praecox, states that: "The mental qualities of these anomalous children never correspond to their sexual or bodily development; either they have the child-like psychical qualities of their age, or they are unusually dull, mentally defective, or even idiotic. . . . The higher tissues and organs also are much more slowly evolved and much less prone to precocity than the lower ones. This is especially so with the human brain, which appears not to attain its development maximum until very late in life—even up to the fiftieth year, and it is noteworthy that this organ is hardly ever affected [in the direction of precocious development] in cases of precocity."

Attempts have been made by several writers to ascertain whether specific trends of mental development go hand in hand with specific glandular disturbances. With respect to puberty praecox arising from hypersecretion of the gonads, Tierney (1922) declares that there is no "mental precocity, rather a tendency to decreased function, even to imbecility." Leiner (1920) says that the hypergonadal type shows no mental precocity. The individual is usually retarded and may never develop beyond the level of imbecility or idiocy. With respect to the ovarian type, Neurath, as quoted by Leiner (1920), reports that their psychic development is less advanced than their physical development and that they put away their dolls, etc., because of self-consciousness, rather than mental maturity. The ovarian and testicular groups assembled in our own survey contain five individuals who were said to be above the normal, four who were normal, five subnormal, and two of imbecile grade.

Tierney (1922) observes that the mentality of suprarenal cases tends to be subnormal. As evidence for a similar generalization Leiner (1920) quotes Pitman's adrenal case which was classified as an idiot, and Colcott-Fox and Bullock and Sequiera's (1905) cases which displayed marked intellectual dullness. Of our suprarenal cases, three were reported as being above normal, two normal, three dull, and one defective.

Leiner (1920) believes that mental precocity is not found in puberty praecox except in cases associated with pineal disturbances. "Mental precocity," he says, "is very rare and is found in those in whom the pineal is primarily involved, and then only in the male.

The mentality in the other types is either unaffected or retarded. The manifest mental precocity is of the child-like, imaginative form, and has no real substantial basis." The pineal cases we have considered do not conclusively confirm or refute the statement of Leiner. According to the physicians reporting, two were considered superior, two normal, two dull, and two defective.

Taking these data as they stand and those for the other glandular disturbances as well, it seems most reasonable to conclude, if any conclusion at all is permissible, that there is no evidence for genuine mental precocity regularly associated with any glandular disturbance underlying puberty praecox. Most probably, the rate of mental development is either unaffected by the glandular disorder or is retarded.

In view of the fact that no more than one-third of the cases of puberty praecox reported in the literature contain any reference whatever to mentality, and still fewer contain objective measures of mental development, it seems pertinent to urge that clinicians who have the rare opportunity of examining a case of puberty praecox add to their routine of physical and medical tests a comprehensive psychological examination. *Every case of puberty praecox should be looked upon as a miniature field of research for the clinical psychologist as well as for the endocrinologist, surgeon, pathologist, or internest.* Mental examinations will throw light on such questions as whether a spurt of mental growth is always, sometimes, or never correlated with the spurt of physical development so common in puberty praecox, and whether the mental life of a child who exhibits the somatic signs of puberty is augmented by the emotions, desires, and impulses normal to the pubescent youth of 12 to 16 years. Objective measures of intelligence will nullify the influence of the well-known size-intelligence illusion confronting those who attempt to estimate intelligence without resorting to standardized measures.

Race

While considering the etiology of puberty praecox, physicians have given very little attention to the factor of race. Aside from American-born children, with racial descent of parents seldom mentioned, we found six cases reported as Germanic, two as English, one French, one Esthonian, one Ruthenian, one Finnish, one Irish, one Russian-Jew, one Mulatto, and one American Jew. So far as we could discover, no examiner considered race an important factor in the causation of this disorder.

Heredity

It seems to be the opinion of most endocrinologists that puberty praecox is an acquired disorder, rather than one arising from a distinctly hereditary base. Although statements as to the non-appearance of puberty praecox among near relatives of the patient were found, only two instances of similar disorder within the family were reported. Robinson (1902) writes of two children brought to him for sex identification. At the time one was 2 weeks and the other 2 years old. Both had well developed pubic hair and the external genitals presented a type of development resembling pseudo-hermaphroditism. An older brother of these children had pubic hair and adult-like genitals at the age of five. Another instance was reported by Stone (1852), who states that the father of the pubescent child had become sexually mature at an early age (age not given). From the limited data at hand, however, one could not determine whether the condition of the father properly brings him under the classification of puberty praecox or whether it is only an instance of normal sexual maturity appearing early in the individual.

Summary

From the data of 62 cases of puberty praecox containing remarks on mental development, the following appear to be legitimate deductions:

1. The rate of mental development tends to be normal or subnormal; it is seldom, if ever, really accelerated. Specific trends of mental development are probably not closely correlated with specific types of pathological development or functioning of the endocrine glands underlying the disorder.

2. Such physical traits as height, weight, muscular development, strength, ossification of bones, closure of epiphyseal junctures, and dentition (?) may greatly surpass the norms for children of similar ages and, quite frequently, they surpass the extreme ranges for American school children of their respective ages.

3. Precocious development of the external genitals and the secondary sexual characters is usually found, although exceptions to the rule have been described. The extent of precocious development in various elements of the sexual organs or the secondary sexual characters is variable from individual to individual.

4. Puberty praecox is considered an acquired disorder arising on a basis of pathological development or pathological functioning of the glands of internal secretion. There is no evidence pointing to a distinctly hereditary basis.

Bibliography

Bulloch, William and Sequeira, James H. "On the relation of the suprarenal capsules to the sexual organs." *Path. Soc. London,* 56:1905, 189-208.

One original case, with eleven others presenting suprarenal tumors with precocious physical development.

Gordon, Murray B. "Disorders of the internal secretions in children." *New York Med. Jour.*, 1921, 239-243.

Discussion of thyroid and thymus disorders. Treatment of the growth and sexual changes accompanying pineal tumors, etc., in children. No case studies.

Krabbe, Knud H. "Early synestosis of the epiphyses with dwarfism in pubertas praecox." *Endocrinology,* 3:1919, 459-466.

Briefly reviews six cases which use X-ray examination and presents one original case. One illustration shows precocious physical development of patient, one X-ray of humerus and ulna, another of hand and wrist.

Leiner, Joshua H. "Pubertas praecox with special reference to mentality." *Endocrinology,* 4:1920, 369-380.

Discussion of the mental and physical characteristics of the various endocrine types of puberty praecox. One of the best references on mental development. Gives review of many case studies of puberty praecox, with an organization of the references by endocrine types.

Lord, John R. "The pineal gland: its normal structure; some general remarks on its pathology; a case of syphilitic enlargement." *Tr. Path. Soc. London,* 50:1899, 18-21.

Brings knowledge of pathology of the pineal gland up to 1899.

McCord, Caroy Pratt. "The pineal gland in relation to somatic, sexual, and mental development." *Amer. Med. Assoc.*, 63:1914, 232-235.

Discussion of effect of tumors upon the pineal gland. Experimental study on pineal deficiency in animals.

Reuben, Mark S. and Manning, G. Randolph. "Precocious puberty." *Arch. Pediat.* 39:1922, 769-785. Article continued in *Arch. Pediat.*, 40:1923, 27.

Discussion of puberty praecox from etiological point of view. Mental development treated. Second volume gives ten case studies. Photographs of two of the cases show overdevelopment of the body, attributed to pineal tumor.

Robinson, R. H. *Rep. Soc. Study Dis. Child,* 2:1902, 232.

Paragraph which tells of three members of one family presenting precocious sexual development.

Sequeira, James H. "A case of sarcoma of the left suprarenal body in a girl of eleven, who presented the phenomena of precocious puberty." *Rep. Soc. Study Dis. Child.*, 2:1902, 226.

The case reported here is covered in the reference under Bulloch and Sequeira.

Stone, Robert King. "Extraordinary precocity in the development of the male sexual organs and muscular system in a child four years old." *Am. J. Med. Soc.*, (Phila.), 24:1852, 561-564.

Report very interesting although very old. Report of physical development very detailed.

Strauch, August. "Sexual precocity in the male." *Amer. Jour. Dis. Child.*, 15:1918, 132-140.

Discussion on the mental aspect. Presentation of several cases, with one original case diagnosed as primary hyper-genitalism. Photograph of patient.

Thoms, Herbert and Hershman, A. A. "A case of sexual precocity." *Amer. Jour. Obst. Gynec.*, 6:1923, 349-350.

Report of very complete case history. Precocity attributed to hypersecretion of the ovary. Excellent illustration of patient.

Tierney, J. L. "Pupertas praecox." *Med. Clin. No. Am.*, 6:1922, 31-60.

Etiological discussion of puberty praecox. Review of early history. Excellent tabular summary of differential signs attributed to the different glandular disorders. Mental development is given considerable attention. Presentation of four original cases with eleven illustrations which show physical development and X-rays of bone development.

WHEELON, HOMER. "Precocious menstruation: observations on two cases of puberty praecox." *Endocrinology*, 9:1925, 353-371.

Discussion of the incidence of puberty praecox among females, including etiological factors and mental development. Presentation of two original cases.

WILLIAMS, ROGER. "Precocious sexual development, with abstracts of over one hundred authentic cases." *Brit. Gynec. Jour.*, 18: 1902, 85-114.

Extensive discussion of precocious sexual development in children, relation of sexual desire and mental development. Presentation of abstracts of 104 case histories with complete bibliography.

PREFATORY NOTE TO CHAPTER XIX

Clinical pictures are presented of two cases of puberty praecox, both of which are girls. The conclusion reached in the previous chapter (a summary by Stone and Doe-Kulmann of published observations upon puberty praecox) is exemplified by the cases reported in detail here; namely, there is no tendency for mental level to respond to, or keep pace with, the unusual endocrine status of the children. One of Gesell's cases in particular offers an interesting study because the child (an imbecile) had had a mental examination twice *before* the onset of puberty praecox as well as an examination nearly two years after the onset. No 'adolescent spurt' was observable in the child between an examination taken at age 8-0 (three months before she matured) and one taken at age 10-0, despite the significant change in her physiological status.

CHAPTER XIX

PRECOCIOUS PUBERTY AND MENTAL MATURATION

ARNOLD GESELL, M. D.
Director of the Yale Psycho-Clinic
New Haven, Connecticut

The nature-nurture problem is biologically bound up with the problem of maturation. It is for this reason that the experiments which alter the natural course of development in laboratory animals are full of interesting suggestion. These studies of growing organisms indicate that growth is extremely plastic, that its rate, direction, and pattern can be altered by physical, chemical, and surgical interference. Environment apparently plays an influential rôle. However, it appears that these induced modifications occur within certain limitations which are equally significative of the basic influence of intrinsic growth factors. The studies in developmental mechanics, moreover, indicate that certain features of growth are more resistant than others to extrinsic influence.

It is probable that the most fundamental laws of growth are so nearly universal that they may be sought and found in any form of life—plant or animal, chick or child. Experimental studies such as the embryologist and biologist undertake with the lower forms of life cannot, of course, be carried into the field of human genetics. This fact should make us all the more regardful of those clinical instances in which Nature herself virtually offers an experiment for our elucidation. It is in such moments of abnormality, as Goethe remarked, that Nature reveals her secrets.

In this sense puberty praecox is a phenomenon of scientific importance. It is a developmental anomaly in which the onset of puberty is markedly advanced. The average age of physiological maturation in 487 high-school girls studied by M. Abernethy was approximately 13 years, 6 months. The lowest age of maturation in this group was 10 years, and only five instances occurred at that age. We report below two well-defined cases of puberty praecox in which menstruation began respectively at 3 years, 7 months, and at 8 years, 3 months. These cases, which we shall designate as J. B. and B. E., therefore, fall entirely outside the curve of normal distribution. In J. B. the precocious displacement amounts to a whole decade.

What is the effect of such precocity upon the ordinary course of mental development? How does it influence the manifestation of endowment? Full details concerning these cases are reported elsewhere;[1] what follows is an abstract of that report.

I. J. B., Examined at 4, 5, 6, and 7 Years of Age

J. B., the younger child, is well-nourished, plump, and normal in her facial appearance. Her physique, however, resembles that of a mature woman of child stature. This child showed early symptoms of accelerated physical development. She began to walk at eleven months and at about the same time enlargement of her breasts was detected. She began to menstruate at the age of 3.5 years, and after a very brief interval regularity of menstruation became established. J. B. has attended school for several years and in her general deportment and school progress she is regarded as being an ordinary pupil. Her exceptional physiological status has had no obvious effect on her school career.

We have examined this child at approximately annual intervals for four years and are now able to consider the outstanding facts in regard to her mental and physical development. These facts are brought into comparative view in Table I. This table lists about a dozen developmental items and indicates the level of maturity attained in these items at the four successive examinations. Both physical and psychological items are included. The ratings are given in terms of "developmental age," as determined by psychometric tests or by clinical estimate. The summarizing estimate of general mental level is based on a careful clinical weighting of numerous data furnished on each examination, by an application of the diagnostic schedules in use at the Yale Psycho-Clinic. The

[1] The Influence of Puberty Praecox Upon Mental Growth. *Genetic Psychology Monographs*, No. 6, Vol. I, November, 1926. The editor of Genetic Psychology Monographs, Prof. Carl Murchison, has kindly granted permission to abstract from this article.

schedules and norms are based on a study of several hundred normal children in the early stages of development.[3]

The General Growth Graph (Fig. I) plots in a simplified way the lines and levels of development in several fields. Chronological ages are represented in months on the horizontal axis. Developmental ages or maturity levels are represented on the vertical axis. The heavy diagonal therefore represents the ideal line of normal or average psychophysical growth.

TABLE I.—APPROXIMATE DEVELOPMENTAL AGE-LEVELS, IN YEARS, REACHED BY J. B. WHEN CHRONOLOGICALLY 4, 5, 6, AND 7 YEARS

Chronological Age at Examinations	4 yrs.	5 yrs.	6 yrs.	7 yrs.
Developmental Items				
Height	5.5	7	....	10
Weight	7.5	9.5	....	11.5
Grip (right hand)	6	....	....	11
Carpal Ossification	6	4	6	7
Drawing	4	4.5	5.5	7
Drawing a man	3.5	4	6	6
Language and Vocabulary	3.5	4	5.5	6
Digit Recall and Knox Cube Test	4	4	7	10
Number Sense	3.5	4	5.5	6
Play Interests	4	5	6	7
School Report	....	....	Average	Average
Reading	....	....	6	7.5
General Developmental Level	3.5	4	5.5	6.5

It is clear from Figure I that puberty praecox has had a dislocating effect upon the total growth complex. The statural development is pursuing an unwonted course and the even tenor of mental maturation has been somewhat disturbed. There is no evidence, however, for believing that the precocious puberty has markedly accelerated general mental development. On the contrary, the mental examination at the age of 5 years showed a consistent retardation. This examination was made with great pains and there was ample rapport with our subject. The subnormality which reflects itself in her 'subnormal' five-year drawing of the man was compensated for by more rapid growth in the ensuing year. Other

[3] Gesell, Arnold. *The Mental Growth of the Pre-School Child. A Psychological Outline of Normal Development from Birth to the Sixth Year, Including a System of Developmental Diagnosis.* The Macmillan Company. 1925. 447 pp.

irregularities express themselves in the memory span for digits and the span of recall in the cube-tapping test, where the subject was definitely beyond her years.

