DOCUMENT RESUME

ED 041 638 24 PS 003 660

AUTHOR Parker, Ronald K.
TITLE The Effectiveness of Special Programs for Rural Isolated Four-Year-Old Children. Final Report.
INSTITUTION Florida State Univ., Tallahassee.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
BUREAU NO BR-9-D-018
PUB DATE Sep 69
GRANT OEG-4-9-190018-0030-057
NOTE 96p.

EDRS PRICE EDRS Price MF-$0.50 HC-$4.90
DESCRIPTORS Cognitive Development, Compensatory Education, *Curriculum Evaluation, Early Experience, Evaluation Techniques, Intelligence Differences, Language Development, *Mobile Classrooms, *Preschool Education, *Program Effectiveness, Readiness, *Rural Areas
IDENTIFIERS Peabody Language Development Kit, PLDK, Walkulla County Florida Schools

ABSTRACT

The objective of this study was to develop and evaluate two procedures for providing preschool education for rural 4-year-olds by using a mobile laboratory. The project used "readimobiles" to determine the effectiveness of a structured, psycholinguistically-based preschool curriculum on black, disadvantaged children. There were three treatment groups: (1) advantaged white children receiving a general enrichment program, (2) disadvantaged black children receiving three months of lessons from the Peabody Language Development Kit, and (3) disadvantaged black children receiving nine months of the Peabody program. Each child in these treatment groups was matched to a control child by age, race, sex, and sociceconomic status. Both groups were posttested twice (to determine reliability) on the Stanford-Binet, the Caldwell Preschool Inventory and the Illinois Test of Psycholinguistic Abilities. The results showed that the experimental subjects were superior to the control subjects on all measures. The treatment groups also differed significantly from one another. Finally, subjects in all groups scored significantly higher on the second posttest. The final report of this document submitted to the Office of Economic Opportunity appeared as ED 039 022. (MH)

Project No. 9-D-018
Grant No. OEG-4-9-190018-0030-057

The Effectiveness of Special Programs For Rural Isolated Four-Year-Old Children

Ronald K. Parker, Ph. D.

The Florida State University

Tallahassee, Florida 32306

September 1969

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Small Grants Research Program

ACKNOWLEDGEMENTS

The completion of this final report represents the cumulative contribution of two funding agencies and the help of many talented individuals. The Small Grants Program of the U.S. Office of Education and the research division of the Office of Economic Opportunity shared equally in providing financial support for the project.

The Southeastern Educational Laboratory and the Wakulla County School System provided the teaching staff and mobile classroom. The principal of the Wakulla County Schools, Mr. William Payne, is a public servant committed to providing the best in preschool education for the children of his county. Hopefully, this project is one step in the right direction to determine what is "best" for preschool children in Wakulla County, Florida.

Mr. Rex Toothman, director of the Preschool Program of the Southeastern Educational Laboratory, served as the administrative backbone of our efforts in Wakulla County. It was through his wise guidance that the program was launched and included a research component. Rex represents one of the "new breed" in early education with a strong committment to empirical evaluations of preschool efforts. The Southeastern Educational Laboratory is fortunate to have Rex Toothman as the administrative head of their Preschool Program and this Wakulla research project profited in numerous ways through its association with Rex.

The daily classroom responsibilities were covered by two teachers, Peggy Gray and Lilian Taylor, and two observers, Margie Lewy and Bill Jennings. Peggy Gray is a master teacher of preschoolers and living proof of a paraprofessional's professional competence.

Dr. Joyce Roll coordinated the evaluation of the program. This was a difficult job which she handled masterfully. One additional person should be mentioned for a significant contribution during the evaluation; Mike Griffey aided the project in a dual role as tester and data manager. Additionally, Dr. Henry Lippert served as a valuable statistical consultant.

Lastly, two bright and professionally competent psychologists helped in preparation of the final report. Without the help of Mary Carol Halbrook and Sue Ambron, I might still be working on the final report.

R.K.P.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (Continued)

First Administrations

LIST OF TABLES (CONTINUED)

First Administrations (Continued)

LIST OF TABLES (CONTINUED)

Second Administrations (Continued)

LIST OF FIGURES

PROBLEM

During the last decade numerous federal and state agencies have developed programs designed to improve the lives of socio-economically disadvantaged Americans. Special assistance has come through increased job opportunities, medical aid, social services, and educational programs. The young child represents the target population for both preventative (e.g., Schaefer, 1965) and remedial (Gordon, 1967) educational programs designed to modify the effects of poverty on the individual.

Support for the interest of early childhood educators and psychologists in modifying the cognitive-intellectual abilities of young children has a base in contemporary theorizing (Hunt, 1961) and the empirical research literature (Elkind, 1967). Hunt (1961) carefully outlines a conception of intelligence in which intelligence is not viewed as constant, nor is it necessarily doomed to develop in a fixed, unmodifiable way. Considerable data are cited to support the contention that intelligence and intellectual development can be modified by means of environmental events. On a more applied level, programs like Project Head Start are based on the assumption that preschool experiences can facilitate school performance of the young socio-economically disadvantaged child.

The rapid growth of preschool educational programs in North America (Reidford, 1968) dramatically underscores the major problem of preschool education today: it is an edifice without a foundation! More specifically, due to the lack of scientific research in the area, we can only make educated guesses about the important variables that influence cognitive and intellectual development during the preschool years (White, 1968). The current lack of scientific information stems from two sources, one historical and one contemporary. Historically, little research was conducted on preschool programs except for those investigations of kindergarten programs for socio-economically advantaged children (Swift, 1964). Almost all of this early research suffered on several important counts (e.g., confounded experimental design), so that generalizations about the merits of nursery school attendance cannot be made because of the inconclusive and contradictory research data.

The contemporary status of preschool education presents a somewhat mixed picture. On the one hand, numerous psychologists and educators are turning their attention to the general area of preschool education and to the specific area of preschool education for socio-economically disadvantaged children (Deutsch, Katz, & Jensen, 1968; Helmuth, 1967; Hess & Bear, 1968; and Webster, 1966). There are at least two noticeable negative by-products from this surge of interest in preschool education: (1) Hundreds of preschool programs exist

that either have no clear statement of curriculum and/or do not have an adequate evaluation component. These efforts are, therefore, basically useless in contributing to a scientific understanding of the important variables in preschool education. (2) Dozens of commercially prepared preschool educational materials have appeared that have not been adequately field tested. It is understandable that publishers develop materials for the current "fad" of preschool education and that authors wish to share their "educated guesses" with the world; however, extreme caution needs to be exercised in the use of these materials.

During the last few years, dozens of programs in urban settings have been developed that look promising. An urgent need now exists for independent assessments of all major preschool curricula and an evaluation of their effects on a variety of subject populations (e.g., rural children). The following prototype programs, in my judgment, illustrate promising research directions that need to be independently evaluated and extended: (1) birth to two-year-olds - Language Tutorial (Schaefer, 1965); Parent Education Project (Gordon, 1967); Project Know How (Parker & Dunham, 1968); Painter Home Tutorial (1968); (2) two- to five-year-olds - Early Training Project (Gray & Klaus, 1965); Structured Psychodiagnostic Program (Karnes, Kirk, Bereiter & Englemann, 1968); Academic Preschool Approach (Bereiter & Englemann, 1966); Learning to Learn School (Sprigle, 1967).

These programs were mentioned because, in general, they have (1) specified their curricula, (2) outlined their theoretical orientation, (3) provided a research evaluation of their program, and (4) produced encouraging results in terms of facilitating cognitive-intellectual development of the participants.

The present research uses the Peabody Language Development Kit (PLDK), Level #P, because (1) the materials are based on psycholinguistic theory, (2) the Peabody materials for older children (Levels K, 1, and 2) have produced impressive results (see the manuals for these three kits for a comprehensive review of the relevant research, American Guidance Service), and (3) the format of the lesson is ideal for paraprofessional teachers to use. The PLDK model was built on Osgood's linguistic theory (1957) which also formed the base of the Illinois Test of Psycholinguistic Abilities (Kirk & McCarthy, 1961). The theoretical model on the nature and training of human intellect by Guilford (1967) was drawn upon in addition to the work of Torrance (1962) in the area of creative thinking. In all four levels (Level #P, #K, #1, and #2) the training of global oral language rather than specific training on selected psycholinguistic processes is stressed. While activities exist for all three components of language, namely reception, expression, and conception, in Level #P stress was placed on auditory reception and on vocal expression. Emphasis is placed on the establishment of an automatic level of sentence structure reflecting basic syntactical rules.

