EVALUATING THE EFFECTS OF INTERVENTION IN THE MILWAUKEE STUDY[1]

Howard Garber[2] and Rick Heber, University of Wisconsin

The research design of the Milwaukee study stands in contrast to previous longitudinal studies. Previous longitudinal studies have been either 1) descriptive, e.g., Fels and Berkeley Growth Studies; or 2) treatment oriented, e.g., Klaus and Grey, Weikart. The longitudinal studies in the first category were thorough descriptions, producing a large amount of correlational data, but were without a particular focus, perhaps because the population was not carefully selected according to set criteria. By considering a large number of variables simultaneously, this research was essential in establishing grounds and guidelines for later work.

The second group of studies was limited longitudinally, for their onset was not at the child's birth, i.e. they studied selected groups of children for several years or less. These studies have come under severe criticism because of their lack of adequate control. For the most part, the major selection criteria were low income of the families and age of the child, while maternal intelligence and a host of other important variables were not considered. Clearly, the focus of this second group of studies was remedial, not preventive. Often the treatment was short-term both in hours per day and in total duration. Specific program goals were sometimes lacking.

While the previous longitudinal research has provided segmental evidence of the importance of early development, it has never clearly coordinated the selection and the longitudinal aspects in such a way as to clearly evaluate development of a particular group of children as a function of a prescribed treatment.

The Milwaukee project was designed to determine whether "cultural-familial" or "socio-cultural" mental retardation could be prevented through a program of family intervention beginning in early infancy. This project differs from previous enrichment or intervention efforts in at least two ways. First, the subjects were selected on the basis of epidemiological studies which indicated that children born to parents who are poverty-stricken as well as of low intelligence are at high risk of being identified as mentally retarded. Secondly, the program begins in very early infancy and continues intensive intervention until the children enter first grade. The intention of this program is the prevention of mental retardation, in contrast to attempts aimed at remediation.

Before summarizing the results, I would like to review briefly the background and design of the Milwaukee project.

Approximately twelve years ago, the University of Wisconsin Research and Training Center established the High Risk Population Laboratory to study mental retardation among low income populations. The cultural-familial mentally retarded individual generally remains undetected until he enters school, since such mild intellectual deficiency is difficult to detect in the very young, especially when there is no evidence of organic damage.

The High Risk Population Laboratory is an area of Milwaukee, Wisconsin, previously found to have an extremely high prevalence of retardation, which we began to monitor by continuous door to door surveys. Though this area comprised about 2½% of the population of the city, it yielded approximately 1/3 of the total number of children identified in school as educable mentally retarded. According to U.S. Census Bureau data, the tracts comprising this area were in the lowest category of population density per living unit, percent housing rated as dilapidated, and unemployment.

All families residing in this area with a newborn infant, and at least one other child of the age of six, were interviewed and received individual intellectual appraisal. Through this survey we found clues for identifying families among the economically disadvantaged group with a high probability of producing a retarded child.

Specifically, we found a differential course of intellectual development for children born to mothers with different IQ levels. Furthermore, although there were no significant differences on the early infant intelligence tests between children born to mothers with above 80 IQ, and those born to mothers with below 80 IQ, after the infancy period, the children whose mothers had IQs greater than 80 maintain a steady intellectual level, while the children whose mothers had IQs less than 80 showed a marked progressive decline. (See Figure 1.) This trend toward a decline in measured intelligence for children in disadvantaged environments is widely accepted as a general characteristic of a "slum" environment population, yet these data indicate that the trend of declining intelligence with increasing age is restricted to offspring of low IQ mothers. In fact, we found the variable of maternal intelligence was the best single predictor of low intelligence in the offspring. The data indicated that the lower the maternal IQ, the greater the probability of the children scoring low on intelligence tests, particularly for the offspring of mothers with IQs below 80.

These observations from our survey data suggested our strategy to approach the prevention of socio-cultural mental retardation by attempting to rehabilitate the family rather than simply the individual retarded adult. The ability to select families "at risk" for mental retardation on the basis of maternal intelligence made it possible to initiate a program to study "high-risk" children before they become identified as mentally retarded.