Closely related to these irregularities are atypical characters in the emotional sphere. Clinically these characters resist precise

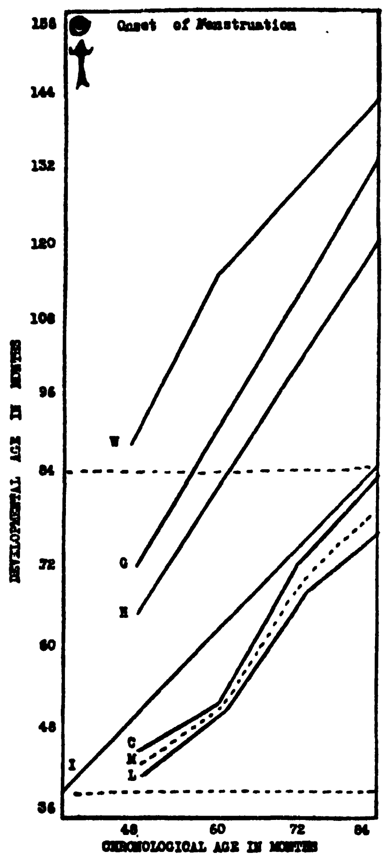


FIG. I.—GROWTH GRAPH OF J. B., SHOWING WIDE DISPERSION OF DEVELOPMENTAL COMPONENTS

Maturity levels are indicated vertically; chronological ages, horizontally in months.

In the figure W represents weight; G, grip; H, height; I, the ideal average psychophysical growth; C, comprehension; M, mental level by clinical estimate; and L, language.

Note that mental development was not accelerated by the precocious pubescence.

measurement, but they make a definite impression. J. B. speaks with a voice rich in timbre, deep in tone, and has a grave, restrained poise unusual for her years. This subdued gravity has a thick, lethargic quality which is unchildlike. This same quality asserts itself in her 'mature' adjustment to certain psychological test situations. It seems to enter into her tolerance for, and memory for, a long row of digits.

She appears to have no vivid self-consciousness. She remarks that at the age of 7 she is much bigger than her 9-year-old brother; but she plays with children of her own age. She has shown no special interest in the opposite sex or in younger children or in dolls. She is not demonstrative with her family. She amuses herself with simple forms of play. Superficially, she leads the life of children of her age; but one cannot escape the impression that the emotional complexion of this life has been altered by her unusual physiological status.

A detailed analysis of these alterations is not required for the present discussion. It is sufficient to say that in the general development of intelligence this girl has consistently cleaved to the average or fallen somewhat below in spite of the precocious onset of puberty and the regular recurrence of full menstruation.

II. B. E., Examined at 6, 8, 10, and 11 Years of Age

B. E., the older child, began to menstruate at the age of 8 years, 3 months. We had made a mental examination of this child precisely on her eighth birthday and also when she was 6 years of age, and we made one when she was 10 years of age. In this instance the precocious puberty intervened in the middle of a four-year period during which the child was known to our clinic. The displacement of sexual maturation, therefore, took on the aspect of an experimental alteration of a mental growth complex which was already under observation. We could compare with some objectivity the rate of mental development indicated by the first two psycho-clinical examinations, with that revealed at the age of 10 years. Although this girl is mentally deficient, the findings are none the less significant for the problem under discussion. If puberty is a decisive, dynamic factor in the development of the cerebral cortex, we should expect to see its influence revealed, perhaps even more patently in the retarded growth of a mentally deficient subject.

This patient first came under our observation at the age of 5 years, 9 months, when she was brought to the clinic by her mother, with diurnal and nocturnal eneuresis as the major complaint. The mental examination showed that the child was definitely feeble-minded, and that her general developmental level was slightly over 2 years. She could with a crayon differentiate between a straight stroke and a scrawl, but she could not copy a circle. She could name familiar objects, but could not name simple objects in a picture. We may express her general developmental age as being 2.25 years and her developmental quotient as 32, or, in round numbers, 30.

B. E. was re-examined on her eighth birthday, after an interval of two years. We found that she had made a perceptible gain in her mental development. She passed all but two of the test items in the two-year developmental schedule, but failed on many of the items on the three-year schedule. She now enumerated three objects in a picture, put three blocks together to build a bridge, made a crude, imperfect copy of a circle. Our clinical estimate for the second examination was summarized in the following statements: "She has barely gained six months mentally in an interval of over two years. She presents a fairly consistent picture of low-grade three-year development." Descriptively, we may express her developmental level on this second examination as being 2.75 years, which yields a developmental quotient of 34.

This means that B. E., in the interval between 6 and 8 years, was developing true to the expectation inferred from the first examination. Her gain was merely absolute, but not relative. In the relative sense she showed neither advance or decline; she was maintaining the slow, but consistent, mental growth of an imbecile with a developmental quotient between 30 and 35.

Puberty praecox supervened three months after the second mental examination, that is, at the age of 8 years and 3 months. The results of the third mental examination, made when the subject was a little over ten years of age, are therefore of critical interest. We found, on this final examination, that she was clinically maintaining a consistent behavior picture. She achieved a measurable absolute gain, but not a relative one.

She approximated a 3.5 years level of development. She could use simple descriptive words, draw an acceptable copy of a circle, and showed a slight advance in her ability to build with blocks. She passed nearly all of a score of items on the three-year developmental schedule, and many of those on the four-year schedule. The ratio of 3.5 years to 10 years, yielding a quotient of 35, is descriptive of her present mental status, and is also indicative of the same rate of development which was in evidence at the initial examination four and a half years earlier.

TABLE II.—SUMMARY OF DEVELOPMENTAL EXAMINATIONS OF B.E. AT 6, 8, 10, AND 11 YEARS OF AGE

(The figures in parentheses indicate the estimated developmental level)

	Exam. I	Exam. II	Exam. III	Exam. IV
Date	Nov. 30, 1921	Feb. 27, 1924	April 29, 1926	Mar. 29, 1927
Age	5 yrs. 9 mos.	8 years	10 yrs. 2 mos.	11 yrs. 1 mo.
Height	41.5 in. (5 yrs.)	....	52 in (10 yrs.)	54.5 in.
Weight	40.5 lbs. (5 yrs)	....	93 lbs. (14 yrs.)	98.5 lbs.
Language	Words and phrases Names 5 objects (2)	Sentences, pronouns past, plural (3)	Uses descriptive words (4)	Volume of speech somewhat greater than on Exam. III
	Points to nose and eyes Tells sex (2+)	Comprehends prepositions (3)	....	Comprehends 3 commissions (5)
Performance	Solves form-board with error (2−)	Solves without error (3−)	Builds partial gate (3+)	Attempts to fold paper diagonally but fails
Drawing	Differentiates betw. stroke and scribble (2−)	Partial copy of circle (3−)	Draws good circle (3+) Draws crude man (3+)	Very slight improvement
Personal habits	Feeds self (2−)	Puts on shoes (3)	Goes on simple errands (4)	Moves about less
Approximate developmental level	2.25 yrs.	2.75 yrs.	3.5 yrs.	3.5 yrs.
Developmental quotient	32	34	35	32

Just recently B. E., now 11 years of age, was again referred to the clinic. The re-examination gave results which are altogether consistent with the previous report and added new data concerning the maturity of her emotional level. The following excerpts are taken from the clinical memorandum.

There has been almost no perceptible mental progress in the interval of a year. Her general developmental level may still be rated at 3.5 years, yielding a quotient of 32.

The response of B. E. to the man completion test[8] furnishes very clear evidence of the mildness of the intellectual improvement which has taken place during the past year. She supplied three completions last year. She supplies no more this year. She adds two buttons by way of elaboration, and draws a very crude second leg, as a year ago, but the level and quality of the response remain virtually unchanged. This test reveals rather faithfully the fact that she has undergone very little improvement with respect to critical judgment.

A similar crudity of response is evident in the bubble test, in which she is asked to draw one bubble under the boy, two bubbles overhead, three behind. She was unable to make four bubbles in response to one part of the test, and the rating of her performance on this test remains practically unaltered. She is unable to draw a square. She is unable to draw a simple cross, either on paper or on the blackboard, although she evidently discriminates between a horizontal and a vertical stroke and brings two of them into relation.

These latter tests, therefore, have furnished very conclusive data concerning the rate of mental growth. There can be no doubt that there has been no radical alteration in her intellectual capacity or in the operations of her judgment.

The question arises whether there have been changes in the field of personality, particularly with reference to her affectivity and social reactions. Fortunately, we were able to put this to a more objective and almost experimental test.

Our method of determining her emotional characteristics was to introduce B. E. into the Observation Nursery of the clinic, which is equipped with attractive toys appealing to a wide range and variety of interests. In one corner is a large Noah's Ark on wheels. On the floor were

[8] Gesell, *op. cit.*, p. 130, Developmental Item A67.

two or three express trains with inviting strings. In another corner were a doll bed and doll; nearby a doll carriage and ironing board. In still another corner was a bookshelf with a large, attractive picture book in lithographic covers. B. E. made a relatively robust exclamation of delight as she entered the door, saying: "Gee! All the toys!" Her next remark was: "Gee! Where did you get all the toys?" For a full quarter of an hour she had an opportunity to exploit, without suggestion or interference, the manifold possibilities of this room. We have a record of her reactions, with marginal time entries which indicate that she flitted from one possibility to another, at the rate of about one shift per minute. There was no organic relation between these shifts: she pulls the trains about for approximately 30 seconds; then she takes a book from shelf and carries it across the room, puts it on a table; then she gets a purse and puts it on the table; then she pulls a little hat out of the bureau drawer, puts it on the table; pushes the doll carriage back and forth, but only for a moment; opens the Noah's Ark and plays with the animals for a somewhat longer time.

We were particularly interested to note her reactions to the doll and to the domestic corner of the nursery. Was there any increment of parental tenderness or of home-making attitude because of the precocity of her physiological maturity? We could detect no evidence of marked maternalism. She handled the doll in an uncritical way, once even seizing it by the foot. She did place the doll in the carriage, as though putting it to sleep, but somewhat later she also deposited the toy locomotive in the carriage, with perhaps as much tenderness.

The level of B. E.'s sociality likewise was determined in a simple, but rather effective manner by observing her reaction when another child came into the room. The first newcomer was a normal boy four years of age, who at once was interested to see her, scrutinized her, and then, in a very sociable manner, walked up to her and offered her his hand. His was a very well-defined social response, the pattern of which one could reconstruct clinically with confidence. The reaction of B. E. to this same situation was, however, in comparison, much more feeble, less defined, more shallow, more narrow, more transient. She walked up, gave him passing heed, and then moved on to her play. This test was repeated somewhat later when a three-year-old child walked into the room. The same type of sociality reaction took place. The three-year-old girl made an approach and evinced a definite interest in the social situation. There was a definite social outgoing on the part of the little three-year-old, but with only a weak response on the part of B. E.

There is no way of quantitatively expressing the affective reactions of these two girls, one three years old chronologically and mentally, the other of similar age mentally, but eleven years of age chronologically. The difference, however, in quality and complexity of social response was manifestly in favor of the younger child. It is our clinical conclusion, then, in this instance, that puberty praecox has not conferred any radical increase of affectivity or sociality. The curve of mental growth in this instance remains relatively unaltered even in the emotional sphere.

Comment

If pubescence has an essential dynamic relation to the maturation of general intelligence, marked cases of puberty praecox should show a measurable deviation of the usual curve of mental growth. So far as we may generalize on the basis of the two cases before us, we can conclude that precocious displacement of pubescence does not carry with it a coördinate deviation in the cycle of mental growth. Such precocity may apparently alter psychic patterns and introduce affective alterations in the attitudes and in the temperamental susceptibilities. There may even be an unusual increment in the sphere of social development, but there is no corresponding increment in the sphere of mental ability. The changes concern personality as contrasted with intellectual factors; and even these changes are not proportionate to their physiological occasion.

A psycho-clinical study of individual cases of puberty praecox confirms the dynamic importance of the endocrine complex in the determination of behavior. It does not, however, warrant the view which has been advanced that the whole period of growth may be regarded as a function of sexual development and differentiation. There is a high degree of specificity, even of independence, in the components of the growth complex. Pubescence plays its part, but not with unlimited autocracy.

The nervous system, among all the organs of the body, manifests a high degree of autonomy, in spite of its great impressionability. It is remarkably resistant to adversity, even to malnutrition. This relative invulnerability gives it a certain stability in the somatic competition between the organ systems. *It tends to grow in obedience to the inborn determiners, whether saddled with handicap or favored with opportunity.* For some such biological reason, the general course of mental maturation is only slightly perturbed by the precocious onset of pubescence.

PREFATORY NOTE TO CHAPTER XX

The data compiled in this chapter have been treated more analytically than has usually been the case in retest studies. Thirty-eight children whose measured I.Q.'s changed by as much as five points on retests were classified according to the possible causes of I.Q. shift. In all but seven cases there appeared to be some unusual factor in the child's make-up or in the conditions surrounding the test which conceivably could result in instability of intelligence quotient. It would be desirable, as a supplementary research problem, to treat retest data from the opposite point of view, *i. e.*, first pick out a group of children of unusual make-up and then investigate the stability of their I.Q.'s.

CHAPTER XX

NOTES ON FACTORS THAT MAY ALTER THE INTELLIGENCE QUOTIENT IN SUCCESSIVE EXAMINATIONS

JANET MATTHEW
AND
BERTHA LUCKEY
Psychological Clinic, Cleveland Public Schools,
Cleveland, Ohio

The Stanford Revision of the Binet Scale has been used for individual testing in the Psychological Clinic of the Cleveland Public Schools since 1920. For various reasons, more or less retesting was done; in fact, in 1924-25, 1118, or 12.8 percent, of the 8701 tests were retests, and the following year 973, or 8.1 percent, of the 11,976 tests were retests. All these examinations were conducted by the trained examiners of the Clinic staff. Whenever an examination was felt to be inaccurate or uncertain, a memorandum was made and the I.Q. was merely queried, instead of being recorded in precise figures.

In January, 1925, and again in January and in February, 1927, we selected for study certain groups of cases in which there had been applied two or more Binet tests.

In the 1925 study fifty cases were selected at random in which the retest had been given not less than one year nor more than two years after the first test. The average I.Q. on first test was 77; on retest, 79. The two series correlated .92 (P.E. .01). Twenty-four percent of the cases showed a divergence of more than seven points of I.Q. between the first and the second test.

In January, 1927, 25 cases in which a difference of five points or more appeared in the retest were selected for intensive study, including scrutiny of the child's scholastic and medical records, interviews with child, teacher, and principal. Only 17 of the cases turned out to be fully available for our purpose. These 17 are among the 38 reported in this article.

In February, 1927, a correlation of .74 was found in the I.Q.'s of 100 pupils selected at random. This correlation, lower than the one previously obtained, is doubtless due to the policy, put in force in 1925-26, of retesting only such children as seemed almost certain, in the opinion of teachers or original examiners, to show an altered I.Q. on retesting. Most of the cases had I.Q.'s between 70 and 90. Twenty one of them showed an alteration of from 11 to 30 points of I.Q. on the retest. These 21 cases are included with the 17 already mentioned, to constitute the 38 cases reported in this article. The salient features of each of the 38 cases are set forth briefly, with the idea of showing the sorts of factors that appear to be at work in bringing about shifts of I.Q. on retesting in public-school work. Roughly, the 38 cases may be classified, in terms of the factor that seemed most prominently concerned, thus:

I.	Cases presenting linguistic difficulties	14
II.	Cases presenting psychopathic tendency	6
III.	Cases presenting medical problems	9
IV.	Cases presenting behavior problems	5
V.	Cases not classifiable	4
	Total	38

I. Cases Presenting Language Difficulties

Case 1

Tested 9-24-24	Mental Age 4-6	I.Q. 75
Tested 10-2-25	Mental Age 6-2	I.Q. 85

Original examiner states: "Immature, foreign, suspect defective hearing."