The rationale for the Kits was based, as well, on theory and research related to verbal learning (McGeoch & Irion, 1952). An attempt was made to cast the lessons in keeping with the behavior modification techniques of Skinner (1957). In addition to the use of tangible and token reinforcements, motivation was also built in (1) by having many of the daily lessons contain an activity which allowed for free movement on the part of the group; (2) by providing attractive full-color pictures as well as novel and intriguing records, puppets, magnetic shapes and other materials; (3) by pacing the activities so as to move on when interest lagged; (4) by having as many as possible of the children engaged in all activities at all times; and (5) by selecting elements which were found in field testing to be of high interest value to most children for whom this level of the Kit was devised. The various aspects of language taught by the lessons were programmed for increasing difficulty, though future field testing will probably demonstrate the need for further refinements in this regard. Finally, behavior theory and research was called upon in building overlearning into the lessons (Ellis, 1963; Vergason, 1964).

No attempt is made here to review the research on Level #K, Level #1, and Level #2 of the Peabody series. This literature is carefully summarized in the manuals of the appropriate level of the Kits. Levels #K, #1 and #2 of the PLDK series appear to be effective in stimulating oral language development. The evidence is less clear on the usefulness of the lessons in training intellect and enhancing school achievement--with some notable successes in both cases.

Regarding the research on Level #P of the PLDK, approximately 45 Kits of experimental edition of PLDK Level #P were field tested. Of this total, 14 Kits were placed in situations in which extensive data were collected prior to and following the experimental use of the materials.

These data were derived from measurements using one or more of the following tests: Peabody Picture Vocabulary Test (PPVT); Stanford-Binet Form L-M (S-B); Illinois Test of Psycholinguistic Abilities (ITPA); and the Test of Grammar adapted from Berko (1958).

The most extensively researched study was conducted in Nashville at a day care center for four- and five-year-old children. The 25 experimental children at this center were compared with a control population of 28 children in the same age range at another Nashville community center. The children in both groups were exposed to daily programs typical of the approach used in day care centers in the Nashville area. In addition, the experimental group received lessons from the experimental version of Level #P of the PLDK. The children in the two groups were compared on the following tests: PPVT, S-B, ITPA and the Test of Grammar. After a seven-month treatment period, a comparison of gain scores from pre- to posttesting on the Binet yielded a gain of 9.6 points in the experimental group, as contrasted to a loss of 2.3 points in the control group. The changes in PPVT performance were +12.0 and +7.8 IQ points for the experimental and control groups respectively. On the ITPA, the experimental group gained an average of 7.2 months, while the control group gained an average of 3.9 months in language age. The subtests on which the experimental group made the greatest gains were Visual-Motor Association (+23.4 months), Auditory Decoding (+16.8 months), and Auditory-Vocal Association (+12.8 months). They showed regression in one subtest, Auditory-Vocal Sequential (immediate auditory memory for digits). By contrast, the control group showed regression in the following four of the nine subtests: Vocal Encoding (-5.8 months); Motor Encoding (-3.95 months); Auditory Vocal Automatic (-5.1 months); and Visual Motor Sequential (-1.5 months).

A more extensive analysis of the data on three of the nine subtests of the ITPA was carried out by Morris (1967). She compared the pre-and postprogram performance of the experimental and control groups on the Auditory-Vocal Association, the Auditory-Vocal Automatic, and the Vocal Encoding subtests of the ITPA. In addition, she studied the postprogram performance of the experimental group on the Test of Grammar (adapted from Berko, 1958) and correlated the scores from this latter test with the three ITPA subtests scores. The results of the Morris study were as follows:

1. Prior to the experimental period, both the experimental and control groups were substantially retarded in language age on each of the three ITPA subtests. The groups were essentially equivalent on the Auditory-Vocal Automatic and the Vocal Encoding subtests, but the language age of the experimental group exceeded that of the control group by about 11 months on the Auditory-Vocal Association subtest.

2. After seven months of daily instructions with the Level #P lessons, a marked improvement in performance was noted among the experimental group on two of the three subtests. The mean language age score reached approximated their age norm in one instance (Auditory-Vocal Association), and was four months nearer it in the other (Vocal Encoding). The control group, while gaining on the Auditory-Vocal Association subtest in the same period, demonstrated statistically significant decreases in performance on the other two subtests.

3. Neither the group receiving language instruction nor the control group showed statistically significant improvement on the Auditory-Vocal Automatic subtest, which assessed the ability to apply grammatical rules, e.g., "Here is a ball; here are two _____." The performance of both groups was also poor on the Grammar Test, which was designed to assess the same type of linguistic skills but with different test material. Both groups were substantially below their age norms on both tests.

A summary of the conclusions of Morris (1967) are as follows: The language teaching device known as Level #P of the PLDK may be expected to contribute to the development of certain expressive language skills to a much greater degree than the traditional program offered in day care centers for culturally deprived preschool children. However, in its present experimental form, it probably will not produce positive changes in such children with regard to their application of grammatical rules for inflectional endings. Therefore, while facilitation of the language development of culturally deprived preschool children can be best enhanced with special language training, such as that provided by Level #P of the PLDK, additional types of teaching media will be required to improve specific grammar skills unless such additional materials are incorporated into the final version of the Kit.

These findings led the authors to place heavy emphasis, in the final rewriting of Level #P, on activities which would promote positive changes in the syntactical and grammatical aspects of language. Further research to test the effectiveness of the final version of the Kits in this regard is needed.

A second study of the experimental edition of Level #P involved a group of 29 mentally retarded children in residence in a school in Wisconsin. These children were also given seven months of training with PLDK, Level #P. Their performance on the ITPA, prior to the program and immediately after it, was compared. At the outset, all of the children fell in the M.A. range of three to five years and had IQ scores which generally clustered around 50. The total group was divided into three subgroups by virtue of their level of placement in the school program. The mean language age gain scores in months after seven months of the experimental edition of PLDK #P lessons were 10.2, 14.2 and 14.0 months for the Adjustment, Pre-Kindergarten and Kindergarten groups respectively. The mean gains in language age for the total group on the nine ITPA subtests ranged from a low of 8.8 months (Auditory-Vocal Automatic and Auditory-Vocal Sequential) to a high of 30.5 months (Visual Motor Sequential).

The Research to date on Level #P of the Peabody Language Development Kits has been based on the experimental version of the Kit and only on the first part of that version. Generally, the findings were heartening in terms of stimulating overall growth in oral language and verbal intelligence. However, the experimental edition did not stimulate grammatical-syntactical aspects of language to the extent desired. Therefore, in developing the final version, a much heavier concentration of exercises was included in this area and a series of songs was devised to make certain syntactical rules automatic. Too, the final edition was expanded by about one-third. Each of the 180 daily lessons was divided in a Part A and a Part B, with two activities generally provided in each. Thus, the Kit now contains what could be described as 360 sublessons. It is hoped that the increased emphasis on syntax and the extension of the training program will overcome weaknesses discerned in the experimental edition. It remains for future research to advance knowledge about the effectiveness of the Kit in its present form, especially with regard to fostering grammatic skills in disadvantaged and retarded children. The present research will provide an independent evaluation of Level #P using rural four-year-old disadvantaged children as subjects.

In summary, the major problem of preschool education is to build itself a strong foundation based on empirical research. The following four-step approach appears to be a reasonable plan: (1) to continue developing prototype preschool curricula from various theoretical positions; (2) to design instructional systems to implement these curricula (e.g., multimedia, use of paraprofessionals, etc.); (3) to carefully evaluate these curricula before premature widespread adoption; and (4) to develop imaginative procedures to implement curricula in special settings with different populations (e.g., rural children, school system with a low budget, advantaged and disadvantaged children, etc.).

The objective of the present proposal was to develop and evaluate two procedures for providing preschool education for rural four-year-olds by using a mobile lab.

The Southeastern Educational Laboratory started a "readi-mobile" program during the 1967-68 school year (Toothman, 1968). The program's purpose is to "design, field test, and demonstrate the application of a mobile instructional unit in providing readiness experiences to preschool age children in geographically isolated areas." The following guidelines exist for comparable implementation of the Readimobile program in six Southeastern locations:

1. Sites should be located which provide easy access to the Readimobile for groups of about 15 children.

2. The Readimobile program will visit each site twice weekly.

3. Exclusive of Readimobile travel and preparation in time, each stop will be two hours in length.

4. Each Readimobile is to be staffed by two para-professionals (indigenous high school graduates).

In Wakulla County, Florida (the poorest county in Florida in terms of per capita income), the Readimobile stops at five locations--Shadeville, Sopchoppy, Crawfordville, Panacea and Buckhorn. The usual weekly Readimobile "Curriculum" can best be described as general cultural enrichment experiences provided basically through the use of films with supplementary introductory or follow-up activities.