As babies were born in our study area, trained surveyors employed by the University of Wisconsin Survey Research Center contacted the family within a few weeks of birth and completed a family history questionnaire which included a vocabulary screening test administered to the mother. Those mothers falling below a cut-off score on the vocabulary test were administered a full-scale WAIS by a trained psychometrist. A maternal IQ on the WAIS of less than 75 was the selection criterion in accumulating a sample of 40 families. These 40 families were assigned to either the Experimental or Control condition. It was not possible to accumulate at once a sample of 40 families where the mother met the WAIS selection criterion and then randomly assign to Experimental and Control group because of the design requirement that intervention be initiated as early in infancy as possible. Our projections suggested that our screening procedures would identify about three families a month meeting criterion, requiring a little better than one year to accumulate our full sample. In actual fact, our projections were somewhat off; a total of eighteen months were actually required to generate the total sample.

Although this procedure constituted a deviation from strictly random assignment, it should be emphasized that only the happenstance of month of birth dictated group assignment. At no time did factors such as condition of the infant at birth, economic or domestic status of the family, etc., dictate group assignment. In fact, statistical analysis of differences in all measures present and known at time of birth, such as birth weight and height, recorded abnormality of delivery or condition of the infant at birth, marital status of family, economic status, and number of siblings were not significant. In addition, subsequent medical evaluations of the children as they grew have been carried out independently by staff of the Children's Hospital and Marquette Dental School. The analysis of these data revealed no statistically significant differences between groups in height, weight, serum lead levels, or other blood analyses.

Obviously, a major hazard for a longitudinal study of this kind is the potential for substantial attrition. We have been able to minimize this figure. Up to the present time, only two Control families have been lost and all efforts to locate them have failed. The Experimental group lost two subjects very early; one infant died as a result of a sudden crib death, and the second was lost by withdrawal of the mother from the program. This latter case represented the only instance of refusal to participate in the intervention program. Since both these losses occurred while the samples were still being accumulated, they were replaced, bringing the total N to 20; however, more recently, three Experimental families have been lost due to relocation to southern states. In two of these cases, the families left after the children had reached four years of age, and in the third case, after the child was 4½. Contact has been maintained with these three families however, and the children will receive the same comprehensive evaluations at the age of seven, as scheduled for all subjects.

The design of the Milwaukee project study for the Experimental group called for a comprehensive family intervention effort beginning in the home. The Experimental program was comprised of two components: (1) the infant, early childhood stimulation program and (2) a maternal rehabilitation program.

For the newborn infant, the program's objective was to provide intensive language and sensory-motor stimulation, and thereby facilitate the development of cognitive skills. Each day, beginning as soon after birth as was feasible - usually between three and six months of age - the child was picked up at home and brought to the Infant Education Center for the entire day.

The general educational program is best characterized as having a cognitive-language orientation implemented through a structured environment. Individualized prescriptive teaching techniques were utilized in the daily program (7 hours per day, 5 days per week). There was a high teacher-to-child ratio, which gave flexibility to the program and allowed for teacher feedback on the effectiveness of methods as well as individualization of instruction.

The program for the Experimental mother was designed to prepare her for employment and increase her awareness of her environment. This program included vocational training and classes in homemaking and fundamental academic skills.

The Control children, drawn from the same group of families as the Experimentals, were seen only for testing, which was done on a prescribed schedule for both the Experimental and Control groups of children. The testing schedule consisted of a comprehensive array of standardized and non-standardized measures of behavioral development, and was set from infancy to age seven where independent behavior evaluations are scheduled at the project's terminal point.

Our schedule of measurement included (1) developmental schedules of infant adaptive behavior; (2) experimental learning tasks; (3) measures of language development; (4) measures of social development; and (5) standardized tests of general intellectual functioning.

The Experimental and Control infants were on an identical measurement schedule, with assessment sessions every three weeks. The particular measures administered at a given session depended upon the predetermined schedule of measures for that age level. A particular test or task was administered to both Experimental and Control infants by the same person; the testers were not involved in any component of the infant stimulation or maternal program.

The Gesell Developmental Schedules were administered to both the Experimental and Control infants, beginning at age six months. Through the 14 month testing, the groups responded comparably on the four schedules: Motor, Adaptive, Language and Personal-Social. These data are represented as a composite of the four schedules, plotted with the mean scale developmental age norms for each age level tested. (See Figure 2.) At 18 months the Control group began to fall 3-4 months below the Experimental group, although still performing close to Gesell norms. At 22 months the Experimental group scores were from 4½ to 6 months in advance of the Control group on all four schedules while the Control group had fallen below the Gesell norms on the Adaptive and Language schedules.