Second examiner states: "Evidently language was factor in earlier failure."

Case 2

Tested 10-8-24	Mental Age 4-10	I.Q. 78
Tested 10-7-25	Mental Age 6-8	I.Q. 108

Language evidently cause of first failure. All failures were on language questions.

Case 3

Tested 10-1-25	Mental Age 4-8	I.Q. 88
Tested 5-26-26	Mental Age 6-4	I.Q. 106

Foreign. Failures due to language. Original examiner requested retest.

Case 4

Tested 10-2-24	Mental Age 4-6	I.Q. 75
Tested 10-2-25	Mental Age 6-4	I.Q. 89

Original examiner states: "Decidedly foreign. Demoted to kindergarten to acquire the language."

Second examiner states: "Very poor work in 1B. Shy, quiet type. Must pull every response. Ability to draw diamond suggests higher I.Q."

Case 5

Tested 11-20-22	Mental Age 4-0	I.Q. 60
Tested 10-21-25	Mental Age 8-2	I.Q. 85

Original examiner states: "Extremely foreign. Used interpreter for all tests. Still feel he should not be excluded until he has been longer in school environment. Teacher says he learns fairly well. No blank sent. Retest in a year."

Case 6

Tested 12-11-24	Mental Age 3-6	I.Q. 57
Tested 10-16-25	Mental Age 5-6	I.Q. 80

Original test questioned by examiner. Very evidently the original low I.Q. was due to language.

Case 7

Tested 3-9-23	Mental Age 4-0	I.Q. 78
Tested 4-14-24	Mental Age 4-6	I.Q. 72
Tested 1-19-25	Mental Age 5-1	I.Q. 72
Tested 4-12-26	Mental Age 7-0	I.Q. 85

Original report: "Provisional test due to language and home factors. Very phlegmatic."

1924 report: "Child has not learned a rhyme so far. Very stupid in school. Would exclude. Very small for age."

1925 report: "Note previous tests. Nothing after one year in 1B. Also counting coins, patience, mutilated pictures, all just passed and no more. A retest should show a drop in I.Q. Recommend special class now. Has the appearance and behavior of a true defective."

1926 report: "This boy has been fitted with glasses and his progress has been remarkable. Unquestionably vision was the chief factor responsible for his failure in school and in the three previous Binet tests. This defective vision was not discovered by the school doctor or nurse, although the boy was given a physical examination. A sight-saving candidate. The teacher reports the boy is partly drunk practically every day causing him to become listless, inattentive, indifferent, and dopey."

Case 8

Tested 10-20-24	Mental Age 4-0	I.Q. 62
Tested 5-19-26	Mental Age 6-10	I.Q. 84

Original test questioned. Language.

Case 9

Tested 5-14-25	Mental Age 4-2	I.Q. 73
Tested 4-12-26	Mental Age 5-6	I.Q. 84

Results show language difficulty. Difficulty is overcome by school environment.

Case 10

Tested 4-23-25	Mental Age 4-0	I.Q. 75
Tested 6-7-25	Mental Age 5-8	I.Q. 87

Failure due to language and home training.

Case 11

Tested 4-23-25	Mental Age 4-6	I.Q. 82
Tested 6-7-26	Mental Age 6-6	I.Q. 98

Language difficulty cause of first failure.
Retest requested by examiner.

Case 12

Tested 3-13-24	Mental Age 6-8	I.Q. 79
Tested 10-13-25	Mental Age 9-0	I.Q. 90

Original report: "Candidate for sight-saving class."

1925 placed in sight-saving class. "Has made an unusual gain although he is still handicapped by language. Would advise another promotion."

Case 13

Tested 5-15-24	Mental Age 4-6	I.Q. 76
Tested 3-25-25	Mental Age 6-11	I.Q. 101

Physical record: 6-10-25, tonsillectomy; 5-7-25, rickets.
Nutrition minus four pounds.

Is a dull looking child. His head is slightly rickety. Looks slower than I.Q. 101 indicates.

School record: Non-promoted 6-19-25. Effort only fair. Scholarship poor but improving. Variation in I.Q. may be due to language. Home Hungarian, very little English spoken.

Case 14

Tested 11-10-21	Mental Age 3-10	I.Q. 62
Tested 4-20-26	Mental Age 9-1	I.Q. 85

Father talks some English. Mother only Slavish. Probably language difficulty at time of first test.

Physical record: 11-14-21, height 44 in., weight 53 lbs., 11 lbs. overweight; 11-27-22, height 46½ in., weight 60 lbs., 12 lbs. overweight; 12-10-24, height 51 in., weight 79 lbs., 18 lbs. overweight.

Nutrition eight pounds overweight. Still looks overweight. Very rosy cheeks. Gazella has probably reached the limits of her mental capacity now. She has been demoted twice.

II. Cases Presenting Psychopathic Tendency

Case 15

Tested 9-22-22	Mental Age 6-6	I.Q. 76
Tested 10-29-25	Mental Age 7-4	I.Q. 69
Tested 6-05-26	Mental Age 8-3	I.Q. 73

In 1921 and 1923 physical examinations were normal. In 1925 was ten pounds underweight.

Five times non-promoted. Seven times promoted. Scholarship always poor. Effort only fair. Attendance good. She was doing about 3B academic work in special class when she was sent to a private school in Pittsburgh. Her tendencies are decidedly psychopathic.

Case 16

Tested 2-18-25	Mental Age 5-0	I.Q. 94
Tested 2- 1-26	Mental Age 5-0	I.Q. 81

Very scattered test. Original test basal 3. Success through eight years. Second test basal 4. Success through the seventh year. Same mental level for both tests. Failed on comprehension in last test which he passed in the original test. Erratic responses show psychopathic tendencies.

Case 17

Tested 10-3-24	Mental Age 5-0	I.Q. 81
Tested 5-11-26	Mental Age 7-2	I.Q. 93

Physical record: 12-17-21, height 43 in., weight 37 lbs. Nutrition four pounds underweight. Teeth very bad. 11-4-26, height 47 in., weight 44 lbs. Nutrition six lbs. underweight. His father is alcoholic and insane. Brother sent to reform school. 3-17-27, was referred to an open-air school. No further report. His scholarship is only fair. Demoted to 1A. He seems to have no number sense. He is a behavior problem, trouble maker. Erratic and unreliable, according to teacher.

Case 18

Tested 4-10-22	Mental Age 6-0	I.Q. 87
Tested 4-24-24	Mental Age 8-4	I.Q. 94
Tested 12-15-25	Mental Age 8-4	I.Q. 79

Was referred to Child Guidance Clinic 10-15-26 because of his peculiar behavior. Their findings are: height 54 in., weight 67 lbs. Four percent underweight. Two inches above age average in height. Very dirty. Scars of old scabies on back. Slight impairment to hearing. Nasal obstruction. Tonsils enlarged. Shows signs of previous infection. Child Guidance Clinic says his behavior is not psychopathic. He suffers from inferiority. Is antagonistic in school. Environment very bad. The scattering in mental test is over 7 years.

Case 19

Tested 10-12-23	Mental Age 6-2	I.Q. 89
Tested 10-15-25	Mental Age 6-10	I.Q. 77

Original tester states: "Child behavior problem. Cruel to children. Had hard time testing him. Cried and refused to do anything."

Second tester states: "Child does not like to try, especially if a thing is a little hard. A very odd-behaving boy."

Test and remarks show child is probably psychopathic.

Case 20

Tested 2-6-23	Mental Age 7-4	I.Q. 76
Tested 3-11-26	Mental Age 8-2	I.Q. 64

Very slow gain. Only a half year in three years. Father's brother has been in insane asylum. Child very small for age. Evidence of malnutrition.

III. Cases Presenting Medical Problems

Case 21

Tested 2-14-22	Mental Age 5-2	I.Q. 71
Tested 4-25-24	Mental Age 7-2	I.Q. 76
Tested 11-1-26	Mental Age 10-10	I.Q. 91

Was sent to Fresh Air Camp 9-2-26. Her father is an alcoholic and her mother tubercular. At that time her nutrition was very poor. 9-15-26 tonsillectomy. Is not an active child. Behavior is good. Does fair 5B work. The girl has been at Fresh Air Camp for 12 weeks. She was twenty-one pounds underweight and has gained twelve pounds since she has been there. Worker feels that the child has improved a great deal since she has been with them. Her appetite has improved. Tonsillectomy accounts for slow gain at first. Tubercular 4 plus.

Case 22

Tested 5-31-23	Mental Age 4-8	I.Q. 76
Tested 2-18-24	Mental Age 5-2	I.Q. 75
Tested 11-5-25	Mental Age 6-0	I.Q. 70

In 1923 was three percent underweight. In 1925 was normal. Very little information, as boy had left school. Always gave a great deal of trouble. From a very shiftless family.

Case 23

Tested 11-21-21	Mental Age 1-7	I.Q. 70
Tested 2-27-23	Mental Age 7-8	I.Q. 67
Tested 3-20-26	Mental Age 10-10	I.Q. 75

Physical examination made 5-24-21. Height 53 in. Weight 59 lbs. Seven lbs. underweight. Very bad tonsils. 3-15-23 tonsillectomy. 5-5-26 height 63 in., weight 115 lbs. Three lbs. overweight.

Case 24

Tested 4-6-22	Mental Age 8-11	I.Q. 75
Tested 2-21-23	Mental Age 8-4	I.Q. 65
Tested 2-18-26	Mental Age 9-8	I.Q. 61

Failed in 1923 in ten-year test where he had had previous success. Three different examiners. Marked stigmata. Lack of mental gain. May be a case of having reached mental level.

3-9-23, doctor's report suggested presence of hypothyroidism.

Case 25

Tested 4-27-25	Mental Age 4-2	I.Q. 62
Tested 4-22-26	Mental Age 4-2	I.Q. 51

Comparison of tests shows both had same basal age of 3.

Same failures in 4th year, except that he could count only to 4 on the first test and could not count at all on the second test. Same failures in 5th year. Knew coins in the 6th year in the last test. No mental gain in 1.5 year's time. Marked stigmata. Difficult to test. When in school did nothing. Language a decided factor. Decrease in I.Q. probably due to deterioration in the child.

Case 26

Tested 11-21-21	Mental Age 5-0	I.Q. 84
Tested 5-1-25	Mental Age 6-6	I.Q. 68

Malnutrition. General physical make-up poor.

Three terms 1B, two terms 1A, two terms 2A, two terms 3A. Placed in special class.

Case 27

Tested 12-14-21	Mental Age 7-4	I.Q. 85
Tested 10-5-23	Mental Age 8-6	I.Q. 80
Tested 5-20-26	Mental Age 9-5	I.Q. 70

Gradual deterioration. Physical condition very poor. Lack of mental ability more apparent in schooling than in the test.

Case 28

Tested 9-27-21	Mental Age 5-8	I.Q. 99
Tested 10-28-25	Mental Age 7-2	I.Q. 73

Hereditary syphilis. No gain in school work. No scattering in mental test.

Case 29

Tested 2-14-22	Mental Age 9-0	I.Q. 80
Tested 11-5-25	Mental Age 8-10	I.Q. 59

In 1923 an illness diagnosed as *petit-mal*. 11-30-25 had had no spells for two years. Nutrition has always been poor. Vision questionable. One eye very peculiar. School record fair. No scattering on mental test.

IV. Cases Presenting Behavior Problems

Case 30

Tested 9-22-25	Mental Age 4-4	I.Q. 90
Tested 9-13-26	Mental Age 6-6	I.Q. 111

Was tested to see if she could enter kindergarten under age. Not entered until March, 1926. In kindergarten her handwork was very poor. Is a healthy looking, pert, spoiled child. Her failure on first examination was probably due to stubbornness and temper. In the beginning she would do nothing in the first grade. Now she responds more readily and is in the first class.

Case 31

Tested 4-13-25	Mental Age 6-4	I.Q. 115
Tested 11-5-26	Mental Age 7-8	I.Q. 108

6-15-26, height 42 in., weight 46 lbs.

Scholarship only fair in kindergarten. Good in 1B. Probably a behavior condition caused slump. Attention is very easily distracted. Mother does everything for the child. Takes her to a great many movies. Her interests are not in school. General unrest. The principal and teachers feel she is capable of much better work.

Case 32

Tested 6-3-25	Mental Age 5-2	I.Q. 91
Tested 5-5-26	Mental Age 6-4	I.Q. 97

11-25-26, height 46 in., weight 48 lbs. Teeth corrected. Tonsils need attention.

In kindergarten 4 terms. Very lackadaisical. Absent a great deal of the time. His parents evidently encourage him in this.

Case 33

Tested 1-25-24	Mental Age 4-6	I.Q. 74
Tested 4-1-26	Mental Age 7-0	I.Q. 85

Original examiner states: "Very immature. Gives up easily. Should really do more than test shows."

Case 34

Tested 3-9-25	Mental Age 8-1	I.Q. 136
Tested 6-17-26	Mental Age 8-4	I.Q. 115

Original examiner states: "Very keen youngster. Place in 1B at once and advance according to accomplishment and health. Credit for ball and field in 12."

Second examiner states: "Grades in school only G. G. G. P. Has done very little this term. Grade P. A retest asked for. Has an older brother in High I.Q. who does well. Gets much help at home. He cannot read; only very simple words."

Some of his success on the original test, such as the ball and field, may be due to coaching.

V. Cases Not Classifiable

Case 35

Tested 2-14-24	Mental Age 6-4	I.Q. 75
Tested 11-1-26	Mental Age 7-4	I.Q. 66

11-5-26, height 57 in., weight 77.5 lbs. Nutrition 3.5 lbs. underweight.

In kindergarten, eight years of age. 1A, ten and one-half years. Was in overage 2B. Now being placed in special class.

Case 36

Tested 3-5-25	Mental Age 4-10	I.Q. 91
Tested 1-21-26	Mental Age 6-6	I.Q. 105

5-27-26, height 45 in., weight 44 lbs. Nutrition three lbs. underweight. Teeth and tonsils bad.

12-1-26, sent home with pediculosis. A very dirty child. Effort only fair. Scholarship poor.

Case 37

Tested 11-5-23	Mental Age 6-10	I.Q. 113
Tested 6-17-26	Mental Age 10-6	I.Q. 121

Nutrition normal. Has trouble with his eyes. Had pink eye when he was five years old. Hard working, serious, and direct. Studies piano, Hebrew, and reads history. A typical Jewish child pushed to the limit.

Case 38

Tested 3-9-25	Mental Age 4-10	I.Q. 86
Tested 3-23-26	Mental Age 5-0	I.Q. 75

Decrease in I.Q. probably due to the fact that the examiner of the original test allowed credit for language.

Original examiner said: "Child looks scared to death all the time. Never smiles. Wretched home. Mother drunk, so neighbor says. Would recommend exclusion were not home conditions so bad. Kindergarten teacher considers him better off in school for any length of time. Retest a year from now."

PREFATORY NOTE TO CHAPTER XXI

This chapter summarizes the author's investigation of the effect of coaching upon Stanford Binet scores. The results show that two hours of special training upon the actual material used in the test may bring about an average net gain in *ability to score* of 20 to 30 I.Q. points over gains made by control groups; but that two hours of training upon material similar to, but not identical with, that of the test causes an average net gain of only four or five points. At subsequent retests the I.Q. level, both of the groups coached on identical material and of those coached on similar material, diminished. By the end of three years the coached groups showed no significant advantage over the control groups—a fact which the author attributes partly to forgetting, and partly to the use of Binet material on higher levels than those upon which the children had been coached earlier.