In June of 1968, Mr. Rex Toothman, Director of the Readi-mobile Program, asked me to react to the program. The essence of my comments can be summarized as follows: The program, while possibly providing a socialization function and serving to develop positive interpersonal relations, will probably fail to have any meaningful impact on the disadvantaged participants' cognitive-intellectual-language development and consequently his "readiness" for school. The general cultural enrichment experiences appear as vague and unstructured as those of similar programs that have failed to improve school readiness (Alpern, 1966).

These comments, focusing on cognitive variables, are not meant to minimize the importance of gains in areas such as social-emotional development. Bereiter and Englemann (1966), however, provide convincing arguments for focusing on specific deficits (e.g., language behavior) of the disadvantaged children during the brief preschool day. Their argument, simply put, is that we cannot help these children in all areas of development, so we must concentrate on those areas most likely to have high pay-off in terms of stimulating cognitive-intellectual-language development and, consequently, school readiness.

The purpose of this project was to compare the effectiveness of a structured psycholinguistically based preschool curriculum on disadvantaged black four-year-old children. One group (#3) received instruction across a nine-month school year while another group (#2) received instruction for only three months. Additionally, the performances of these two groups were compared to a group (#1) of advantaged (by local standards) white children receiving the general enrichment curriculum of the Readimobile. Even though race and curricula are experimentally confounded when comparing these three groups, our interest in adding the white children was to provide local "norms" for comparison purposes. In essence, we were wondering if our structured treatments would mask the often reported differences between black and white children on a variety of dependent measures.

PROCEDURES

Three groups of eight four-year-old children served as the treated population in this study. Group 1, the general enrichment curriculum, is represented by children who participated in the standard 1968-1969 Readimobile Program at the Panacea location in Wakulla County. A second group (Group 2) of the children at the Buckhorn location received lessons from the Peabody Language Development Kit, Level P (American Guidance Service, 1968) for the last three months of the 1968-69 program. Group 3, also the structured curriculum, is represented by the children who participated in the 1968-69 Readimobile Program at the Shadeville location using the Peabody Language Development Kit for 9 months. The children in Group 1 were white children from families with a median income of $4,500 and whose parents had a median of 12 years of education. The children in Groups 2 and 3 were black children whose families' median income was below $3,000 and their parents' median education was 8 years. Each group was composed of five males and three females.

Each child in the treated population was matched with an untreated control child with respect to age (within three weeks), race, sex and socio-economic status. The control population was obtained from rural portions of adjoining Leon and Gadston Counties, which do not have a preschool program. None of the control children had ever attended a nursery school or any type of preschool program.

The programs for Groups 1 and 3 started in September, 1968, and continued until June, 1969. The program for Group 2 lasted from March, 1969, until June, 1969. The Readimobile paraprofessional teachers, Mrs. Gray and Miss Taylor, were the same for all three groups.

Contact hours for Group 1 were 8:00 - 12:00 a.m. on Wednesday mornings. The daily schedule was quite flexible with a general enrichment curriculum including films with supplementary, introductory or follow-up activities. At the end of the 9-month program, the contact hours totaled 144.

The contact hours for Group 2 (Peabody curriculum for three months) were 8:00 - 12:00 a.m. on Friday mornings with a schedule similar to that of Group 3, only including more lessons due to the four hours of contact in one day rather than four hours divided over two days. Group 2 met for a total of 48 hours of contact during the three months.

The contact hours for Group 3 (Peabody curriculum, nine months) were 9:00 - 11:00 a.m. on Tuesday and Thursday mornings. A typical day's schedule is outlined as follows:

9:00 - 9:20	Peabody Lesson 14A
9:20 - 9:40	Peabody Lesson 14B
9:40 - 10:00	Outside structured Play (e.g., learn parts of the body, concepts such as near -- far, up -- down, etc. while playing)
10:00 - 10:20	Peabody Lesson 15A
10:20 - 10:40	Peabody Lesson 15B
10:40 - 11:00	Remedial work on earlier lessons.

Since Group 3 met only twice per week, the children did not cover all of the 180 lessons of the Peabody Kit during the nine months. However, Group 3, like Group 1, met for a total of 144 contact hours.

Evaluation

Both internal and external evaluations were employed to document the changes across time of Group 3 (Peabody Curriculum, nine months) and to determine if differences existed among the three treated groups and the three untreated groups on measures of intelligence, language, school readiness, and cognition.

I. Internal Evaluation. The internal evaluation of Group 3 (Peabody Curriculum) was accomplished by having two observers (Miss Lewy and Mr. Jennings) record each child's responses to the Peabody Lessons (see Madsen & Madsen, 1969 for procedures). These data were to serve two purposes: (1) to document attainment levels of each child throughout the year and (2) to serve as diagnostic data for the teachers. With regard to the first purpose, this approach enabled us to not only keep accurate, up to date records on each child's learning progress on each concept, skill, or task, but also to identify the strengths and weaknesses of the curriculum materials on this subject population. For example, how many "trials" were necessary for these children to learn the meaning of "under -- over," "up -- down," "big -- little." Many of these concepts and tasks were presented as a twenty minute lesson, yet we already had sufficient data to show that, on the average, much more time needs to be devoted to each of these lessons.

The second purpose of the internal evaluation, diagnosis of attainment levels, enabled the teachers to group the children on each occasion to capitalize on past learning. For example, consider the problem of teaching children to identify (receptive language) and name (productive language) the primary colors. Initially, none of the children could identify or name more than one color accurately. After only two sessions, our records indicated that five children had made rapid progress in color identification and naming. These children were then advanced to more challenging tasks, while the remaining children continued the elementary review on color concepts. This was a deliberate attempt to maximize the use of the child's time since the "Readimobile Preschool" only lasted four hours each week. This was, of course, the essence of some experiments in individually prescribing instruction (ERIE, 1968) and the approach taken in computer assisted instruction (Hansen, 1966). In this regard, our daily diagnosis and structured approach to preschool education was instituted to insure that these children in four hours per week had more opportunities for specific learning than children in a conventional preschool setting that meets three hours daily or fifteen hours per week.

II. External Evaluation. The external evaluation represents the more traditional approach in which children are assigned to groups, given or withheld a treatment (independent variable), and then the effects of the experimental or control placement are assessed (dependent variable). The children were evaluated in May and June of 1969, using the following instruments:

Intelligence:	Stanford - Binet
Language:	Illinois Test of Psycholinguistic Abilities (Revised form, 1968)
Behavior Inventory:	Caldwell Preschool Inventory
Cognition:	Englemann's and Bereiter's Concept Inventory Scale
School Readiness:	Metropolitan Readiness Tests

The external evaluation did not follow a random assignment of Ss to treatment groups and a pretest-posttest design for two reasons. First, the group composition was determined by where the Readimobile stopped, and there was no opportunity to randomly assign children to location. Second, pretests were not administered because there were no funds available for pretesting. It is probably true that the lower-class black children in rural Wakulla, Leon and Gadston Counties form a relatively homogenous group since poverty is so widespread in Northern Florida among rural blacks.

Particular caution was exercised in evaluating the distal control subjects. Much research exists emphasizing that non-intellectual variables, such as rapport between the examiner and child, markedly influence children's responses in testing situations (Bereiter & Engelmann, 1966; Glick, 1968; Zigler & Butterfield, 1968).

The following precautions were taken to ensure valid test results:

1. A team of experienced examiners was hired.

2. The race and sex of the examiner was the same as the child's race and sex.

3. The Readimobile children (Groups 1, 2 and 3) were tested at their usual preschool sites.

4. The control children were tested in some suitable location in their homes. (An attempt was made to test the first two control children in a near-by elementary school, but the children became upset and did not respond well to the test items.)

5. Each child received a maximum of 45 minutes of testing per day to avoid fatigue and restlessness. Frequent rest breaks were also provided.

6. The examiner devoted the amount of time necessary to establish rapport with each child. Particular caution was used with the control children.

Three of the instruments, the Stanford-Binet, ITPA, and the Caldwell Preschool Inventory were administered a second time to each child. The purpose of the second administration was to determine the test-retest reliability with this particular population. The second administration followed within one month of the first administration of each instrument.

III. Teacher Evaluation. The teacher evaluation consisted of direct observation of their behavior in the structured curriculum setting, Group 3. The fundamental question was: Can bright high school graduates who are highly motivated be taught the principles of behavior modification and how to implement a "packaged" preschool curriculum. The teacher training program was as follows:

1. Didactic orientation. This included reading and discussion of the use of behavior modification principles (Madsen & Madsen, 1969), needs of the socio-economically deprived preschool child, and rationale for the Peabody Lessons.

2. Role modeling. During September and October, the research director (Parker) and the observers (Jennings and Lewy) demonstrated how each lesson was to be used with children. After that time, the teachers were responsible for introducing the lessons.

3. Planning daily activities. After four months (September, October, November, December) the teachers slowly assumed more and more of the responsibility for planning and sequencing each day's activities.