Beginning at 24 months, increased emphasis was given to experimental, direct measures of learning and performance, as well as to the standardized tests of general intelligence.

The learning research program was designed to assess the longitudinal learning-performance characteristics of young children and to determine the role of these characteristics in the learning process. Furthermore, the role of this part of the assessment program was to provide more comprehensive information about cognitive growth than we were deriving from the IQ tests and various language measures.

We were concerned with delineating some of the characteristics of early learning behavior that either facilitate or interfere with performance. We wanted information on the response patterns or behavior styles, and how a child's simple response choice may reveal his general response tendencies and his ability to select and order incoming stimulation.

We employed a series of tasks including color form and probability matching. Our concern was with the child's strategy of responding: i.e., did he adopt a developmentally sophisticated strategy of consistent responding, either to color or form, or did he respond randomly, or perseverate to position? These learning measures have been administered every year since the children were 2½ years of age.

Our data revealed more developmentally sophisticated responding on all these measures by the Experimental group. Generally, the Experimental children have utilized a response strategy of altering successive responses according to the outcome of their previous responses. The Controls showed a tendency to perseverate on a response, e.g. to choose one position or to alternate from left to right, indicating that the children are insensitive to previous feedback and make no attempt to adopt a strategy.

We feel that in spite of the apparent simplicity of such tasks, they powerfully demonstrated the association of early intellectual development with the ability to impose order on the environment. Piaget (Inhelder and Piaget, 1958) suggests that response stereotyping is a manifestation of logical immaturity, and is a developmentally related deficiency in the use of higher order cognitive strategies. Even at five and six years of age, the percentage of Control children showing a tendency to perseverate was greater than the percentage of Experimental children showing such performance at three and four years of age. Thus, a response behavior which is important for future performance - the strategy or style of responding - appears to develop in the early years. The Control children strategies may interfere with their later learning while the style of the Experimental children should facilitate problem solving performance.

Our second major area of concern was the children's development of language and the measurement of this development.

The first statistically significant difference in language development appeared at 18 months on the Language scale of the Gesell Schedules. By 22 months, the Experimental children were over 4 months ahead of the norms and 6 months ahead of the Controls. This trend of differential language development has continued, in even a more dramatic way. In fact, some of the most striking differences in the performance of the Experimental and Control children are reflected in the research measures of language performance.

The analysis of free speech language samples indicated that the Experimental children between the ages of 1½ and 3 say more in conversation. Using this measurement technique, we find that it is not until three years of age, that the Control group produces a vocabulary comparable to that of the Experimental children. However, since the measure provides a gross picture of increasing language complexity, it actually masks the considerable linguistic differences that existed between the children. These differences show up in the group's performance on the more sophisticated language measures, beginning at age three.

At the age of three we began to test imitation with a sentence repetition test. This is an easily administered instrument which requires the child to repeat 34 sentences of varying length and grammatical complexity. The children's replies are analyzed for omission, substitutions and additions. The omissions are significantly greater for the Control group at every age level from 36 months on, while there is a significant decrease in omissions by the Experimental group every 6 months. Also, the Experimental group has substituted and made additions significantly fewer times in the repetitions. By the age of 4 the Experimental group made significantly more exact repetitions than the Control group, whose performance is comparable to the Experimental group's performance at 3. This same performance differential continues through age 5.

Also beginning at age 3 we tested grammatical comprehension with a modified version of a test developed by Bellugi-Klima (Fraser et al., 1963). This measure is a game in which the child manipulates objects in order to demonstrate his ability to understand 16 grammatical constructions. (The tester gives instructions for the child to fulfill a command, i.e. "Put the ball under the cup".) The results show that the Experimental group's performance is significantly superior at all age levels tested (3, 4 and 5). Their grammatical comprehension is one year, or more, in advance of the Control group.

Our standardized language instrument has been the ITPA, which has been administered to all children over 4½. The results have supported the differential performance of the Experimental and Control groups on our other measures. The mean psycholinguistic age of the Experimental group is 63 months (measured at 54 months) as compared to a mean of 45 months for the Control group: a difference in favor of the Experimental group by over a year and a half.

In describing the language behavior of the Experimental children, one would find them expressive, verbally fluent and according to the ITPA linguistically sophisticated. They speak their own dialect and they are proud of their own speech and yet their performance is developmentally advanced on sophisticated tests of the English language.