Changes in mental age are also discussed. However, there is some question as to the interpretation of this latter discussion, since the author allowed for natural mental growth due to age by subtracting one month of mental age for each month increment of chronological age, regardless of the I.Q.'s of the children, though the average I.Q. of several groups was well below 100.

The special significance of the study lies in its demonstration that many tasks not ordinarily learned by children of a given mental level *can* be learned if intensive measures are used to teach them. It is, of course, in providing positive or negative evidence upon this very point that the bearing of coaching studies upon the *nature-nurture* problem is found. Should the efficacy of coaching, even upon material nearly identical to test material, prove very slight, that would be presumptive evidence that the differential rate of mental development of children was conditioned almost entirely by nature. If, on the other hand, coaching could greatly or permanently stimulate performance, it might properly be inferred that other environmental training, even though less specific in character, might have a definite, measurable effect upon ability.

CHAPTER XXI

THE INFLUENCE OF SPECIALIZED TRAINING ON TESTS OF GENERAL INTELLIGENCE

KATHARINE B. GREENE
University of Michigan, Ann Arbor, Michigan

The experiment reported in this article is a sequel to the work described in "The Influence of Specialized Training on Tests of General Intelligence."[1] Since then it has been possible to retest a considerable portion of the subjects after a three-year interval.

In experiments with growing children that extend over a considerable period of time it is necessary to postulate the normal or expected gains in growth in order to eliminate its part in the total results, leaving a clearer presentation of the effects which we wish to measure. There are two possible methods of making allowances for this expected growth.

First, we may assume that the I.Q. calculated from the original test of an individual indicates his normal rate of growth. On this basis we may calculate his expected growth in units of mental age. Second, we may assume that all individuals in a group are progressing at approximately the average rate of a normal group. Both methods have certain advantages. The first will probably yield the more accurate results, assuming that the first test is a valid indication of his probable growth. Yet other investigators have found that a single test has certain inaccuracies. Rugg[2] says: "The chances are one in two that an I.Q. from a single test will increase as much as six points, or decrease as much as three; that the chances are one in five that it will increase as much as twelve, or decrease as much as six; and that the chances are one in twenty that it will increase as much as eighteen or decrease as much as nine."

In order to get rid of this error due to the inaccuracy of the first test results, the most impartial method would be to say that

[1] Graves, K. B., Teachers College Contributions to Education, 143.

[2] Rugg and Colloton, "Constancy of the Stanford Binet I.Q. as shown by retests." *Jour. of Educ. Psych.*, September, 1921.

all had an equal gain during the period under investigation. The relationship between original success and subsequent improvement in a set task is so little explained at the present time that any other method is open to many dangers from errors made under misapprehension of this relationship. If we were to assume that all had gained equally, it would be best to say that each child gained as though he were a child progressing at the normal rate of growth; that is, a child whose I.Q. was 100 for any and each interval of time. The average I.Q.'s for the three groups varied less than seven points from the average I.Q. of the entire group.

This was the method adopted, *i.e.*, each month of chronological gain was assumed to mean a month of mental gain. By doing this two chances for error are eliminated. The personal bias of the tester is not felt, since the tendency of a tester to give more credit than another would have given does not lessen the credit for the amount actually gained by her subjects subsequently. The condition of the child at the time of the first test does not influence the amount of estimated gain. This method is both impartial and simple. It makes allowance for various disturbing factors with the least possible injury to the true scores.

The first mental age achieved by the child in the first test given in this experiment is taken as the basis for all calculations. It is not, of course, a strictly accurate measure of the child's mental ability. But as the inaccuracy is distributed by chance, we can not eliminate it from these results, since we could not give a second test to make the score a more reliable measurement.

For example, the first retest (second test) came just three weeks later. The effect of this period on mental age is deduced from its effect on chronological age. If the child is regarded as having gained a month chronologically (calculating the monthly birthday nearest to the date of the test), then the three weeks is regarded as a period of one month's advance. If there is no change in chronological age (as calculated to the nearest monthly birthday), the influence of this period is disregarded. In the majority of cases, one month has been added to the chronological age. Then, one month is subtracted from the total number of months made on the second test, in order to make this score commensurate with the first, so far as the influence of time on growth is concerned.

A similar method is used in treating the later test results from both School A and School Y. When the results from both are translated into terms of the original scores, A and Y can be combined, though the difference in length of time elapsing between the tests had made this impossible before.

The groups tested were divided into three sub-groups. One group, here called the Control Group, was left alone between the tests, and no attempts were made to influence their training.

One group was trained between the period of the first and second tests in the material of the Terman Revision. This group had direct coaching, and is here referred to as the Coached Group. The coaching period for this group was divided into periods of group and individual coaching. Three or four children of similar mental age were taken in a group and coached on each test in which they had failed, until two or three were able to give the correct answer, and it seemed a waste of time to keep the successful children waiting while the other child was practised. Each child was then scored plus or minus on his ability to pass this particular test at that time, and the group was directed to the next 'failed' test.

The next period each child had to himself for practice on his own failures. If he seemed to grasp the right answer within the time limit allotted by the coacher to that problem, he was graded plus, and no further reference was made to that test during the period. In the third period the child reviewed his original failures, and practised once more on those he still missed. In this way the time allotted to coaching each test was spread out to avoid fatigue and also to re-impress those answers which were almost but not quite learned. The material used in Schools A and Y for coaching purposes was the same, except that the children at Y had two extra years of material and were coached on Years VI to XII, inclusive, of the Stanford Revision of the Binet-Simon Test, while those at A were coached on Years VI to X inclusive.

The group here called the Similar Group was also coached in the period between the first and second tests. This coaching was in the form of three different sets of form booklets made up of material similar to the material in the Terman Revision. Each booklet was presented three times. The first time the children did the work as well as they could. The second time, the teacher and tester helped the children to get the right answers and to write them down. The third time the children did the work without asking for help from the teacher or tester.

The children in the whole series, including all members of the three sub-groups had, at the beginning of the experiment, an average mental age of seven years and 4.85 months, with a standard deviation of 8.64 months. The report of the results in one of the two groups at the end of the first year is given in Table I.

TABLE I.—RESULTS, IN MONTHS OF MENTAL AGE, AT THE END OF THE FIRST YEAR

(Ages equated to the first test.)

Test	Control Group (21) Average	S.D. of Dist.	Coached Group (28) Average	S.D. of Dist.	Similar Group (22) Average	S.D. of Dist.
I	86.47	7.9	86.61	8.22	92.54	6.30
II	90.24	7.9	109.57	6.68	100.41	7.86
III	92.05	9.6	105.25	7.05	102.05	7.70
IV	90.74	10.03	95.71	6.83	97.27	8.47

Test I made at the start.
Test II made three weeks after Test I.
Test III made three months after Test I.
Test IV made twelve months after Test I.

At the end of the three-year period there were 88 of the 153 children originally reported in the results for Schools A and Y who were still available for testing. If the results for these children are reported in the same manner, we have the results reported in Table II.

The same material, expressed in terms of intelligence quotients, appears as Table III.

The results bear out the conclusions that were made in the previous summary of this survey.[3] There were in all 88 cases which were available at the end of the three-year period, though there were 153 available in the earlier survey at the end of the Test IV. In order to see the similarity between the groups, Table IV was constructed.

The first row, in the results for each test, gives the number of cases, the average and the sigma of distribution for the mental ages of the group at the end of the first year; the second row gives

[3] Graves, K. B. *Op. cit.*

TABLE II.—RESULTS, IN MONTHS OF MENTAL AGE, AT THE END OF THE THIRD YEAR

(Ages equated to the first test.)

School A

Test	Control Group (9) Average	S.D. of Dist.	Coached Group (18) Average	S.D. of Dist.	Similar Group (17) Average	S.D. of Dist.
I	83.77	8.35	85.50	7.19	92.82	6.87
II	89.88	7.39	109.06	6.25	100.24	7.92
III	88.66	6.43	104.67	8.13	104.18	8.19
IV	86.22	8.18	94.88	6.93	98.41	8.68
V	81.11	17.07	86.22	12.22	99.88	17.29

School Y

Test	Control Group (17) Average	S.D. of Dist.	Coached Group (11) Average	S.D. of Dist.	Similar Group (16) Average	S.D. of Dist.
I	87.06	7.75	89.45	6.61	92.75	5.93
II	88.82	8.63	121.27	12.36	99.19	8.55
III	86.94	7.38	104.91	11.65	96.13	7.00
IV	86.80	7.91	104.72	14.55	95.38	7.67
V	84.41	14.70	94.64	21.36	90.50	16.12

Test I made at the start.
Test II made three weeks after Test I.
Test III made thirteen to eighteen weeks after Test I.
Test IV made twelve months after Test I.
Test V made three years after Test I.

TABLE III.—DATA OF TABLE II IN TERMS OF I.Q.

School A

Test	Control Group (9) Average	S.D. of Dist.	Coached Group (18) Average	S.D. of Dist.	Similar Group (17) Average	S.D. of Dist.
I	82.33	12.17	84.22	10.56	101.35	11.92
II	88.22	7.24	107.94	13.64	109.47	12.97
III	87.78	12.43	103.17	13.34	113.41	14.91
IV	86.56	13.39	94.28	11.58	106.76	13.09
V	85.44	16.47	88.67	11.62	106.71	17.47

School Y

Test	Control Group (17) Average	S.D. of Dist.	Coached Group (11) Average	S.D. of Dist.	Similar Group (16) Average	S.D. of Dist.
I	98.05	11.92	98.55	8.39	101.06	8.09
II	100.18	13.08	133.09	13.73	107.81	11.00
III	97.76	10.29	114.55	12.27	104.31	8.06
IV	100.40	9.67	113.73	14.35	106.88	7.69
V	96.18	12.82	102.82	17.84	98.75	13.34

TABLE IV.—COMPARISON OF THE AVERAGE MENTAL AGES IN MONTHS OF THE GROUPS USED IN THE REPORTS

(Ages equated to the first test. Tests are given at the same interval as reported in Table II.)

School A

Test	Control Group Average	S.D. of Dist.	Coached Group Average	S.D. of Dist.	Similar Group Average	S.D. of Dist.
I.	(21) 86.47	7.9	(28) 86.61	8.22	(22) 92.54	6.30
	(9) 83.77	8.35	(18) 85.50	7.19	(17) 92.82	6.87
II.	(21) 90.24	7.9	(28) 109.57	6.68	(22) 100.41	7.86
	(9) 89.88	7.39	(18) 109.06	6.25	(17) 100.24	7.94
III.	(21) 92.05	9.6	(28) 105.25	7.05	(22) 102.05	7.70
	(9) 88.66	6.43	(18) 104.67	8.13	(17) 104.18	8.19
IV.	(21) 90.74	10.03	(28) 95.71	6.83	(22) 97.27	8.47
	(9) 86.22	8.18	(18) 94.88	6.93	(17) 98.41	8.68

the number of cases, the average, and the sigma distribution for the mental ages of the group at the end of the third year.

In Table II it is worth noting that the Control Group gains from about two to about six months of mental age. This is followed by an almost identical standing in School A, and a loss of two months in School Y. At the end of the year, School A has lost almost 8 months from the two later repetitions, while School Y has lost about 4.5 months. This would seem to indicate that the later retests are of negative value in advancing the individual's score beyond the advances made in the first retest.

The Coached Group shows quite a different development. Both the A and the Y groups show a marked increase in the second test and a lowering of this gain in the third test. In the A group there is a gain of about 23.5, and in the Y group of about 32 months. While this is a grave error in its effect on test results, it is even more disturbing to find that a large proportion of this effect is still present three months after the tests. In the A group this effect is about 19 months of increase in the average score, and in the Y it is about 15.5 months. If the retests are given after longer intervals, the effect is lessened. This is probably due not only to the forgetting of the coached material, but also to the fact that the children were growing, so that they were tested on the upper ranges of the material where there had been no coaching. That is, in time they outgrew the coaching. At the end of the first year, the A group retained about 9 months of the gain shown in the first retest, and the Y group kept most of the increase shown at the same retest.

The difference in the amounts retained at A and Y is probably due to the difference in the coaching process. At A the time was spent in more concentrated work on fewer tests; at Y the time was scattered more by the inclusion of two extra years of coached material. This meant that the material at Y could not be remembered so well over a short time, but that it extended farther up the scale and so could effect the results over a longer intervening period of growth. The children at Y could not outgrow their coaching so rapidly as was the case at A.

At the end of three years, this difference between the two groups is still apparent. At A there is scarcely 1 month of mental age shown over and above the expected growth (though it must be remembered that at the same time there is an actual loss in the Control Group of about 2.5 months), while at Y there is an actual gain of a little over 10 months.

The Similar Group shows the same rise in score due to familiarity with the tests, plus the actual learning that is transferred to the material of the test on its next presentation. The increase is not so large as in the case of the Coached Group. In the first retest, the A group showed an increase of about 7.5 months of mental age, and the Y group showed an increase of about 6.5 months. At the end of three months, the A group showed an increased improvement, with a gain of 11.5 months over the first test, while the Y group still showed a gain of 3.5 months. At the end of the year, the A group still showed a gain of 5.5 months, and the Y group showed a gain of 2.5 months. At the end of three years, the A group still showed a gain of 7 months, and the Y group showed a loss of about 2 months. The standard deviations in these last two cases have increased to 17.29 at A and to 16.12 at Y, so that the results must be scanned carefully. It is possible that another period of testing would have given identical results in the two groups.

In conclusion, it seems that the effect of coaching children in the tests of the Stanford Revision of the Binet Scale will have an effect on later tests for an extended period of time. It is not possible to predict with any great accuracy the amount of alteration which will be produced by such coaching. In general, its effect is still felt in no small measure at the end of a year. At the end of three years the issue is somewhat clouded by the large standard deviations found for these groups. There is some effect still present, as can be seen by comparing the results of the Coached Group with those of the Control Group.

It must be remembered that the amount of time given to the coaching of these children is relatively very small. No child was coached for more than two hours. If so little coaching can produce so great an effect, it behooves those who are interested in the exact determination of mental ability to seek always for tests which will not be largely influenced by specialized experiences in the lives of some children. It may well be that even tests thus carefully selected should be supplemented by tests which will gauge the examinee's ability to learn. It may also be that tests will be devised which will call for a universal coaching period as a preliminary, so that there will be some control of the experience of the child in dealing with the type of material used in the final test.

PREFATORY NOTE TO CHAPTER XXII

The studies summarized in this chapter are masters' theses recently completed at Stanford University. They were undertaken as contributions to a program which it is hoped may be continued. The program aims to establish a consecutive, well-organized field of knowledge regarding the limits placed by mental age upon the performance of tasks of varying grades of difficulty. Can items of knowledge or processes of thought not ordinarily mastered by children before attaining a mental age of six or ten or twelve be grasped at a much earlier mental age if they are skillfully imparted to the children? These studies make only a beginning in this field of investigation.

CHAPTER XXII

THREE STUDIES ON THE EFFECT OF TRAINING IN SIMILAR AND IDENTICAL MATERIAL UPON STANFORD-BINET TEST SCORES[1]

MARY L. CASEY, HELEN P. DAVIDSON, AND DORIS I. HARTER
Stanford University, Palo Alto, California

I

STANFORD M. A. THESIS, 1926, BY MARY L. CASEY

Thirteen trained children and 13 control children, all of them in the first grade, served as subjects. The trained and control children were paired by mental age and intelligence quotient upon the basis of Stanford Binet tests administered shortly before the experiment began. Seven in each group were of M.A. 5 to 6 years; and six were of M.A. 8 to 9 years. The I.Q.'s of these sub-sections averaged about 89 and 127, respectively.