Records were kept on how often the research director or observers had to "intervene" with constructive criticism or were asked by the teachers for help. Obviously, the goal was for the teachers to become completely autonomous in selecting and implementing the materials. This research documents how long it takes for this type of paraprofessional to be trained in the use of behavior modification techniques and "packaged" curricula.

RESULTS

Throughout the remainder of this paper the following code will be used to differentiate the curricula groups and their controls.

	Experimental	Control
General Enrichment	GE	GE-C
Peabody - 3 months	P3	P3-C
Peabody - 9 months	P9	P9-C

Analysis of variance tables are in the appendix. Tables summarizing the Duncan's Multiple Range Test are included in the text.

Administration of Dependent Measures

The Binet, ITPA, and Caldwell were administered as posttests on two separate occasions in order to determine the stability (i.e., reliability) of these test scores on this population of Ss. Table 1 presents the test-retest reliability coefficients between the first and second administrations of the Binet, ITPA, and Caldwell total batteries and subtests. It is clear that all of the coefficients are large and statistically significant.

In order to carefully examine the group mean scores, a 2x3x2 analysis of variance was computed on each dependent measure. Each analysis included the following variables: Treatment (Experimental vs. Control) x Curriculum (GE, P3, P9) x Administration (First Administration vs. Second Administration). Tables 21, 22, and 23 summarize the results of these analyses.

The results can be briefly summarized as follows: (1) the experimental Ss were superior on all measures on both occasions to the control Ss; (2) the treatment groups differed significantly from one another (these two findings are to be thoroughly discussed later in the results section); (3) the Ss in all groups scored significantly higher on the second administration of the measures than on the first administration; and (4) there was a significant curriculum x administration interaction on the Caldwell measure.

TABLE 1

Test-Retest Reliability Coefficients

Between First And Second Administrations

Total Batteries	
Stanford-Binet	.9093
Illinois Test of Psycholinguistic Abilities	.9061
Caldwell	.8621
Subtests of the Illinois Test of Psycholinguistic Abilities	
Auditory Reception	.5926
Visual Reception	.5281
Visual Sequential Memory	.5781
Auditory Association	.7848
Auditory Sequential Memory	.8238
Visual Association	.6525
Visual Closure	.7040
Verbal Expression	.6842
Grammatical Closure	.5267
Manual Expression	.7457
Auditory Closure	.7305
Sound Blending	.4855
Subtests of the Caldwell Preschool Inventory	
1. Personal-Social Responsiveness	.7362
2. Associative Vocabulary	.6658
3. Concept Activation-Numerical	.8082
4. Concept Activation-Sensory	.8215

$p < .05$ if $r \geq .288$

$p < .01$ if $r \geq .372$

Since the Ss in all groups scored higher on the second administration of the measures than on the first administration, a comparison of the mean scores for the groups on each measure should be enlightening. On the Binet, the IQ scores for the experimental Ss were 97.00 and 100.58; the control Ss scored 86.79 and 90.92. In each case the gain was approximately 4 IQ points. On the ITPA, the raw scores for the experimental Ss were 136.50 and 154.71; the control Ss' mean scores were 114.17 and 129.00. The gains for the experimental and control Ss were approximately 18 and 15 raw score points respectively. On the Caldwell, the scores for experimental Ss were 49.20 and 53.88; the control Ss mean scores were 41.04 and 45.54. While the gains were approximately 5 points for both the experimental and control Ss, the interaction between the curricula groups and the administration of the Caldwell provides the opportunity for a more refined examination of these data. Figure 1 presents this interaction, revealing a dramatic increase in performance of the P3 group between the first and second administration of the Caldwell.

All groups improved in performance on the second administration of Binet, ITPA, and Caldwell; therefore, the subsequent analyses will use scores obtained on the second administration of these tests.

Stanford-Binet Intelligence Quotient

An analysis of variance of the Stanford-Binet IQ scores revealed that there were main effects of treatment and curriculum. There was almost a 10 point difference between the means of the treatment groups ($\bar{X}$ = 100.58) and the control groups ($\bar{X}$ = 90.92). In Figure 2 the mean IQ of each of the six groups is depicted.

The Duncan's New Multiple Range Test was applied to the means of the six groups and is summarized in Table 2. The GE group ($\bar{X}$ = 108.13) scored significantly higher than the P3 group ($\bar{X}$ = 93.00), the P3-C group ($\bar{X}$ = 87.38), and the P9-C group ($\bar{X}$ = 84.63). However, the GE group did not score significantly higher than its own control group ($\bar{X}$ = 100.75) or the P9 group ($\bar{X}$ = 100.63). The GE-C group ($\bar{X}$ = 100.75) scored significantly higher than either Peabody control group. There was no significant difference between the P9 group and the GE group.

These results reveal differences in the effectiveness of the GE and P9 curricula on their respective populations. The children involved in the GE curriculum did not score significantly higher on the Stanford-Binet than did their controls, while children in the P9 curriculum did score significantly higher than their controls. Since the mean IQ's of the two Peabody control groups were

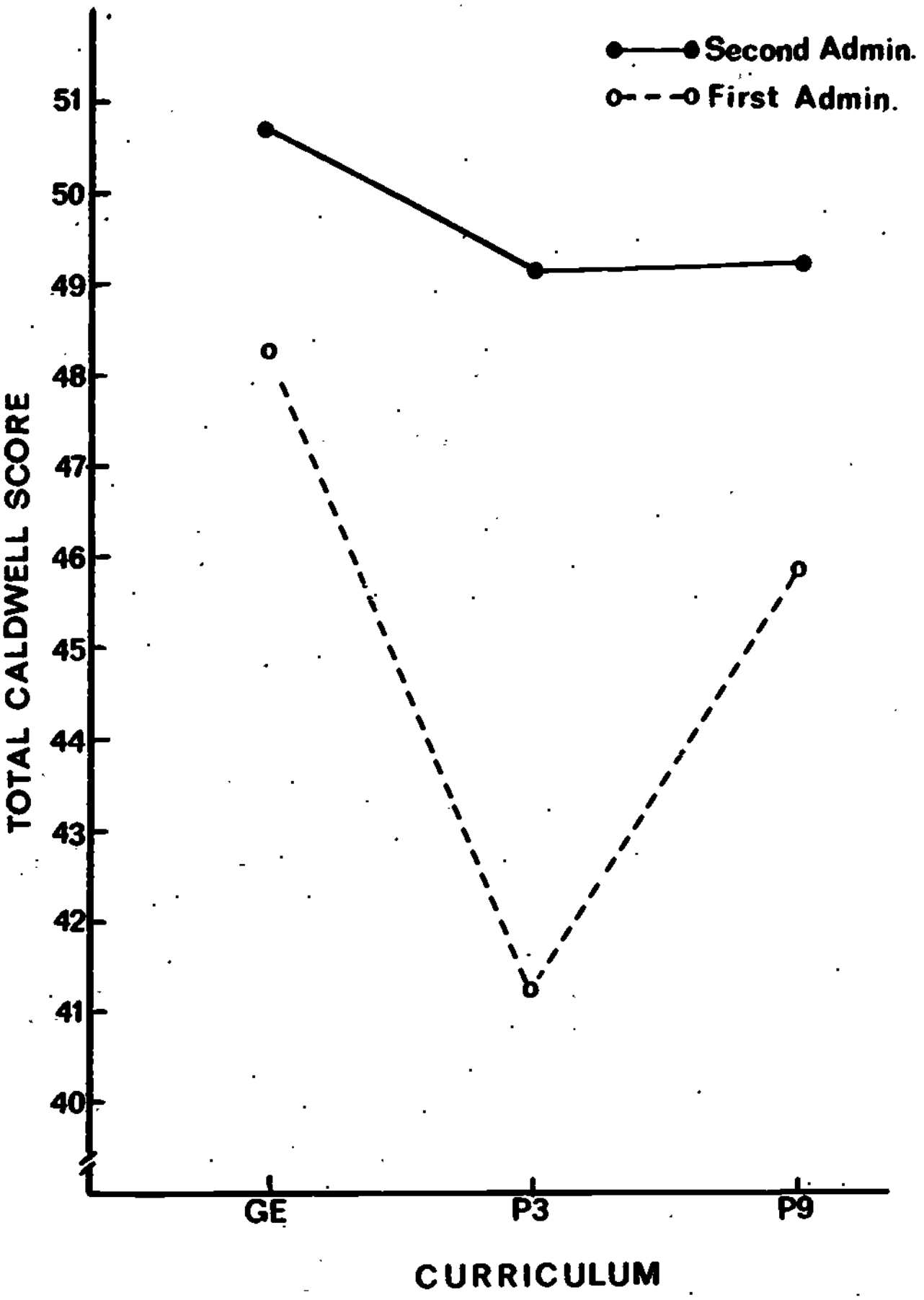

Fig. 1. A comparison of the three curricula groups, including their respective control groups, on the first and second administrations of the Caldwell Preschool Inventory.