The next area we have given attention is mother-child interaction. We were concerned with the effects the intervention program may have had upon the family, particularly the mother. Previous research (e.g. Hess and Shipman, 1968) found that the mother's linguistic and regulatory behavior induces and shapes the information processing strategies and style in her child and can act to either facilitate or limit intellectual growth.

In the mother-child interaction most sophisticated behavior - such as the initiation of problem-solving behavior by verbal clues and verbal prods, or the organization of tasks with respect to goals in problem-solving situations, etc. - is done by the mother. However, where the mother has low IQ, the interaction is more physical, less organized and less direction is given to the child. Indeed, while this was the case in the Control group mother-child dyads, it was quite different in the Experimental dyads.

We found that the Experimental dyads transmitted more information than the Control dyads, and this was a function of the quality of the Experimental child's verbal behavior. The Experimental children supplied more information verbally and initiated more verbal communication than the Control dyads. The children in the Experimental dyad took responsibility for guiding the flow of information - providing most of the verbal information and direction. The mothers of both dyads showed little differences in their teaching ability during the testing session. However, in the Experimental dyads, the children structured the interaction session either by their questioning or by teaching the mother. Also, the Experimental mothers appeared to be modelling some of the behaviors of their children. Consequently, they used more verbal positive reinforcement and more verbal responses.

As a result, a developmentally more sophisticated interaction pattern has developed between the Experimental children and their mothers, which contributed to faster and more successful problem completion.

It became apparent from these data of the mother-child interaction, that the intervention effort has effectively changed the expected pattern of development for the Experimental dyads. Moreover, the result of what might be termed a reciprocal feedback system initiated by the child has been to create a more sophisticated, more satisfying interaction pattern in the Experimental dyad. In fact, there is some evidence that the Experimental mothers might be undergoing some changes in attitude and self-confidence. The Experimental mothers appear to be adopting more of an "internal locus of control" - an attitude that 'things happen' because of their decisions and actions and not purely by chance or fate. Thus, the intensive stimulation program, in which the Experimental children participated, has benefited both the Experimental child and the Experimental mother by broadening their verbal and expressive repertoire.

A clearer picture of the differences between groups is given by the results from standardized measures of intelligence.

We have presented the summary data from intelligence testing in Figure 3.

We have derived data from 12 to 21 months from the Gesell Developmental Schedules. The standardized intelligence scores at 24 months are from Cattells and from Binets, thereafter.

As you can see, the mean IQ of the Experimental group is consistently 25 to 30 points above that of the Controls. For example, at 60 months the mean IQ of the Experimental group was 118 in comparison to the Control mean IQ of 92, a difference of 26 points. We have calculated IQ at the 72 and 84 month points, but they include scores from less than the complete group of Experimental or Control children. These points are particularly important for they are some of the first evaluative data obtained since the children have been out of the intense education program and on their own. Although there was a drop in the scores of the Experimental children to 112 at 72 months, and 110 at 84 months, where some of the children have completed first grade, there is a comparable drop for the Controls, lowering the mean score to 87 at 72 months and 84 at 84 months. It is particularly significant that with the decline in test scores, there has remained a large differential in mean IQ, which the Experimentals have maintained throughout the testing: at 84 months there is still a 30 point difference between the groups. We are encouraged by these preliminary results for if you remember, the purpose of this program was to prevent a decline in intellectual functioning with age increases to the retarded mean IQ level of their siblings and mothers. This decline is in evidence for the Control group, whereas it seems likely that the Experimental group will level off at mean IQ level about 100. Of course, next year's testing will give us a more complete picture of the progress of both groups.

The tendency for declining IQ in this population is further illustrated by the comparison data in Figure 4. The bottom dotted curve is the original survey group. The longer solid line curve is the mean IQ of the siblings from both the Experimental and Control families. In general these older siblings of our actual subjects show the same trend toward declining IQs with increasing age, as do the actual Control children, whose mean IQ data is represented in the shorter solid line curve. Thus, it appears that we have prevented in the Experimental group the relative decline in intellectual development that we see now in the Control group, and that we found in the siblings of both groups and in the original survey groups.