The experiment consisted of putting the trained children through 16 half-hour training periods (two a week) for eight weeks. Training was given the pupils in groups of two or three children at a time. The training periods were divided into three parts of 10 minutes, each devoted to different parts of the training program. The training material was similar to the items of the Stanford Binet, and consisted of the following tasks:

1. Count 20-0 by 1's, 2's, 3's, 4's, 5's
2. Healy Formboard
3. Alphabet forwards and backwards
4. Days of week forwards and backwards
5. Months forwards and backwards
6. Learn date
7. Write Roman numerals from 1-30
8. Count piled blocks, as in Army Beta
9. Name colors (20 colors)
10. Tell time
11. Block T puzzle
12. Write capital letters upside down

[1] These studies were carried out under my direction. They have not before been published, but are here summarized for the *Yearbook* by Barbara Stoddard Burks.—L. M. T.

13. Reproduce designs from dots
14. Learn names of 10 new objects
15. Interpret pictures

The Control Group, as its name implies, was given no training, but was retested at the same times that the Trained Group was retested. Retests were given all the children immediately after, and again six weeks after, the end of the training period.

Gains in M.A. and I.Q. may be summarized as follows:

	M.A. 5 to 6 years		M.A. 8 to 9 years	
	Av. M.A. gain	Av. I.Q. gain	Av. M.A. gain	Av. I.Q. gain
Test II over I				
Trained Group	15.71	14.00	12.16	6.17
Control Group	7.29	3.43	6.74	1.67
Difference	8.42	10.57	5.42	4.50
S.D. of diff. (approx.)	10.68	......	4.65	....
Test III over II				
Trained Group	2.72	1.14	6.64	6.5
Control Group	4.0	3.71	6.55	6.5
Difference	−1.28	−2.57	.09	....
S.D. of diff. (approx.)	7.26	......	1.94	....

Interesting clinical data were also reported upon the response of individual children to the training. For example, one child of M.A. 5-2 and C.A. 6-4 could not learn to count from 20 to 0 by 1's, though "he could count from 1 to 20 by mere rote memory. Names and numbers had never been associated in any way." However, the other six children between M.A. 5 and 6 were able to learn this task, though another of them was not able to learn to count from 20 to 0 by 2's.

To ascertain whether or not the gains of the trained children were due to general mental growth or to improvement only in the tasks similar to those of the training material, the Binet tests were all rescored so as to eliminate the items of the scale that were very similar to those of the training material. The following were deleted: digits (forward and backward), naming colors, counting, and giving date. The average *excess* of gain, in months of M.A., of the Trained Group over the Control Group now became:

Test II over I

8.89 for children of M.A. 5 to 6 years (S.D. of gain, 11.48)
2.94 for children of M.A. 8 to 9 years (S.D. of gain, 6.39)

Test III over II

—1.82 for children of M.A. 5 to 6 years (S.D. of gain about 7.0)
2.72 for children of M.A. 8 to 9 years (S.D. of gain about 7.0)

The author suggests that the larger gains found in the Trained Group of M.A. 5 to 6 years than in the Trained Group of M.A. 8 to 9 years are due to the fact that the tests at the 5- to 6-year level are more similar to the training material than those at the 8- to 9-year level.

Conclusions were drawn by Miss Casey as follows:

1. "Training in material similar to the Binet test material results in gain in mental age immediately after the training period.

2. "The gain was greatest for the group having mental ages between 5 and 6 years.

3. "There was a small gain even after a six weeks' period without training, but it can hardly be accounted for as a result of the eleven weeks' training.

4. "Repetition of the tests brings about a gain in mental age. The gains of the Control Group would indicate this.

5. "The training seems to have given the children something in the way of an attitude or interest that enables them to gain in mental age immediately after the training and to retain that gain over a six weeks' period, showing, in addition, a slight gain on the third testing.

6. "Results of the training suggest the importance of early training in accelerating mental growth."

II

STANFORD M.A. THESIS, 1925, BY HELEN P. DAVIDSON

Twenty-six pairs of third-grade children were matched for M.A. and I.Q. (which averaged 96). These constituted a Trained and a Control Group.

The children were tested on the Stanford Binet, and, regardless of their mental level, upon the following tests of 10-year-old difficulty:

a. Drawing designs (as on item X-2 of the Stanford Binet Scale, and also on a battery of 16 designs ranging from simple to complex, devised by the author)

b. Memory span for digits

c. Naming 60 words in three minutes

d. Memory span for letters

Training was then begun and given in half-hour lessons three times a week for six weeks on digits, naming words, and a practice battery of designs (different in content from the test battery). Each type of coaching was given 10 minutes a day to the whole group at once; and each child had a chance to 'recite' about once a week.

The tests listed above (including the Stanford Binet) were immediately repeated, and repeated once more three months later.

1. Results

a. *Designs.* The percentages of children passing the ten-year designs test on the first, second, and third tests were:

Trained Group 28, 76, 60;

Control Group 36, 60, 72.

According to the standard errors of the differences in percent, there are 998 in 1000 chances that the difference in improvement of Trained over Control on Test 2 is significant; 62 in 100 chances that the loss of Trained Group in comparison with Control Group on Test 3 is significant.

b. *Battery of 16 designs.* The Control Group improved on the first retest from an average score of 807 to 867—a gain of 60 points, as contrasted with an S.D. of 157 points on the first test. The S.D. of the gain is 20 points.

The Trained Group improved on first retest from an average score of 787 to 1086—a gain of 299 points as contrasted with an S.D. of 206 on the first test. The S.D. of the gain is 30 points, and the S.D. of the difference (239) in gains (Control and Trained) is 42 points.

The improvement of the average on the second retest over the average on the first test was 899 minus 807, or 92 points, for the Control Group; and 1037 minus 787, or 250 points, for the Trained Group. The difference in gains was thus reduced to 158 points, and represents a retrogression after training stopped. S.D.'s of differences are about the same magnitude here as immediately above.

c. *Digits.* Digit spans on the first, second, and third tests averaged:

Trained Group 5.12, 5.76, 5.80;

Control Group 5.40, 5.48, 5.52.

The S.D. on the first test was .84 for the Control and .91 for the Trained Group. The difference in gains on the first retest was .56, with an S.D. of the difference of .24. The difference in gains on the second retest (over the first test) was .56, with an S.D. of difference of .22.

It is interesting to note that on the first retest in the Control Group, 5 gained, 3 lost, and 17 remained the same; in the Trained Group 14 gained, 2 lost, and 9 remained the same. On the second retest, corresponding figures were 7, 4, 14 for the Control, and 15, 3, 7 for the Trained Group.

d. *Letters.* Transfer of training was investigated with respect to span for letters. On the first, second, and third tests, the letter span averaged:

Trained Group 4.80, 5.28, 5.20;

Control Group 4.88, 5.16, 5.32.

On the first retest the difference in gain was .193, with a standard error of the difference of .187; and there were thus 81 chances in 100 that the gain was significant. On the second retest the difference in gain (over the first test), .04, was in favor of the Control Group.

e. *Naming words.* The percentages of children passing the word-naming test on the first, second, and third tests were:

Trained Group 36, 92, 84;

Control Group 28, 40, 56.

According to the standard errors, the difference in the percent of Trained and Control children passing on the first retest has 998 in 1000 chances of being significant, and on the second retest, 894 in 1000 chances.

On total number of words named, the Control Group improved on the first retest from an average score of 50.96 to an average of 54.27. On the second retest the Control Group averaged 61.4. The Trained Group improved from 55.38 to 78.39, but fell to 76.38 on the second retest. On the first retest, therefore, the difference in gains was 19.69 in favor of the Trained, with a standard error of difference of 4.71. On the second retest, the difference in gains was 10.16 in favor of the Trained Group, with a standard error of difference of 4.97. The S.D. of the Control Group on the first test was 16.21; and of the Trained, 15.3.

f. *Average Intelligence Quotient (25 pairs).*

	Test 1	Test 2	Test 3
Trained Group	95.0	97.8	103.2
Control Group	96.1	101.1	102.3

There is no evidence here of transfer effect upon the test as a whole.

2. General Conclusions

Miss Davidson draws the following conclusions:

1. "Training in material similar to the Binet test material brings about gain in these particular tests immediately after the practice period.

2. "The effects of training were still evident after three months' interval in the case of the digits, and to a lesser degree in the case of the word-naming test. On the other hand, the Control did better than the Trained in the Binet designs." [This latter is true, to an insignificant degree, of the designs of the Stanford Binet Test, but not of the author's battery of 16 designs, upon which the Trained Group was still ahead of the Control after three months.]

3. "There is a very slight indication of transfer of training from the digits to the memory span for letters in the first retest . . . . There is no evidence of transfer of training in the second retest.

4. "The training had no apparent effect on the I.Q. scores immediately after the training period, for the Control gained more than the Trained . . . . After a three months' interval, the gain in I.Q. by the Trained Group is not sufficiently large to justify drawing conclusions either one way or the other."

III

Stanford M.A. Thesis, 1925, by Doris I. Harter

1. Method

Twenty first-grade and second-grade children were divided into the following groups:

Group A. Five trained subjects and five control subjects, matched for I.Q. and C.A., with C.A. ranging from 6-5 to 6-10, and I.Q. ranging from 106 to 146.

Group B. Five trained subjects and five control subjects, matched for I.Q. and C.A., with C.A. ranging from 6-6 to 7-5, and I.Q. ranging from 91 to 127.

The Stanford Binet was administered to all subjects before and at the end of the experiment, which ran over a period of three weeks.

Two lists of words had been drawn up, each list containing 100 pairs of nouns "bearing some relation to each other." No words were included in these lists having frequencies in the Thorndike Word Book lower than the word of lowest frequency in Tests VIII-4 and VII-4 (similarities and differences), of the Stanford Binet Scale.

One list (referred to as the test list) was now given individually to all the children of both groups at the rate of 20 words a day; but Group A children were asked to give similarities, and Group B, differences.

The other list (practice list) was given to the Trained Group the following week (20 words per day) with thorough coaching upon the answers and underlying principles. The test list was then given again to both the Trained and Control Groups the week after.

The test list contained: (a) 12 word pairs which had no similar comparisons in the practice list; (b) 83 words having similar comparisons in the practice list; (c) 5 word pairs which were in the practice list as well.

2. Results

The results are summarized in tabular form below:

Group A. Average Gains in Giving Similarities

	(a) 12 "indirect"	(b) 83 "similar"	(c) 5 "identical"	Total
Trained Group	4.8	47.2	3.2	53.2
Control Group	—0.8	3.8	.2	3.2

(Average scores are not reported; but none of the children 'hit the top' on the retest, except on the 5 "identical" pairs of words; so the comparisons of gains, at least, seem valid.)

Group B. Average Gains in Giving Differences

	(a)	(b)	(c)	Total
Trained Group	1	0.8	—0.2	1.8
Control Group	—0.4	1.0	0.2	0.8

The author suggests that the gains in Group A are due partly to the learning of new words but chiefly to the learning of new methods of attack; and that the failure of Group B to show significant gains may be due to the fact that their average M.A. is lower than in Group A, so that they may be less able to profit from coaching.

Gains in I.Q. During the Experiment

	Trained	Control
Group A	15	4.75
Group B	11.6	11.4

In Group A, using those who passed VII-4 or XII-8 (similarities) on the retest but not on the first test, and calculating I.Q.'s which delete credits for these tests on the retest, the average gain of the trained group is still as high as 10 points.

3. Conclusions

The following conclusions were drawn by Miss Harter:

1. "With children of C.A. 6-5 to 6-10 and I.Q. from 106 to 146 it seems:

a. Possible to impart to them through drill and coaching an attitude or method of approach to the problem of finding similarities between the two members of a comparison which enables them to find similarities between pairs of words on which they have not been coached;

b. That direct coaching on finding similarities between the two members of a comparison is a more efficient method than indirect coaching;

c. That through the drill and coaching of the practice list used in this study, children acquire something which enables them to make a greater gain in I.Q. than those children who do not have this drill and coaching.

2. "With children of C.A. of 6-6 to 7-5 and I.Q. from 91 to 127 it seems:

a. Possible to impart to them through drill and coaching an attitude or method of approach to the problem of finding differences between the two members of a comparison which aids them somewhat in finding differences between pairs of words on which they have not been coached;

b. That children of this mental level do not profit much from drill in finding a difference between the two members of a comparison;

c. That practice and coaching in giving differences between the two members of a comparison have little effect on the ability of the child to make a gain in I.Q., but

d. That sometimes when a child makes a low I.Q., it is possible to increase his I.Q. through familiarity with the tester and with the method of examination."

PREFATORY NOTE TO CHAPTER XXIII

The question raised in this chapter refers to the ultimate nature of the improvement brought about by training. Admitting that practically any continued process of drill or practice will produce an improvement in proficiency, the question still remains: Has the training, the educational effort, really altered the person's intrinsic capacity, either directly or indirectly, or has it only furnished him with devices, methods, ideas which enable him to use his capacity more effectively?

The author of the chapter sketches these and related theoretical possibilities, then shows how an experiment would have to be contrived to investigate the possibilities. Two experiments are reported, the one in the field of motor skill, the other in the field of memory.

From these experiments follow certain implications concerning the probable nature of capacity and the effect on it of processes of training, also certain implications for educational theory and practice. In substance, it is held that "the effect of education is not to change capacity directly nor to modify the growth of capacity, but only to give the recipient useful information, techniques, methods of work, and the like." This conclusion obviously reinforces the position of those who believe that the fundamental endowments given by nature are themselves little open to alteration by nurture.

CHAPTER XXIII

THE NATURE AND LIMIT OF IMPROVEMENT DUE TO TRAINING

Arthur I. Gates
Teachers College, Columbia University, New York City

I. Nature of the Problem

Though many facts have been accumulated, they have been too diverse and seemingly conflicting to provide an unequivocal answer to the question: How much does nature and how much does experience contribute to one's equipment? Out of all the studies, one thing appears certain: The relative potency of these two factors differs with different human traits. The influence of events and circumstances of living have different effects upon the color of the eyes, the length of the skeleton, the preferences for the primary gustatory qualities, the evenness of temper, the speed of running, and the ability to do arithmetic.

Not only it is advisable to narrow the problem in the effort to determine the relative contributions of inherited traits and of experience but it is also preferable to ascertain, if possible, on what factors and in what way each influence operates. In this paper will be discussed the problem of determining in what manner and degree native factors—if such there are—and experience each influence human proficiencies in mental and motor activities.

Whatever one's views may be concerning the relative importance of native 'capacities' and education, it must be admitted that in most mental and motor activities well-managed training does produce changes—changes that we customarily term 'improvement.' This fact has been amply and most objectively demonstrated by the so-called 'practice experiments' in psychology. Thorndike, reviewing the literature prior to 1913, wrote: "So far as I am aware of the facts, no mental function has been deliberately practiced with an eye to improving it and with proper opportunity for the law of effect to operate without some improvement as the result."