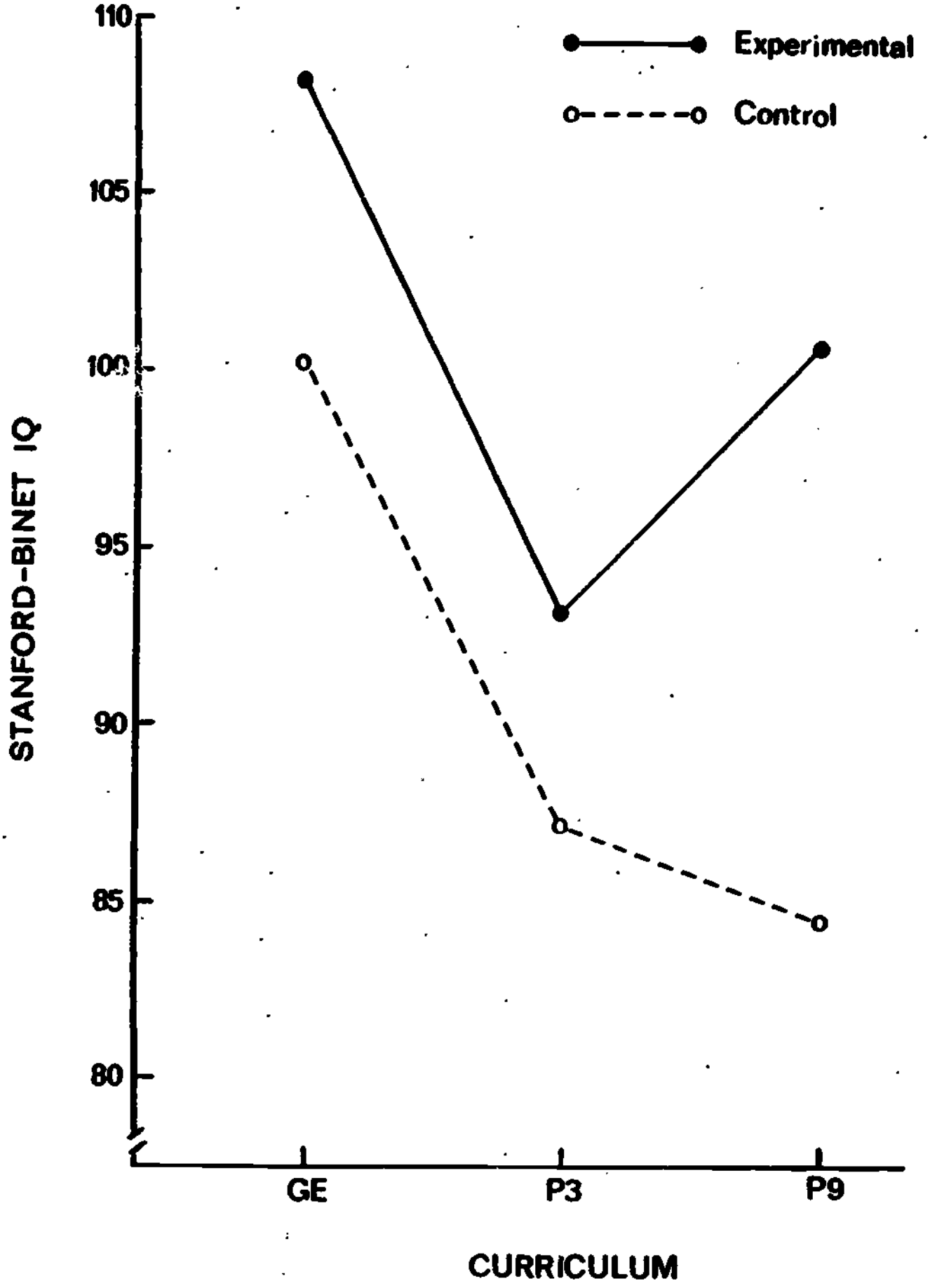

Fig. 2. A comparison of the three curricula groups and their control groups on mean Stanford-Binet IQ scores.

TABLE 2

Binet IQ - Second Administration

Duncan's New Multiple Range Test

of Differences Between Means

		P9-C	P3-C	P3	P9	GE-C	GE
Means		84.63	87.38	93.00	100.63	100.75	108.13
P9-C	84.63		2.75	8.37	16.00**	16.12**	23.50**
P3-C	87.38			5.62	13.25*	13.37*	20.75**
P3	93.00				7.63	7.75	15.13**
P9	100.63					.12	7.50
GE-C	100.75						7.38

P9-C P3-C P3 P9 GE-C GE

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

** $p < .01$

significantly lower than the mean IQ of the GE-C group, it is of note that (1) the P3 and P9 groups were not significantly different from the GE-C group and (2) the P9 group was not significantly different from the GE curriculum group as well as the GE-C group.

Illinois Test of Psycholinguistic Abilities

An analysis of variance of the total ITPA scaled scores indicated that there were main effects of treatment and curriculum. The mean for the treatment group was 396.67 and for the control group 356.79 Duncan's Multiple Range Test (Table 3) shows that the GE group ($\bar{X}$ = 477.13) scored significantly higher than all other groups. The GE-C group scored significantly higher than both Peabody control groups and the P3 group. However, the scores of the GE-C group ($\bar{X}$ = 415.75) and the P9 group ($\bar{X}$ = 376.13) were not significantly different. Thus it appears that both the GE and the P9 curricula were effective in increasing language skills.

Figure 3 illustrates the comparison of the mean total ITPA scaled score for each curriculum group and control group. ITPA scaled scores rather than psycholinguistic ages are used because the examiner's manual for the ITPA gives composite psycholinguistic age norms based only on 10 subtests rather than on the 12 subtests comprising the total ITPA test battery used in this study.

Subtests of the ITPA

Figures 4, 5 and 6 compare each curriculum group with its control group on the twelve subtests of the ITPA. It can be seen in Figures 4 and 5 that the GE group and the P9 group had a higher mean profile than did their respective control groups. Figure 6 demonstrates that the P3 group did not have a higher profile than its control group. Apparently, the Peabody curriculum did not significantly increase language skill when implemented for only three months.

Figure 7 compares the profiles of the three curricula groups. The GE group obtained the highest mean profile. The P9 group obtained a slightly lower profile and the P3 group had the lowest mean profile. However, it is interesting to note the similarity of the P9 profile and the GE-C profile (Figure 8).

Analyses of variance were applied to scores from each of the subtests, and where significant effects were indicated, the Duncan's Multiple Range Test was used.

TABLE 3

Total ITPA Scaled Score - Second Administration

Duncan's New Multiple Range Test

of Differences Between Means

		P3-C	P9-C	P3	P9	GE-C	GE
Means		329.00	325.63	336.75	376.13	415.75	477.13
P3-C	329.00		6.63	7.75	47.13	86.75**	148.13**
P9-C	335.63			1.12	40.38	80.12**	141.50**
P3	336.75				39.38	79.00**	140.38**
P9	376.13					39.62	101.00**
GE-C	415.75						61.38*