I think these data answer one of two pivotal concerns about the study at this time. One concern, obviously, is the basis for predicting that these children are at risk for retardation. In other words, can we be sure that the downward trend in IQ for this population is reliable? The data I have just shown indicates that it is; successive samplings from four generations of offspring have shown the same tendency to declining IQs. The second major concern at this time is whether the present differential performance favoring the young Experimentals is merely an artifact of training. The strength of the present differential performance in favor of the Experimental group is borne not only by their standardized test scores and their performance levels on the various experimental tasks, but also by the differential behavior patterns displayed by the two groups. The pattern of the Experimental children indicates a sincere and concerned effort to work the task, while the Control children have tended to be apathetic and perseverate their response.

Even with such a comprehensive assessment program, interpretations of, and generalizations based upon present data must be tempered not only by recognition of the test sophistication which has obviously been acquired but also by knowledge of previous enrichment studies where treatment gains have not been maintained over long post-treatment periods. We have planned independent, comprehensive behavioral evaluations to be conducted a year beyond the termination of intervention. These data may prove a more reasonable basis for evaluation of effects on intervention. This is not to suggest that subsequent changes on relative performance levels would not occur beyond that level, but rather, it would provide a more solid basis for evaluation of the treatment effects. Any ultimate evaluation, of course, must be based on the performance of these children as they move through the educational system. We are encouraged by the preliminary results from the members of the Experimental group who have completed first grade. However, final interpretations will be put off until more of the group have reached this point.

We shall continue to be quite cautious in the interpretation of our data. This is not peculiar, particularly when one considers the numerous pitfalls and hazards of infant measurement. The Experimental children have had training, albeit fortuitously, on items included in the curriculum which are sampled by the tests, while the repeated measurements have made both groups test-wise. We have tried very hard to answer whether it has been simply a matter of training and practicing specific skills. In fact, extraordinary precaution has been taken to separate the development of the curriculum and the assessment program. Two separate staffs have been employed. It is obvious to most researchers that, to some extent, infant intelligence tests must contain material which approximates material used in preschool curriculum, primarily because of the limited variety of material for this age. To circumvent this problem somewhat, we employed other measures of performance, which minimized the stock item, and thereby afforded additional insight into the differential development of these children. As could be seen in the measures of learning and language development, the differential performance discrepancy is consistent with the IQ measures, indicating advanced intellectual development of the Experimental group. What's more, there is considerable difference in the pattern or style of behavior between the groups -- particularly the tendency to stereotype in the responses exhibited by the Control group, which certainly is antagonistic to successful learning performance.

[1]Research supported in part by Grant 16-P-56811/5-09, formerly RT-11, from the Rehabilitation Services Administration of the Social and Rehabilitation Service of the Department of Health, Education and Welfare.

[2]A more comprehensive report is available from the first author at the Waisman Center on Mental Retardation and Human Development, 2605 Marsh Lane, University of Wisconsin, Madison, Wisc, 53706.

Fig. 1 IQ Decrements in Disadvantaged Children Whose Mothers are Mentally Retarded