Beyond the bare fact that education may and does produce improvement along the line practiced lies the question, of great importance to education and psychology, concerning not only the magnitude but especially concerning the nature of the changes which training brings about. What, precisely, is the character of these changes which constitute improvement in a function? Are these changes wrought in a person's most fundamental fiber and tissue? Are they additions to, or increases in, or betterments of, the substances which the hereditarians term 'innate capacity'? Are the changes due to training identical in kind with modifications due to 'innate maturation' or development? Or are they of a very different sort? Are these changes solely a result of training *per se* or are they conditioned in some important way by the nature of the person on whom education operates? These seem to be very vital questions for the sciences of psychology and education which are concerned with understanding, measuring, and changing human behavior.

So far as the writer knows, few previous investigations have been so designed as to shed much light on these central problems.[1] There are, nevertheless, certain general theories that have been offered. These theories will be mentioned in brief before certain studies designed to contribute to the solution of the crucial problems are presented.

II. Three Possible Theories of the Nature of Improvement

Concerning the character of such improvement as is brought about by practice, there are three types of theories. Since these three types have no traditional designations, each will be suggested in a phrase and later explained in a paragraph. The three theories may be termed:

(1) The theory that improvement results from the *direct* increase in fundamental capacities.

(2) The theory that improvement results from the indirect increase in fundamental capacities by means of stimulated growth.

---

[1] The nearest approaches are the studies of G. M. Whipple, *Journal of Educational Psychology*, 1910, pp. 249-262, W. S. Foster, same journal, 1911, pp. 11-22 and K. M. Dallenbach, same journal, 1914, pp. 321-334, 387-404, on the development of the range of "visual apprehension" of letters, digits, dots, and other figures. They discussed some of the same general issues but utilized experimental techniques which differed in important respects from those used in this present work.

(3) The theory that improvement results from the acquisition of techniques, information, 'tricks of the trade.'

1. Direct Increase of Fundamental Capacities

The first theory, which assumes that practice may bring about an increase or improvement in the fundamental capacities exercised, may take many forms. An old form, known as the 'Faculty Theory,' assumed that memory, perception, retentiveness, motor adaptability, etc., were fundamental powers, capacities or 'faculties' which could be improved as a whole by training in one or more particular lines. Thus, memory as a whole, it was held, could be trained by practice on names and dates, series of numbers, and the like. The old type Faculty Theory is now rarely upheld because of the antagonistic results of numerous experiments on the transfer of training. Modified in application and form, the same theory may be, and still is, maintained. It may—to illustrate—be freely admitted that there are many particular memories instead of one as the Faculty Theory assumed, but that each memory depends upon the operation of various human capacities. It may be argued, furthermore, that practice in memorizing any kind of material brings about improvement in whatever neurones, muscles, or other machinery or 'capacities' are exercised and that these mechanisms, when fundamentally improved, will contribute to increased efficiency in other tasks. This is not the same as saying that specific practice improves memory, in general and for all purposes. It may be that every other type of memory includes mechanisms not involved in the type trained. It need merely to be urged that training does increase certain capacities and that this increase will be apparent only when one engages in another task which requires the function of identical mechanisms. As a matter of fact, the 'transfer of training' demonstrated in experimental studies could reasonably be explained in this way.

Those who uphold this theory, in general, would assume that general intelligence, whether conceived as one or two or innumerable capacities, could be improved by exercise. They would believe, most probably, that intelligent action depends upon the operation of a great many particular mechanisms and capacities and that as these are each trained, singly or in combination, they may function better in any other task in which they are engaged. In this way intelligence is gradually improved by means of diverse types of intellectual training.

2. Indirect Increase by Stimulated Growth

The second type of theory is similar to the first except in regard to the means whereby capacities or mechanisms are improved. In the first theory it is assumed that the effects of training are direct and immediate and that they may appear in adults of full maturity as well as in growing children. The second theory seeks to account for an increase in a capacity by the indirect means of the stimulation of growth. It is first pointed out that general intelligence, speed of learning, rate of motor response and the like develop gradually from birth to a maximum or maturity reached at some age, presumably, usually before twenty. Since they are apparently subject to growth, like bodily tissues, it is conceived that *continued exercise before the time of maturity will stimulate a more rapid and prolonged growth of the capacities than would otherwise occur.* Continued exercise, by the production of nutritive after-effects, or by stimulating the production or release of 'hormones' or in other ways now unknown, might increase the rate of development and raise the mature level of the factors which underlie capacity, whatever they may be.

3. Improvement in Technique Only

According to the third theory, education or training in a function produces improvement by means of the development of techniques, mental and motor, without any change in what are usually termed fundamental capacities. Thus, it is held that in learning to play the piano or to dance improvement results, not from improved motor capacity, not from increased speed, steadiness, facility, etc., of the motor responses in a general way, not from a general improvement of the motor and neural machinery involved in the tasks, but rather from the acquisition of new methods of managing the mechanisms, from new patterns of reacting, new knowledge of procedure, new methods of work, new techniques, new 'tricks of the trade.' Similarly, in learning to read, it is held that improvement consists not in improved memory, perception, intelligence, not in generally increased mental or motor capacity but in new information, new methods of work, new habits of functioning. The many investigators and theorists who at the present time conceive of 'intelligence' as a native capacity or group of capacities which cannot be appreciably changed by education and experience, however rich, must subscribe, substantially, to this type of theory.

III. Possible Criteria of the Source of Improvement

To determine in which of these three conceivable ways improvement is brought about is of greatest importance. Improvement may, of course, be the joint product of all three or of any two factors. If this is a fact, it remains equally important to determine the contribution of each. Is there available evidence which will enable us to decide what the truth is? Let us consider several possible criteria.

1. Improbability

A native capacity is sometimes defined as a factor which cannot be improved. This definition carries with it the suggestion that one should be able to locate certain activities which depend upon pure capacity—so to say—and could not, therefore, be improved by training. The Binet Test, it is claimed, measures intelligence, a native capacity or group of capacities. Actually, the Binet Test determines how well a person does certain activities. If these activities are native capacities, it might be argued, they should not be subject to improvement by training. It has been demonstrated, however, by Dr. Graves and others that training in the Stanford Binet exercises usually brings increased ability. Such a demonstration—as the investigator herself points out—does not answer the question whether improvement is due merely to acquired information and techniques in all cases or to machinery fundamentally changed, nor does it determine whether intellectual capacity of unmodifiable character is a legitimate conception. The Binet Test may not weigh intellectual capacity directly but merely provides, under certain circumstances, a reliable symptom of it. So, in general, merely to prove that any particular function can be improved by practice does not prove that the notion of native capacities is a fiction. Improvement may everywhere be due primarily or wholly to acquired techniques.

On the other hand, if it were possible experimentally to control or eliminate the contribution, if any, due to acquired information and technique, the influence of continued practice directly upon capacities or upon the growth of capacities might be ascertained. If, under these conditions, practice should bring about improvement, it would appear that the fundamental mechanisms or capacities were favorably changed. If these conditions could be fulfilled experimentally, a fairly crucial test of the theories would be possible.

2. Permanence of Improvement

It seems usually to have been assumed by those who think changes in fundamental capacities are brought about directly by training or indirectly by the stimulation of growth, that these changes possess considerable permanency. It would be expected especially that general or specific intellectual capacities whose growth had been accelerated by training would tend to retain the level achieved. To the extent that improvement is due to acquired information and techniques of functioning, a lesser degree of permanence, on the whole, would apparently be expected by most writers. If this distinction were valid, the degree of permanence of improvement would be a criterion by means of which the nature of improvement could be appraised. Resort to the literature, however, sheds little light on the issue from this point of view. For one reason, there is uncertainty owing to the small number of investigations in which the degree of permanence of an intellectual or motor skill has been determined. There are many examples of practice producing improvement in ability to memorize poetry and nonsense syllables, in speed of reading and so on, but rarely has the persistence of such skills during periods of disuse been determined. There are some instances, however, of rather permanent retention of the ability to type, toss balls, trace mirror-wise, solve geometrical problems, and the like.[2] Theoretically, we know that

[2] For example, B. Bourdon, "Récherches sur l'habitude," *L'Année Psych.* 8, 1901-2, p. 327f; E. J. Swift, "Studies in the psychology and physiology of learning," *Amer. Jour. Psych.*, 14, 1903, p. 201f; E. J. Swift, "Memory for a skillful act," *Amer. Jour. Psych.*, 16, 1905, p. 131f; E. J. Swift, "Relearning a skillful act," *Psych. Bulletin*, 7, 1910, p. 17f; Swift and Schuyler, "The learning process," *Psych. Bulletin*, 4, 1907, p. 307f.

information, although it escapes readily, may, when 'over-learned,' be retained for very long periods[8] and, therefore, to the extent that an intellectual or motor power is dependent upon *knowledge* of methods of procedure, it may be retained for a long period. In no studies of which I am aware have knowledge of procedures to use, and improvement of capacities, if any, been sufficiently isolated from the more subtle or less conscious techniques or skills to make possible an appraisal of the permanence of the latter. Most students, I judge, would expect a variation from extremely subtle and evanescent skills to rather stable ones. Since acquired techniques and knowledge may sometimes be well retained, the permanence of improvement is scarcely a valid criterion for the purpose of determining the nature of improvement except possibly in one way: If improvement, brought about by prolonged practice, is relatively temporary, it is more likely due to subtle skills or techniques than to an increase in the fundamental machinery or capacities.

However this may be, a practical question of great moment is whether improvement in attention to, or in learning or retention of, any given data, or any other intellectual ability tends to be unstable and evanescent or stable and permanent. If it is the latter, great expenditures in time and effort in training would be justified, since a small increase in intellectual ability is of tremendous worth; if the former is true, the futility of much of our educational endeavor will be apparent.

3. The Form of the Curve of Improvement

A third criterion which has been applied in the effort to distinguish the sources of improvement is the nature of the curve of improvement. Like the others, this criterion is not absolute and may be stated only in the form of probabilities. In the main, the suppositions are as follows: To the extent that improvement is due to a direct or indirect increase in capacity, the curve of improvement would probably be relatively smooth and of uniform acceleration. Rapid initial improvement, irregularities, and especially sharp ascents here and there suggest rather the products of insights or information otherwise acquired and the effects of good

[8] See C. H. Bean, *The Curve of Forgetting.* Archives of Psych., No. 21, 1912.

and bad techniques due to accidental factors or tuition. It is possible that improvement resulting from increased fundamental capacity should give a convex or concave or irregular curve and that improvement due to acquired information and methods of work would be smooth and of uniform acceleration, but in the light of available knowledge these possibilities are less probable than the reverse.

These three criteria are the main ones that have been suggested as means of determining the nature of improvement. None is alone fully adequate. The investigator should utilize all three but with great caution in appraising his data. Even then he can scarcely expect to achieve an unequivocal interpretation.

IV. An Experimental Technique Designed to Distinguish the Cause of Improvement

The most fruitful approach to a solution of the problem lies in the design of the experimental situation. The truth is most likely to be revealed by an experiment so designed as to keep one or more of the types of development—the acquisition of information and techniques, the direct increase of capacity or the indirect improvement due to the acceleration of growth—under control. In this way, the other factor or factors may be tested.

According to the writer's analysis, a procedure most likely to yield crucial data consists in either eliminating the possibility of improving by means of knowledge of methods of procedure or by the acquisition of techniques or at least in keeping these factors under control and then providing extensive training. Such training should test the possibility of improvement due to increased capacity apart from change due to techniques and methods of functioning. If, with the influence of acquired tricks of the trade eliminated or allowed for, continued practice does bring improvement, it should be attributed to increased capacity directly or indirectly brought about. To distinguish between capacity increased directly and capacity improved indirectly by the acceleration of growth, additional experiments must be performed. A suggestive test would be to apply the same experimental procedure to subjects in a stage of rapid growth and to subjects well beyond the age of maturity. To the extent that improvement is greater in the immature subjects, it is probably due to the indirect effects of stimulation of growth, rather than to direct influences upon the mechanisms which underlie capacity.

The first test should be made to distinguish between acquired techniques of functioning and increased capacity however caused; since, if no evidence of the latter appears, the effort to distinguish between direct and indirect causes will be unnecessary. According to the writer's appraisal, the following conditions should be observed to approximate a crucial test of the first problem.

1. Subjects for training should be secured who have a minimum of ability to improve their performance by insight, by studying their own methods, by observing other performers, by utilizing suggestions of teachers and others. The reason for this condition is that it is desirable to reach as rapidly and fully as possible a stage in which no improvement from superior techniques is possible.

2. The function to be trained should be one in which the possibilities of improvement by means of acquired techniques, knacks, skills, tricks of the trade, are as limited as possible and are likely to be acquired fully in a short time.

3. The function should be one not practiced much in the ordinary life of this subject. This condition is prescribed in order to keep the training under control and to make it possible to provide at will periods of disuse.

4. The subjects for the study should be children at an age of rapid growth, in order to make it possible to secure the fullest effects of practice, if any, upon the growth of capacity.

5. The subjects should be numerous enough to provide two groups equivalent in age, intelligence, and other traits which may contribute to efficiency in the function trained.

6. One of the two equivalent groups should secure a great deal more practice, distributed over a longer period of time, than the other. The crucial matter is the influence of practice continued beyond the time when all or nearly all of the improvement due to working methods or techniques have been secured.

7. The practice should be of the sort known to produce most surely and effectively the greatest improvement.

V. An Experimental Test of the Theories by Means of Practice in Tapping

In two experiments efforts have been made to satisfy these requirements in slightly different ways. In one, the rate of tapping with a blunt pencil was chosen as a function, school children from four to six years of age as subjects and daily practice under the conditions of the 'practice experiment' as the method of training.

Eighty-two children were given daily three short practice periods in tapping for eighteen days. They were also tested once in eight different motor functions requiring speed and precision of control. These tests were: (1) marking a page of oblique lines //// to make each an X; (2) drawing X's as rapidly and as well as possible; (3) making equal signs (=); (4) striking the center of a series of circles (target test); (5) copying a series of simple geometrical figures; (6) sorting a box of pegs according to color; (7) marking between converging line (steadiness test), and (8) tracing a maze pathway (speed and steadiness test).

At the end of the eighteen days of practice in tapping, the pupils were paired so as to yield two groups practically equivalent in each of the following respects: (1) speed of tapping; (2) motor ability in the eight tests just enumerated; (3) sex, *i.e.*, equal number of each sex in each group; (4) chronological age; (5) Stanford Binet mental age; (6) intelligence quotient; and (7) grade in school.

Both groups practiced eighteen days during which improvement was very rapid until toward the end when it appeared that the pupils had reached about the limit of improvement by means of perfecting a technique of tapping. For one group the training was then stopped but for the other it was continued. The pupils were given three short practice periods per day while working in groups of five to eight, for 76 school days, over a period of six months. Then, during a final period of 17 days both groups were given practice. The entire schedule may be summarized as follows:

	Preliminary Period	Main Period	Final Period	Total
Practice Group	18 days	76 days	17 days	111 days
Control Group	18 days	0 days	17 days	35 days

The progress in tapping during this period is shown in the table below:

TABLE I.—AVERAGE SCORES IN TAPPING, ARRANGED BY FIVE-DAY PERIODS (EXCEPT THE STARRED, WHICH ARE FOUR-DAY PERIODS)

Days	Group		Days	Group		Days	Group	
	Practice	Control		Practice	Control		Practice	Control
1– 4*	35.5	38.8	39–43	40.6	....	79–83	43.7	....
5– 9	38.4	37.5	44–48	41.2	....	84–88	44.3	....
10–13*	40.6	40.1	49–53	41.8	....	89–93	44.6	....
14–18	40.1	40.9	54–58	41.0	....	94–98	44.7	42.3
19–23	39.8	....	59–63	42.1	....	99–102*	44.2	42.2
24–28	39.9	....	64–68	42.8	....	103–106*	45.0	45.2
29–33	40.1	....	69–73	43.2	....	107–111*	44.8	44.3
34–38	40.7	....	74–78	41.9	....			