P3-C	P9-C	P3	P9	GE-C	GE

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

** $p < .01$

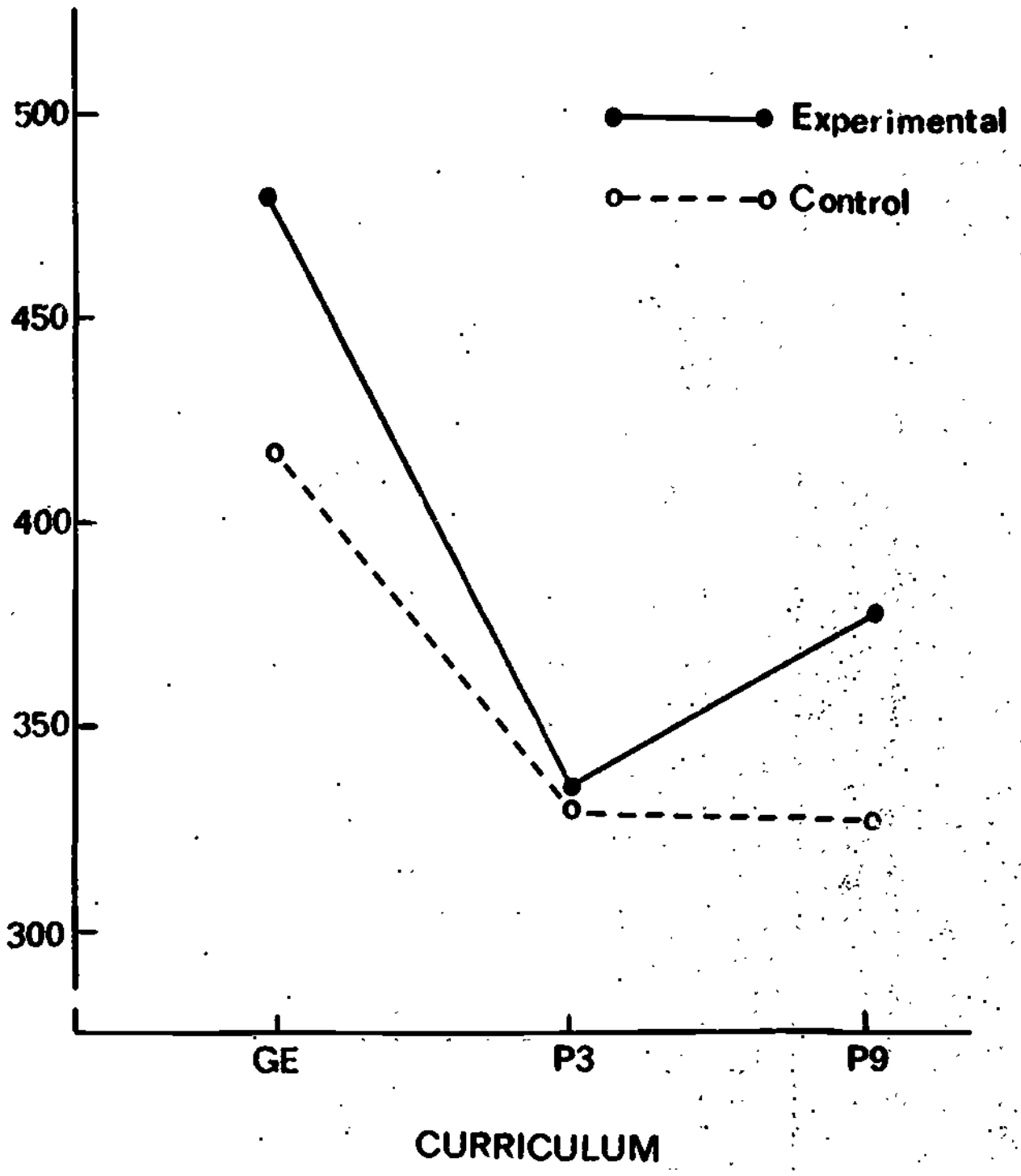

Fig. 3. A comparison of the mean total ITPA scaled scores for each curriculum group and its control group.

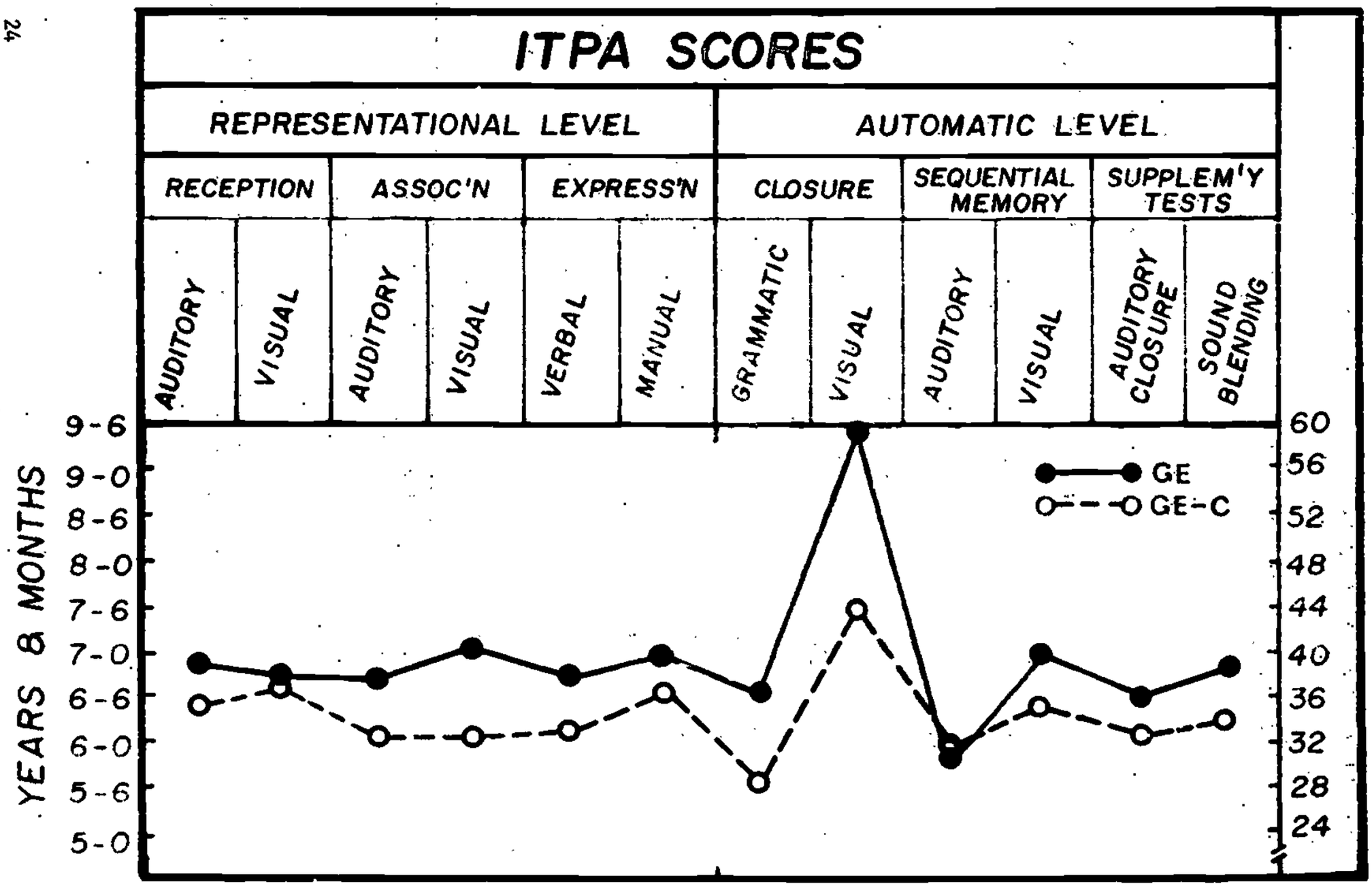

Fig. 4. A comparison of the means of the General Enrichment group and its control group on the 12 subtests of the ITPA.