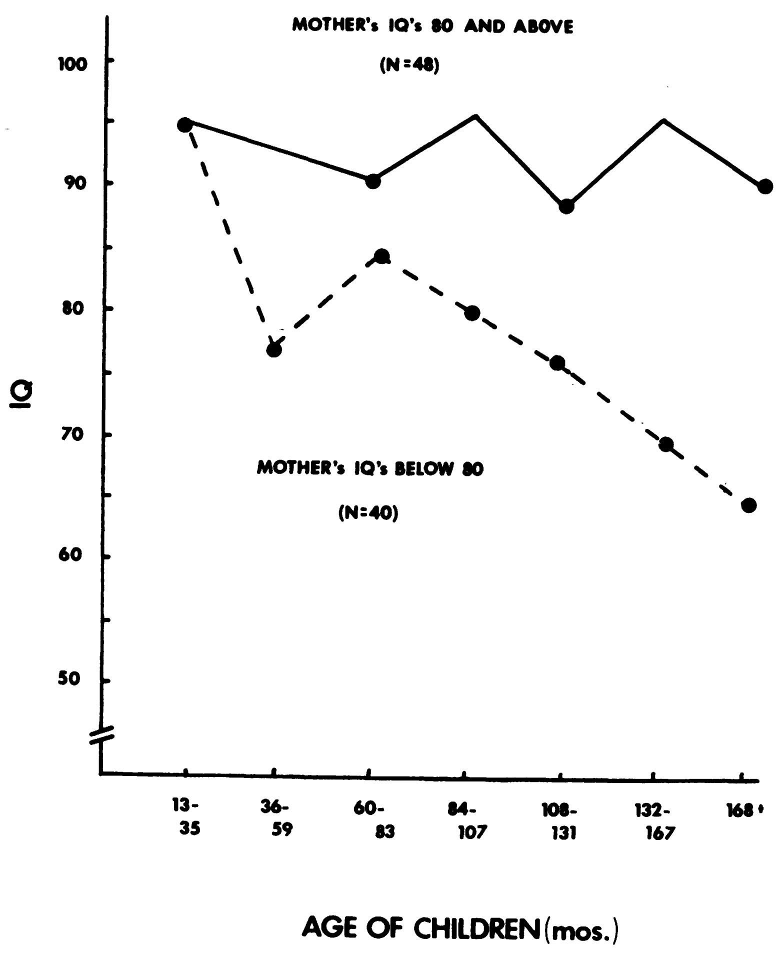
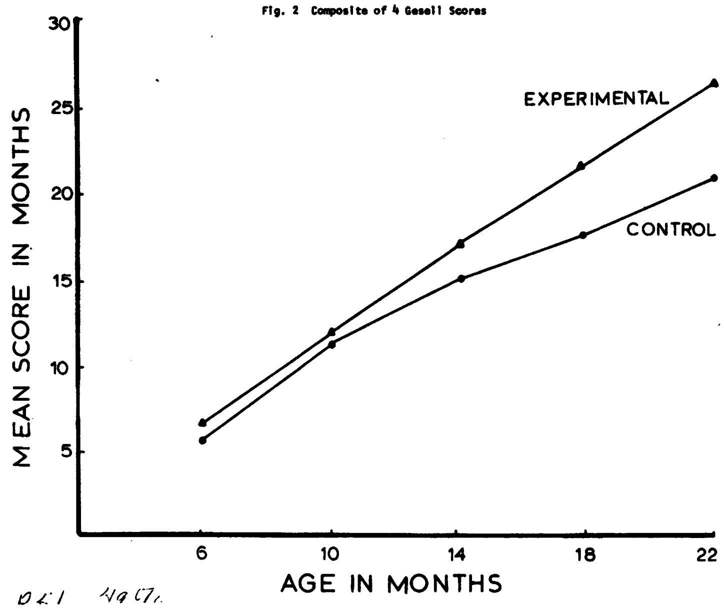


Fig. 2 Composite of 4 Gesell Scores



DZ:1 49 (7/.

Fig. 3 Mean IQ Performance with Increasing Age for the Experimental and Control Groups

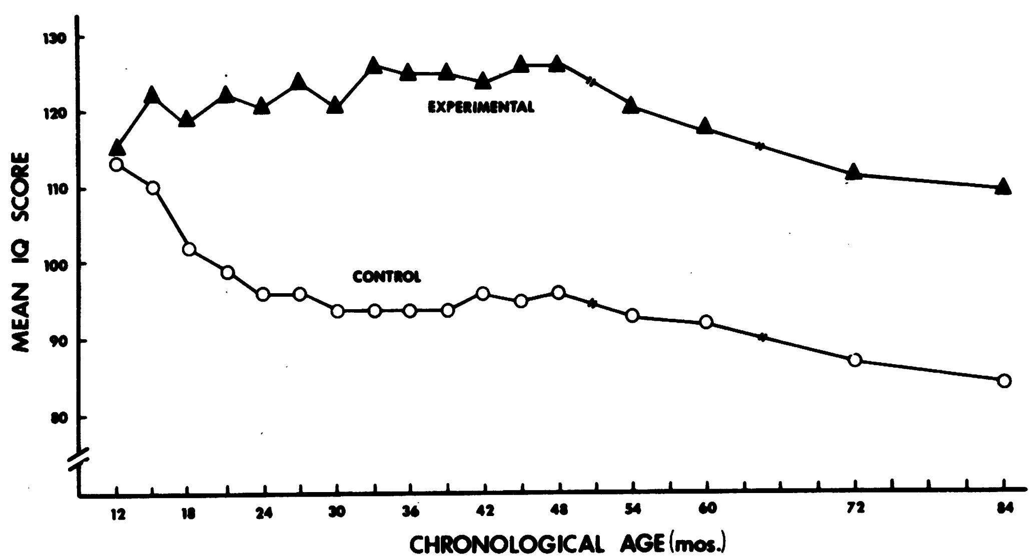


Fig. 4 Sibling and Contrast Group IQ Scores