These figures indicate a fairly rapid rise in tapping ability during the first 18 days, suggestive of the acquisition of various devices or techniques. For the Practice Group there is a slow but steady improvement during the next six months of practice. Tested again after this period of no practice, the untrained, or Control Group, is slightly less proficient than the Practice Group. The Control Group very rapidly improves, however, so that after the 103rd day they are quite equal to the Practice Group. These facts suggest the probability that the Control Group needed at the end only to perfect the method of working, which had suffered from six months of disuse, to equal the highly trained pupils. Both groups were better at the end than at the beginning, probably because they were more mature, but there was no clear evidence that the long period of intensive training had itself increased the capacity of those so trained.

The test of permanence should also be applied to these pupils for several reasons: (1) it is possible that the influence of intensive training on growth is rather slow but that, once started, it may continue; (2) it is possible that, although the trained pupils were no better than the untrained at the end of the practice period, they had an improved capacity that would not deteriorate rapidly, whereas the others might lose their easily acquired skills quickly; and (3) it is possible that the continuous training had resulted in a loss of zest which affected the final performances of the Practice

Group. For these reasons, both groups were given no further practice for a period of six months (which included the summer vacation). At the termination of this period both groups were again given three tests daily for six consecutive school days. The scores showed both groups to be practically equivalent in tapping ability at the end of this period. There was no evidence to indicate that the extensive training of the Practice Group had brought about any greater or more permanent improvement than the less extensive experience of the Control Group.

The final test applied to the two groups was designed to ascertain whether the highly trained group, even if no better in tapping, might not have improved for other, related types of motor capacity. It is not altogether improbable that they had become rather bored with tapping—although such boredom was not manifest in any unusual degree—while their capacities for motor activities were increased and might be apparent in slightly different functions. The eight motor tests enumerated above and given at the beginning of the study were repeated at the end of the six months of practice and again after the six months' period of rest. In these tests, the two groups made substantially the same average improvement.

This study has revealed the following facts or impressions:

1. The curve of improvement, which shows an early rapid rise (at two points for the Control Group) suggests the probability that the increase in ability is due to improved techniques of taking the tapping tests rather than to a steady increase in any fundamental capacity for tapping.

2. The fact that the Practice Group was no better, after nearly six months of continuous practice, than the Control Group after a few days of readjustment also suggests that improvement was due to technique and that even prolonged training had no perceptible influence upon capacity.

3. The fact that, after six months of disuse, the intensively trained Practice Group showed no superiority over the Control Group points to the same conclusion.

4. The achievements of the two groups tested three times during the year in other motor skills lends no support to a belief that any capacities, specific or general, had been increased by intensive training.

VI. A Second Experiment: Immediate Memory

In the first study just reviewed are certain features which conceal several facts of interest in connection with the general problem. In the first place, the Control Group, utilized to eliminate the probable effects of acquired techniques, received considerable training. Strictly speaking, the study revealed the apparent futility of training beyond a certain amount to produce an increase in ability. A similar study in which the Control Group received a much smaller amount of practice should illuminate certain other points. In particular, the first study made it impossible to discover how permanent the influence of acquired techniques may have been, since both groups had apparently secured a maximal influence, and, consequently, there remained no basis of comparison. Secondly, the investigation utilized a function in which motor activities were engaged. Certain critics have stated that the failure of long continued practice may have been due to a physiological limit of motor response and that a more purely mental function might behave differently. While the writer would reply to these criticisms that the essence of the problem of increasing capacity is the question of increasing the so-called physiological limit and that the alleged distinctions between 'motor' and 'mental' functions are mainly fictitious, he nevertheless framed a second study to test the assumptions raised.

The second study was conducted with subjects, practice distributions, incentives, and general technique similar to those in the first, but with certain differences. The function selected was immediate memory for series of digits presented orally. This function, like tapping, was judged to be little subject to practice in the ordinary course of events by young children, to allow of only limited acquisitions of techniques, and to be what is usually termed a nearly 'pure mental' function, depending little on motor skills. This function is used in the Binet Test as one of the symptoms of intellectual capacity. Finally, the Control Group in this study was given only a careful test at the beginning, without several days of practice as in the tapping study.

As in the preceding study, two groups were selected approximately equivalent in traits that might have an influence on improvement. The traits taken into account were: (1) sex, (2) age,

(3) Stanford Binet mental age, (4) intelligence quotient, (5) scholastic maturity as judged by teachers, (6) school grade, (7) memory for digits presented orally—the function to be trained, (8) memory for letters presented orally, (9) memory for unrelated words, (10) memory for related words presented orally, (11) memory for geometrical figures, (12) memory for pictures of common objects, and (13) memory for names of pictures, *i.e.*, printed words accompanying pictures.

The pupils of the Practice Group were given, individually, practice in recalling series of digits which varied in length from an easy series to one beyond their ability on each of 78 days extending over a period of five months. The Control Group was tested only on the first and last days, by the same examiner and by the same technique used with the Practice Group. The results are shown in Table II.

TABLE II.—SCORES IN IMMEDIATE MEMORY FOR DIGITS

	Practice Group	Control Group
Average score, initial test	4.33	4.33
Average score, final test	6.40	5.06
Gain	2.07	.73

The Practice Group progressed steadily (as shown by daily records not given here) from 4.33 digits to 6.40, a gain of 2.07 digits. On the Stanford Binet Scale this advance requires the hypothetical untrained average child approximately 6 years.[4] The Control Group, equal at the beginning, is one and a third digits lower on the scale at the end. Practice demonstrably brings improvement—big improvement. What is its nature?

First, apply the test of permanence. After the final tests, all practice had been dispensed with for four and a half months (a period which included the summer vacation) when, on October 10, the tests were again given to all of the original pairs of pupils available. In this October test the scores of the two groups were almost identical; the slight (but wholly unreliable) advantage was with the Control Group. The two groups were, then, as nearly equivalent as they were at the beginning of the experiment. Im-

[4] Memory for four digits is placed at Year IV and for six digits at Year X.

provement, whatever its nature, is apparently unstable and evanescent in this function and in these subjects. Such instability suggests the probability that the improvement brought about by training in this case is due to subtle techniques rather than to increased fundamental capacities.

If the improvement in ability to recall digits is due to highly subtle techniques it is quite probable that these techniques consist partly in subtle adjustments to the particular test situation in which the pupils were trained. If this is the case, it may be expected that minor changes in the test situation would affect the achievements of the highly trained pupils. Factors in the test situation are the voice, lip movements, facial expressions, etc., of the examiner. All the training was conducted by the same examiner, who used the Stanford Binet technique of testing memory for digits throughout. To have the pupils tested by another person, highly skilled in the same technique, would constitute a test of this hypothesis. The tests after the five months period of disuse were, in fact, given by a new examiner. The fact that the Practice Group showed no superiority may have been due in part to the change in the test situation. The suggestion is that the original improvement of this group was due, not to any fundamental increase in capacity for memory for digits, but to very specific methods of adjustment to a particular test situation.

Further evidence in favor of the last supposition is afforded by the results of a repetition of the tests for recall of letters, unrelated words, pictures, geometrical designs, and picture names. In the tests of these abilities given approximately ten months after the beginning of its training in memory for digits by a new, but skilled examiner, the Practice Group showed no superiority to the Control Group. In no phase of immediate memory work did the extensive training seem to have been of value at this time.

In the tests after the interval of disuse, the examinations for recall of digits were, however, not very extensive. It is conceivable that the Practice Group needed but a little reorientation to demonstrate a superiority. It is possible that the effects of the extensive practice were lying dormant, needing only a short period of further practice to arouse their potency. It is conceivable that the Practice Group would rapidly achieve through further practice an ability to recall digits that the Control Group could not equal. To test this possibility, both groups, 13.5 months after the beginning of the study and 8 months after the end of the training of the Practice Group, were given specially intensive and equal practice for 22 consecutive school days. The average scores were as seen in Table III.

TABLE III.—SCORES IN IMMEDIATE MEMORY FOR DIGITS IN SPECIAL PRACTICE PERIOD

	Score First Day	Score Twenty-second Day	Gain
Practice Group..............	4.73	5.73	1.00
Control Group..............	4.83	5.92	1.09

The Practice Group failed to reveal in this test any fundamental advantage from its previous intensive training.

In sum, all the evidence from both studies seems to favor the theory that improvement brought about by training is due to acquired information and skills, to techniques or methods of attack, to tricks of the task. No evidence has been found, on the other hand, to support the theories that continued practice increases any fundamental mechanisms or capacities underlying a function either in some direct manner or indirectly by means of accelerating the growth of whatever mechanisms or capacities are responsible for the abilities trained.

VII. Objections to These Experiments

Before turning to some of the implications of these studies, it will be advisable to consider certain objections which have been voiced since these studies were originally published.

It has been objected that the functions selected for training are unusually narrow and therefore not typical. This fact is admitted; indeed, such functions were deliberately selected—as already explained—to make possible a crucial experimental situation in which the effectiveness of acquired techniques could be controlled.

It has been objected that because the techniques here found were so subtle, they were less permanent than those found in representative functions. This is also probably true. It was the condition sought. The transitoriness of the skills made it possible, by letting them die from disuse, to ascertain if any residual improvement due to other causes remained.

It has been stated that, whereas little or no transfer of improvement to similar functions was found in these studies, some transfer is usually found and that, consequently, the results of these studies are atypical. All of these statements are also probably true. The writer has argued elsewhere[5] that transfer of improvement is due, not to the increase of 'identical' capacities or neural or other mechanisms, but to the functioning of acquired techniques, skills, devices, and information in situations other than the one in which they were acquired. Some identical or similar feature of a new situation may activate one or more or a combination of these acquired facts or skills. In the present study, where the acquired techniques are exceedingly subtle, evanescent or uniformly acquired by both groups, transfer of this form would not be expected. That it did not occur in these studies indicates both that transfer is probably always due to the spread of information and skills and never to the improvement of fundamental capacities utilized in different functions. Again, the choice of functions in which the techniques were subtle, narrow, and instable made possible insight into the nature of improvement that would otherwise have been obscured.

While there are grounds for believing, then, that experimental techniques of the sort here described may yield crucial data concerning the nature of improvement, there is no disposition to claim that these particular results are conclusive. Too many different influences unobserved by an investigator may affect the outcome of such studies. Repetition of similar researches by other students are needed. Meanwhile, it is merely suggested that these results increase the probability that education gives us information and techniques but does not increase fundamental capacities.

VIII. Implications of the Results Concerning the Concept of Capacity

The results of the study—assuming for the moment that they are valid—yield certain implications concerning the validity of the

[5] *Psychology for Students of Education,* Ch. XV.

concepts 'capacity' and 'proficiency.' The relevant facts are as follows. In the tapping study is was found—though these data were not given above—that individuals differed in tapping ability not only before practice was begun but after a minimal efficiency had been attained by dint of intensive and long continued training. It was found, moreover, that after a certain period, further practice was impotent to increase the rate of tapping among those intensively trained. The Control Group, though unpracticed during a long period, nevertheless showed an increase in ability. This increase was apparently due to growth—to development in no way attributable to intensive training—since it was identical for both the trained and untrained groups. As the result of growth alone, individuals differed in ability. In the memory for digits experiment it was found that ability could be increased by special training but that the increase due to training alone was not permanent. Increased ability also resulted from growth among those not trained and such ability possessed a high degree of permanence. In this study, also, individuals differed in ability due to growth and this growth was apparently not subject to modification by intensive training, since in the long run the trained and untrained groups were equivalent. Though we shall necessarily await further information concerning human mechanisms before we can state precisely what it is that grows and grows differently in different individuals so as to result in various degrees of ability in performance, we may nevertheless term these factors *capacities* and distinguish them from the acquired information and techniques which, though they also contribute to proficiency, nevertheless show clear differences in character when they may be distinguished as in the present studies. By *capacity* we may mean, then, the functional possibilities of the neural or other mechanisms which result in a degree of ability without highly special or intensive practice; factors which, as shown by these experiments, grow slowly but steadily both in those who are, and in those who are not, subjected to special training. A person's *proficiency* normally depends both upon his capacity, as defined, due to inner growth, and upon techniques, methods of work, information, adjustments to the task, etc., which are the results of practice and experience.

If the results of the studies reported are correctly interpreted, they imply that all types of mental and motor abilities or proficiencies are partly to be attributed to capacity and partly to information and skill resulting from education. It is not maintained, of course, that the contributions of training will be of the same amount in all functions as they were here. It is merely stated that the distinction in kind between capacity and acquired factors will be a true, and therefore, probably a useful one.

IX. Implications of the Facts for Education

The implication of the studies is that a person's ability in any intellectual task depends upon his native capacity and upon the acquired information and skills which may influence his performance. A further implication is that the effect of education is not to change directly or to modify the growth of capacity, but only to give the subject useful information, techniques, methods of work, and the like. To certain tender-minded persons, this implication will be distasteful, partly because the suggestion that education is subject to any limitations whatever is annoying, and partly because any suggestion that education is anything but easy and simple is also disturbing.

The implications of the studies are certainly to the effect that education is highly complex and that, to be effective, it will require both extensive information and high skill. Since there is no evidence that even good training will produce stable improvement by the easy method of increasing one's fundamental capacities once and for all, the implication is that merely requiring a pupil to submit to training will not be a most effective procedure. On the contrary, the factors that really add to proficiency are facts, skills, techniques, methods of procedure, tricks of the task. These are not only specific to each task and often highly subtle, obscure, detailed, evanescent, but are also, probably, such that the optimal ones vary with the capacity of the individual who tries to acquire them. If these implications are correct, education cannot be adequately achieved by merely asking or requiring children to practice nor by any simple rules of thumb in teaching. Real education will require not only much more intimate and detailed knowledge of all the facts and techniques which may contribute to proficiency in each line, but also much more comprehensive diagnoses and understanding of individual learners than are now common or possible. The very fact that we cannot increase native capacities, but must develop to the full the information and skills needed by each type of task and individual, implies that education must become an intricate art, which must be grounded in a complex science.

INFORMATION CONCERNING THE NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

1. **Purpose.** The purpose of the National Society is to promote the investigation and discussion of educational questions. To this end it holds an annual meeting and publishes a series of yearbooks.

2. **Eligibility to Membership.** Any person who is interested in receiving its publications may become a member by sending to the Secretary-Treasurer information concerning name, address, and class of membership desired (see Item 4) and a check for \$3.00 or \$2.50 (see Items 4 and 5). Membership may not be had by libraries or by institutions.

3. **Period of Membership.** Applicants for membership may not date their entrance back of the current calendar year, and all memberships terminate automatically on December 31st, unless the dues for the ensuing year are paid as indicated in Item 6. Unless special arrangements are made, applications received between January 1st and November 30th are considered as applications for membership during the current year, but those received during December are considered as applications for the year beginning the following January.

4. **Classes of Members.** Application may be made for either active or associate membership. Active members pay two dollars dues annually, receive two copies of each publication, are entitled to vote, to participate in discussion, and (under certain conditions) to hold office. Associate members pay dues of \$1.50 annually, receive one copy of each publication, may attend the meetings of the Society, but may not vote, hold office or participate in discussion. The names of active members only are printed in the yearbook. There were in 1927 about 1250 active and 1300 associate members.