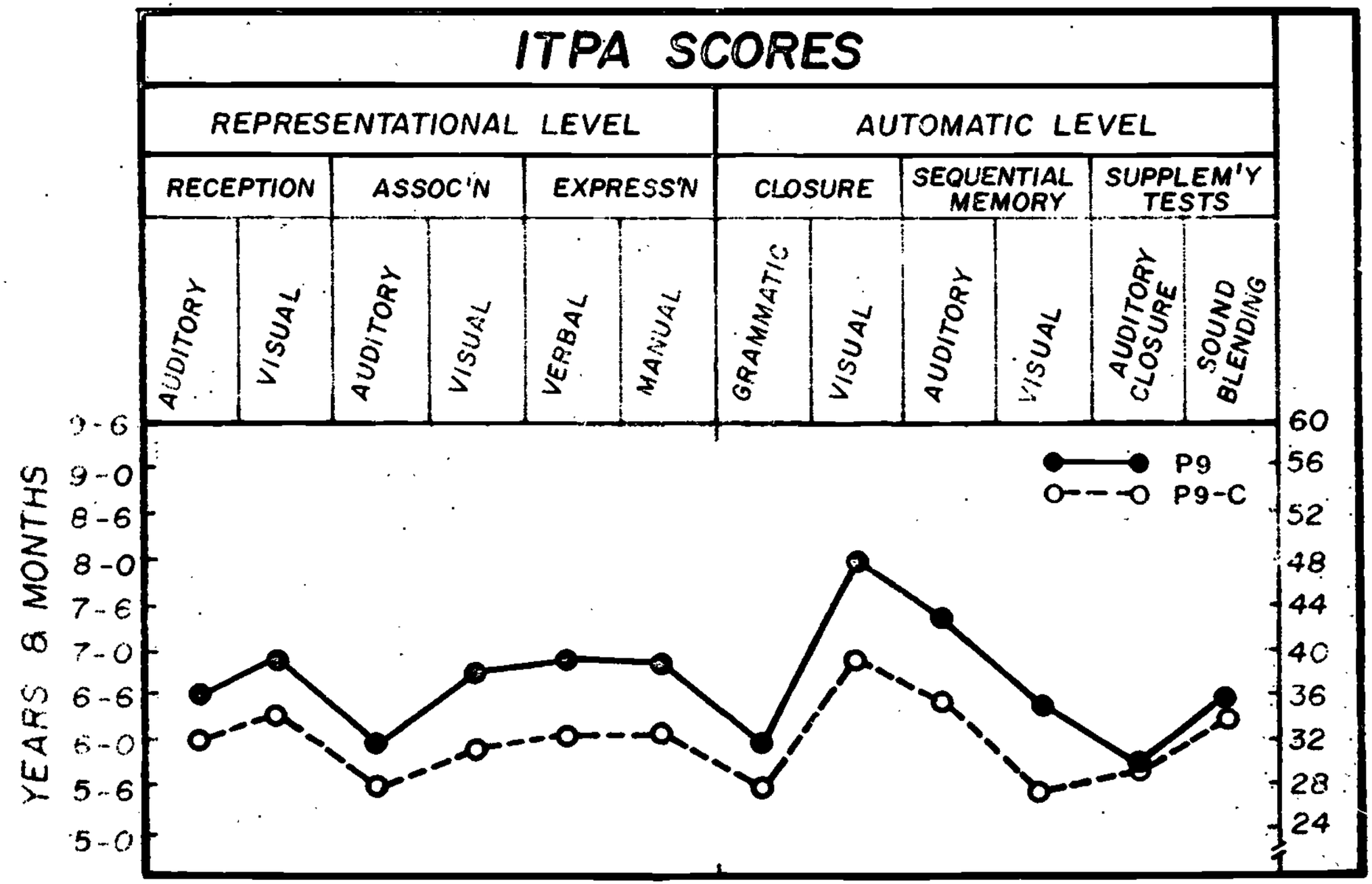

Fig. 5. A comparison of the means of the Peabody 9-month group and its control group on the 12 subtests of the ITPA.

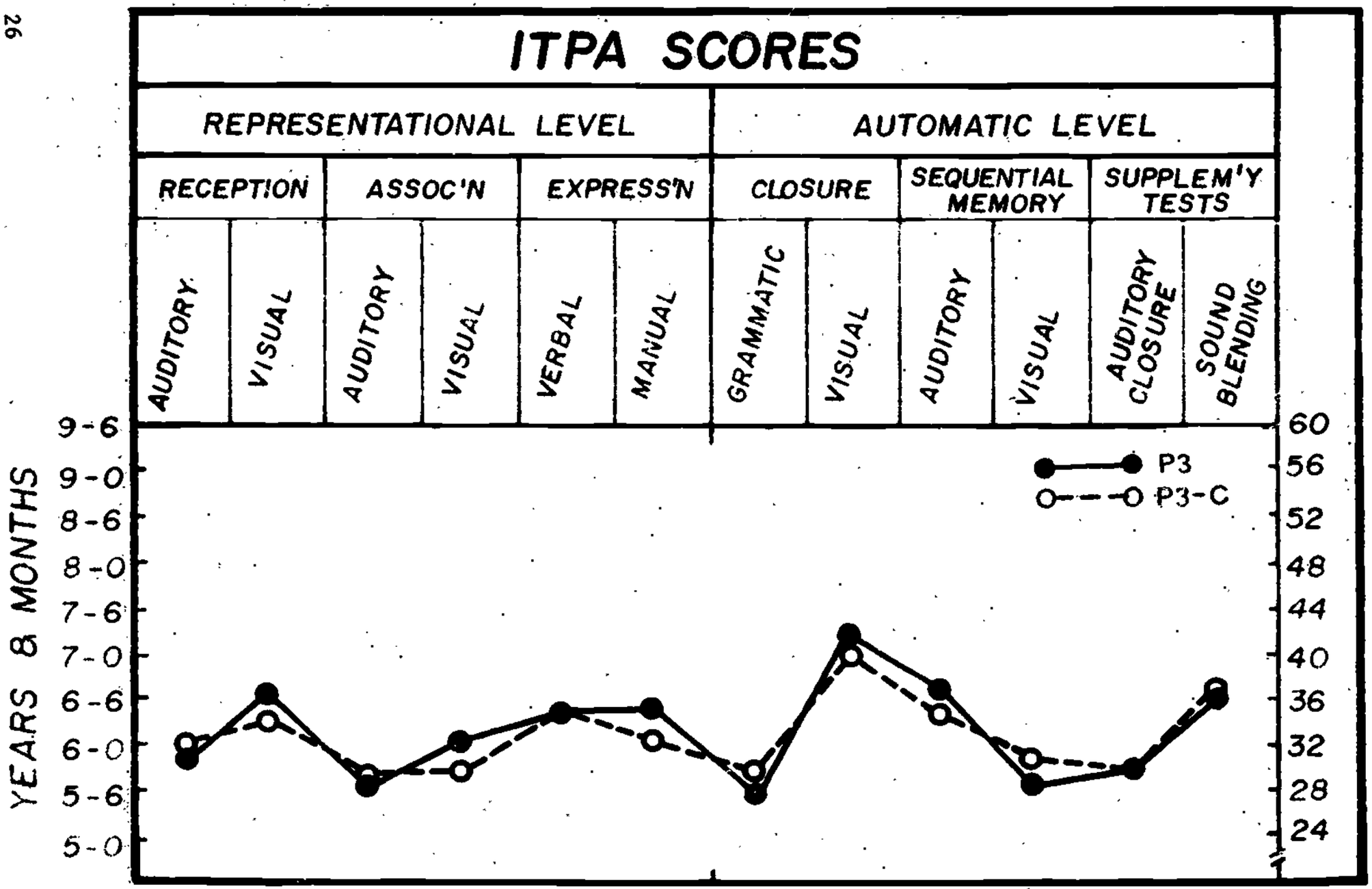

Fig. 6. A comparison of the means of the Peabody 3-month group and its control group on the 12 subtests of the ITPA.

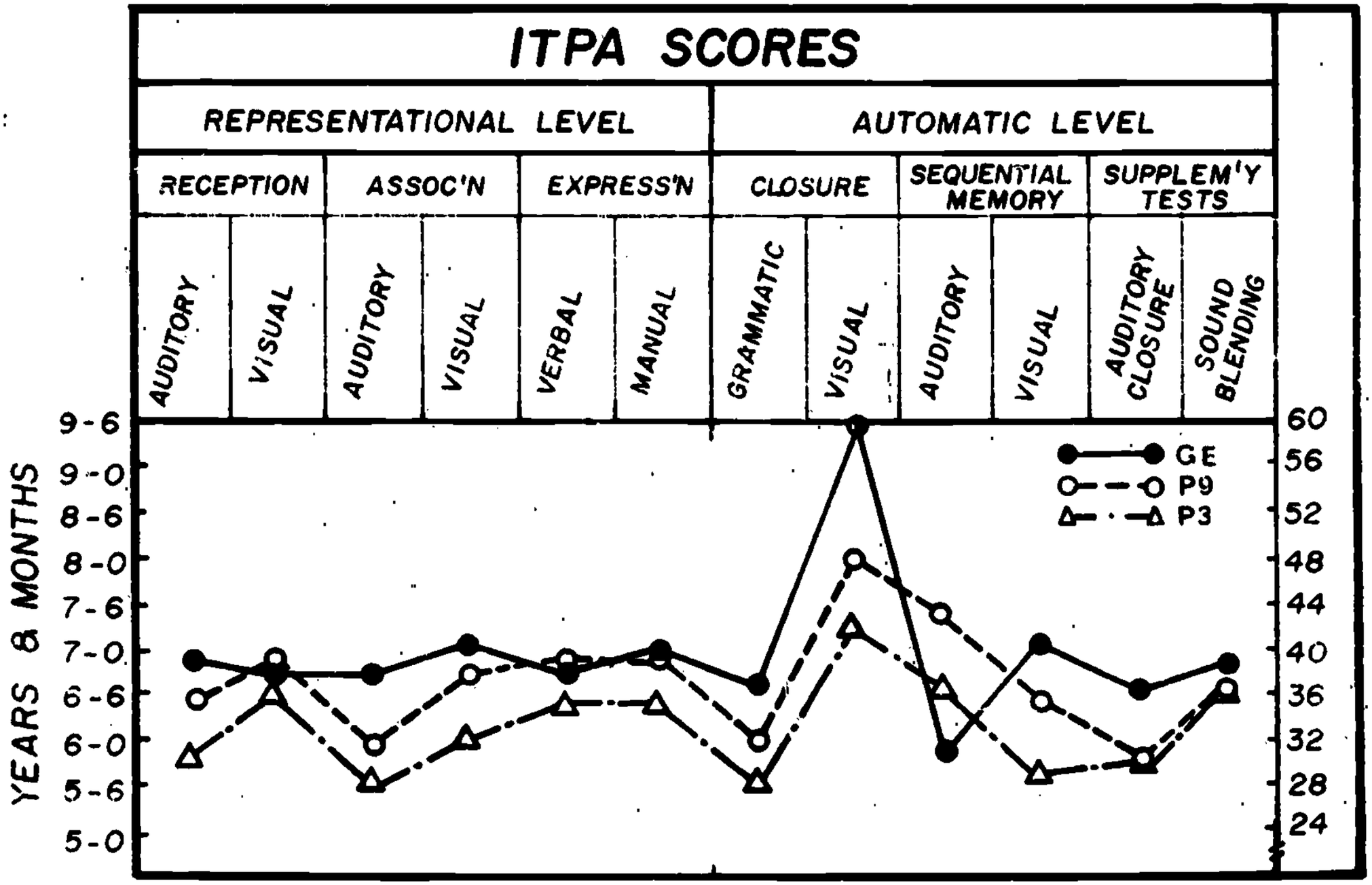

Fig. 7. Comparisons of the means of the General Enrichment, the Peabody 9-month, and the Peabody 3-month groups on the 12 subtests of the ITPA.