5. **Entrance Fee.** New active and new associate members are required the first year to pay, in addition to the dues, an entrance fee of one dollar.

6. **Payment of Dues.** Statements of dues are rendered in October or November for the following calendar year. By vote of the Society at the 1919 meeting, "any member so notified whose dues remain unpaid on January 1st, thereby loses his membership and can be reinstated only by paying the entrance fee of one dollar required of new members."

School warrants and vouchers from institutions must be accompanied by definite information concerning the name and address and class of membership of the person for whom membership fee is being paid.

7. **Distribution of Yearbooks to Members.** The yearbooks, ready prior to each February meeting, will be mailed from the office of the publishers, only to members whose dues for that year have been paid. Members who desire yearbooks prior to the current year must purchase them directly from the publishers (see Item 8).

8. **Commercial Sales.** The distribution of all yearbooks prior to the

current year, and also those of the current year not regularly mailed to members in exchange for their dues, is in the hands of the publishers, not of the secretary. For such commercial sales, communicate directly with the Public School Publishing Company, Bloomington, Illinois, who will gladly send a price list covering all the publications of this Society and of its predecessor, the National Herbart Society. Bound sets can be supplied of the first twenty-six Yearbooks.

9. **Yearbooks.** The yearbooks are issued in parts (usually two) from one to four months before the February meeting. They comprise from 500 to 800 pages annually. Unusual effort has been made to make them, on the one hand, of immediate practical value, and on the other hand, representative of sound scholarship and scientific investigation. Many of them are the fruit of coöperative work by committees of the Society.

10. **Meetings.** The annual meetings, at which the yearbooks are discussed, are held in February at the same time and place as the meeting of the Department of Superintendence of the National Education Association.

Applications for membership will be handled promptly at any time on receipt of name and address, together with check for the appropriate amount ($3.00 for new active membership, $2.50 for new associate membership). New members will receive, about three weeks after their application is mailed, the yearbook slated for discussion at the February meeting of the year in which the application is made (except in the case of December applications, as explained in Item 3).

GUY M. WHIPPLE, Secretary-Treasurer.

10 Putnam Street,
Danvers, Mass.

PUBLICATIONS OF THE NATIONAL HERBART SOCIETY

(Now the National Society for the Study of Education)

	Postpaid
First Yearbook, 1895.—75 cents; postpaid	$0.79
First Supp. to First Yearbook.—25 cents; postpaid	.28
Second Supp. to First Yearbook.—25 cents; postpaid	.27
Second Yearbook, 1896.—75 cents; postage extra (weight 10 oz.)	.85
Supp. to Second Yearbook.—25 cents; postpaid	.27
Third Yearbook, 1897.—75 cents; postage extra (weight 10 oz.)	.85
NOTE.—Ethical Principles Underlying Education, John Dewey. Reprinted from Third Yearbook, 25 cents; postpaid, 27 cents.	
Supp. to Third Yearbook.—25 cents; postpaid	.27
Fourth Yearbook, 1898.—75 cents; postpaid	.79
Supp. to Fourth Yearbook.—25 cents; postpaid	.28
Fifth Yearbook, 1899.—75 cents; postpaid	.79
Supp. to Fifth Yearbook.—50 cents; postage extra (weight 10 oz.)	

PUBLICATIONS OF THE NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

	Postpaid
First Yearbook, 1902, Part I.—Teaching of History. Lucy M. Salmon. 50 cts.	$0.54
First Yearbook, 1902, Part II.—Geography. W. M. Davis. 50 cents	.53
Second Yearbook, 1903, Part I.—The Course of Study in History in the Common School. Rice, McMurry, Lawrence, and Page. 50 cents	.53
Second Yearbook, 1903, Part II.—Relation of Theory to Practice in Education by M. J. Holmes, J. A. Keith, and Levi Seeley. 50 cents	.53
Third Yearbook, 1904, Part I.—The Relation of Theory to Practice in the Education of Teachers. John Dewey, Sarah C. Brooks, F. M. McMurry, et al. 50 cts.	.53
Third Yearbook, 1904, Part II.—Nature Study. W. S. Jackman. 75 cts. net; (wt. 10 oz.)	.85
Fourth Yearbook, 1905, Part I.—The Education and Training of Secondary Teachers. E. C. Elliott, E. G. Dexter, M. J. Holmes, et al. 75 cts. net; (wt. 10 oz.)	.85
Fourth Yearbook, 1905, Part II.—Vocational Subjects in the High-School Curriculum. J. S. Brown, G. B. Morrison, and Ellen H. Richards. 50 cts	.53
Fifth Yearbook, 1906, Part I.—On the Teaching of English. G. P. Brown	.53
Fifth Yearbook, 1906, Part II.—Certification of Teachers. E. P. Cubberley. 60 cts.	.64
Sixth Yearbook, 1907, Part I.—Vocational Studies for College Entrance. C. A. Herrick, H. W. Holmes, T. De Laguna, V. Prettyman, and W. J. S. Bryan. 60 cts.	.70
Sixth Yearbook, 1907, Part II.—The Kindergarten and Its Relation to Elementary Education. Ada Van Stone Harris, E. A. Kirkpatrick, Maria Kraus-Boelté, Patty S. Hill, Harriette M. Mills and Nina C. Vandewalker. 60 cts. net;	.70
Seventh Yearbook, 1908, Part I.—The Relation of Superintendents and Principals to Training and Improvement of Teachers. C. D. Lowry. 75 cents	.78
Seventh Yearbook, 1908, Part II.—The Co-ordination of the Kindergarten and the Elementary School. B. J. Gregory, Jennie B. Merrill, Bertha Payne, Margaret Giddings. 75 cents	.78
Eighth Yearbook, 1909, Part I and II.—Education with Reference to Sex. C. R. Henderson. Each part, 75 cents; both parts $1.50, postage extra (weight 12 oz.)	1.60
Ninth Yearbook, 1910, Part I.—Health and Education. T. D. Wood. 75 cts. net; (wt. 10 oz.)	.85
Ninth Yearbook, 1910, Part II.—The Nurse in Education. T. D. Wood, et al. 75 cts.	.78

Postpaid

Tenth Yearbook, 1911, Part I.—The City School as a Community Center. H. C. Leipziger, Sarah E. Hyre, R. D. Warden, C. Ward Crampton, E. W. Stitt, E. J. Ward, Mrs. E. C. Grice, C. A. Perry. 75 cents .... .78

Tenth Yearbook, 1911, Part II.—The Rural School as a Community Center. B. H. Crocheron, Jessie Field, F. W. Howe, E. C. Bishop, A. B. Graham, O. J. Kern, M. T. Scudder, B. M. Davis. 75 cents .... .79

Eleventh Yearbook, 1912, Part I.—Industrial Education. F. M. Leavitt, G. A. Mirick, M. W. Murray, J. F. Barker, H. B. Wilson, C. F. Perry, A. L. Safford, P. Johnston, M. Bloomfield, B. W. Johnson. 75 cents net; (weight 10 oz.) .... .85

Price for Yearbooks VI to X inclusive, 8vo. cloth, $5.00; (net weight, 4 lbs.)

Eleventh Yearbook, 1912, Part II.—Agricultural Education in Secondary Schools. A. C. Monahan, R. W. Stimson, D. J. Crosby, W. H. French, H. F. Button, F. R. Crane, W. R. Hart, G. F. Warren. 75 cents net; (weight 10 oz.) .. .85

Twelfth Yearbook, 1913, Part I.—The Supervision of City Schools. J. F. Bobbitt, J. W. Hall, J. D. Wolcott. 75 cents net; (weight 10 oz.) .... .85

Twelfth Yearbook, 1913, Part II.—The Supervision of Rural Schools. A. C. Monahan, L. J. Hanifan, J. E. Warren, Wallace Lund, U. J. Hoffman, A. S. Cook, E. M. Rapp, Jackson Davis, J. D. Wolcott. 75 cents net; (wt. 10. oz.) .85

Thirteenth Yearbook, 1914, Part I.—Some Aspects of High-School Instruction and Administration. H. C. Morrison, E. R. Breslich, W. A. Jessup, and L. D. Coffman. 75 cents net; (wt. 10 oz.) .... .85

Thirteenth Yearbook, 1914, Part II.—Plans for Organizing School Surveys, H. L. Smith; Summary of Typical Surveys, Charles H. Judd. 75 cents net..... .79

Fourteenth Yearbook, 1915, Part I.—Minimum Essentials in Elementary-School Subjects. H. B. Wilson, H. W. Holmes, F. E. Thompson, R. G. Jones, S. A. Courtis, W. S. Grey, F. N. Freeman, A. C. Pryor, J. F. Hosic, W. A. Jessup, W. C. Bagley. 75 cents net; (weight 12 oz.) .... .85

Fourteenth Yearbook, 1915, Part II.—Methods for Measuring Teachers' Efficiency. Arthur Clifton Boyce. 75 cents .... .79

Fifteenth Yearbook, 1916, Part I.—Standards and Tests for the Measurement of Efficiency of Schools and School Systems. G. D. Strayer, Bird T. Baldwin, B. R. Buckingham, M. R. Trabue, F. W. Ballou, D. C. Bliss, H. G. Childs, S. A. Courtis, E. P. Cubberley, Charles H. Judd, George Melcher, E. E. Oberholtzer, J. B. Sears, Daniel Starch, and G. M. Whipple. 75 cents net; (weight 12 oz.) .... .85

Fifteenth Yearbook, 1916, Part II.—Relationship between Persistence in School and Home Conditions. Charles Elmer Holley. 75 cts. net; (weight 10 oz.) .87

Fifteenth Yearbook, 1916, Part III.—The Junior High School. Aubrey Augustus Douglass. 75 cents net; (weight 12 oz.) .... .85

Sixteenth Yearbook, 1917, Part I.—Second Report Committee on Minimal Essentials in Elementary-School Subjects. H. B. Wilson, W. S. Gray, C. F. Munson, J. H. Hoskinson, F. N. Freeman, H. C. Pryor, W. W. Charters, W. S. Monroe, G. M. Wilson, W. C. Bagley, Ernest Horn. 90 cents net; (weight 14 oz.) 1.00

Sixteenth Yearbook, 1917, Part II.—Efficiency of College Students as Conditioned by Age and by Size of High School. B. F. Pittenger. 75 cents net...... .85

Seventeenth Yearbook, 1918, Part I.—Section 1, Third Report of the Committee on Economy of Time in Education—*Arithmetic, Banking, Geography, Vocabulary, Composition, English, Civics, History.* Section 2, Purposes of Historical Instruction in Seventh and Eighth Grades. W. C. Bagley, J. Q. Dealey, C. A. Ellwood, E. B. Greene, A. B. Hart, W. H. Mace, D. Snedden, and others. 75 cents net .... .85

Seventeenth Yearbook, 1918, Part II.—Measurements of Educational Products—A Report to the National Association of Directors of Educational Research S. A. Courtis, B. R. Buckingham, George W. Melcher, F. W. Ballou, W. S. Monroe, Executive Committee. Price 90 cents, (postage extra) 1.00

Eighteenth Yearbook, 1919, Part I.—Professional Preparation of High School Teachers— H. L. Miller, T. W. Gosling, S. S. Colvin, L. V. Koos, A. R. Mead, G. N. Cade, F. C. Whitcomb, C. Fordyce, H. H. Foster, and Bibliography. $1.50 net; (weight 22 oz.) 1.65

Eighteenth Yearbook, 1919, Part II.—Report on Economy of Time in Learning (Fourth Report of Committee on Economy of Time in Education)—H. B. Wilson, F. N. Freeman, W. S. Gray, Ernest Horn, W. S. Monroe, F. C. Ayer, C. E. Seashore. $1.00 net; (weight 12 oz.) 1.10

Postpaid

Nineteenth Yearbook, 1920, Part I.—New Materials of Instruction—Prepared by the Society's Committee on the Materials of Instruction. $1.00 net; (weight 15 oz.) 1.10

Nineteenth Yearbook, 1920, Part II.—Classroom Problems in the Education of Gifted Children, by T. S. Henry. 90 cents net; (weight 12 oz.) 1.00

Twentieth Yearbook, 1921, Part I.—New Materials of Instruction—Second Report by the Society's Committee. $1.20 net 1.30

Twentieth Yearbook, 1921, Part II.—Factors Affecting Results in Silent Reading, and Exercises for Making Reading Function; W. W. Theisen; May Ayres Burgess; Wm. S. Gray; J. A. O'Brien; J. H. Hoover; G. A. Yoakum; C. E. Germane; H. A. Greene; J. L. Packer; Daniel Starch; R. R. Heller and S. A. Courtis. $1.00 net; (weight 15 oz.) 1.10

Twenty-First Yearbook, 1922, Part I.—General Principles of Intelligence Testing; E. L. Thorndike, S. S. Colvin, Harold Rugg, G. M. Whipple. Part II.—Administrative Use of Intelligence Tests: H. W. Holmes, W. K. Layton, Helen Davis, Agnes L. Rogers, Rudolph Pintner, M. R. Trabue, W. S. Miller, Bessie Lee Gambrill, and others. The two parts are bound together in one volume of 300 pages. Price $1.50 net 1.60

Twenty-Second Yearbook, 1923, Part I.—English Composition—Aims, Methods, and Measurement; Earl Hudelson. Price $1.00 net 1.10

Twenty-Second Yearbook, 1923, Part II.—Social Studies; H. O. Rugg, C. H. Judd, L. C. Marshall, E. Rugg, C. W. Washburne, H. C. Hill, R. W. Hatch, A. S. Barr, E. Horn, J. J. Coss, F. M. McMurry and others. Price $1.40 net 1.50

Twenty-Third Yearbook 1924, Part I.—The Education of Gifted Children. Report of the Society's Committee, Guy M. Whipple, Chairman. Price (443 pages), by mail 1.75

Twenty-Third Yearbook, 1924, Part II.—Vocational Guidance; and Vocational Education for the Industries. By A. H. Edgerton and others. Price (435 pages), by mail 1.75

Twenty-Fourth Yearbook, 1925, Part I.—Report of the National Committee on Reading. W. S. Gray, Chairman; F. W. Ballou, Rose L. Hardy, Ernest Horn, Frances Jenkins, S. A. Leonard, Estaline Wilson, Laura Zirbes. Price $1.40 net 1.50

Twenty-Fourth Yearbook, 1925, Part II.—Adapting the Schools to Individual Differences. Report of the Society's Committee. Carleton W. Washburne, Chairman. Price $1.40 . . . . . . 1.50

Twenty-Fifth Yearbook, 1926, Part I.—Present Status of Safety Education. Report of Society's Committee, Guy M. Whipple, Chairman. Price $1.60 net. 1.75

Twenty-Fifth Yearbook, 1926, Part II.—Extra-Curricular Activities. Report of Society's Committee, Leonard V. Koos, Chairman. Price $1.40 net 1.50

Twenty-Sixth Yearbook, 1927, Part I.—Curriculum-Making; Past and Present. Report of the Society's Committee, Harold O. Rugg, Chairman 1.75

Twenty-Sixth Yearbook, 1927, Part II.—The Foundations of Curriculum-Making. Prepared by individual members of the committee, Harold O. Rugg, Chairman 1.50

Twenty-Seventh Yearbook, 1928, Part I.—Nature and Nurture; Their Influence upon Intelligence. Prepared by the Society's Committee, Lewis M. Terman, Chairman 1.75

Twenty-Seventh Yearbook, 1928, Part II.—Nature and Nurture: Their Influence upon Achievement. Prepared by the Society's Committee, Lewis M. Terman, Chairman 1.75