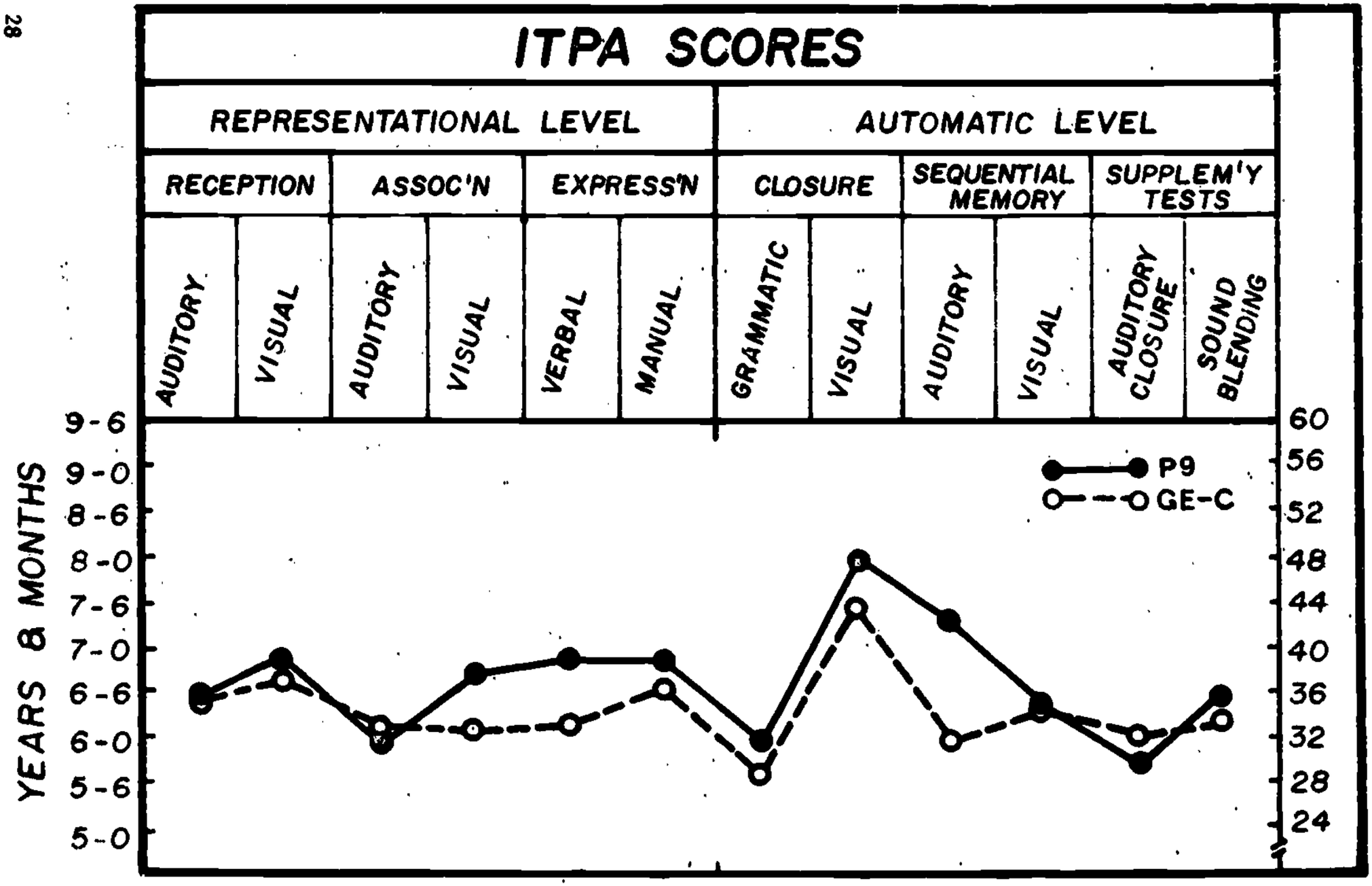

Fig. 8. A comparison of the means of the Peabody 9-month group and the General Enrichment control group on the 12 subtests of the ITPA.

The Auditory Reception subtest analysis revealed a main effect of curriculum. Table 4 shows that the GE group ($\bar{X}$ = 38.75) scored significantly higher than the P3 group ($\bar{X}$ = 30.75) and both Peabody control groups ($\bar{X}$'s: P3-C=32.00, P9-C=32.38). However, the GE group, the P9 group, and the GE-C group were not significantly different.

There were no significant differences between groups on the Visual Reception subtest.

The analysis of the Auditory Association subtest revealed significant differences between curricula groups. The Duncan's test in Table 5 indicates that the GE group ($\bar{X}$ = 37.00) scored significantly higher than both Peabody control groups ($\bar{X}$'s: P3-C=29.63, P9-C=28.25) and the P3 group ($\bar{X}$ = 28.88). However, the GE group did not deomonstrate better auditory association than the GE-C group or the P9 group.

The Visual Association analysis indicated a main effect of treatment. Table 6 shows that the GE group ($\bar{X}$ = 40.75) scored significantly higher than all other groups except for the P9 group ($\bar{X}$ = 38.00). The P9 group scored significantly higher than either Peabody control group ($\bar{X}$'s: P3-C=30.00, P9-C=30.75).

A main effect of treatment was indicated on the Verbal Expression subtest. As Table 7 shows, all curricula groups scored higher than all control groups. The only significant difference revealed by the Duncan's test was between the P9 group ($\bar{X}$ = 39.38) and its control group ($\bar{X}$ = 33.13).

On the scores of the Manual Expression subtest, the analysis of variance indicated a significant effect of treatment (Table 8). The GE group ($\bar{X}$ = 40.13) scored significantly higher than both Peabody control groups ($\bar{X}$'s: P3-C=32.88, P9-C=32.88).

There were no significant differences between treatment or curricula groups on the Grammatical Closure subtest.

The Visual Closure analysis showed main effects of both treatment and curriculum. The Duncan's indicated a significant difference between the GE group ($\bar{X}$ = 61.25) and all other groups (Table 9). There was also a significant difference between the P9 group ($\bar{X}$ = 48.13) and its control group ($\bar{X}$ = 39.25).

A main effect of curriculum was revealed by the analysis of the Auditory Sequential Memory subtest. The Duncan's in Table 10 shows that the P9 group ($\bar{X}$ = 42.50) scored significantly higher than any of the control groups or the GE group.

TABLE 4

Auditory Reception Scaled -
Second Administration

Duncan's New Multiple Range Test

of Differences Between Means

		P3	P3-C	P9-C	GE-C	P9	GE
Means		30.75	32.00	32.38	35.50	35.63	38.75
P3	30.75		1.25	1.63	4.75	4.88	8.00**
P3-C	32.00			.38	3.50	3.63	6.75*
P9-C	32.38				3.12	3.25	6.37*
GE-C	35.50					.13	3.25
P9	35.63						3.12

P3	P3-C	P9-C	GE-C	P9	GE

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

** $p < .01$

TABLE 5

Auditory Association Scaled -
Second Administration

Duncan's New Multiple Range Test

of Differences Between Means

Means		P9-C 28.25	P3 28.88	P3-C 29.63	P9 32.00	GE-C 32.88	GE 37.00
P9-C	28.25		.63	1.38	3.75	4.63	8.75**
P3	28.88			.75	3.12	4.00	8.12*
P3-C	29.63				2.37	3.25	7.37*
P9	32.00					.88	5.00
GE-C	32.88						4.12

P9-C	P3	P3-C	P9	GE-C	GE

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

** $p < .01$

TABLE 6

Visual Association Scaled -
Second Administration

Duncan's New Multiple Range Test

of Differences Between Means

Means		P3-C 30.00	P9-C 30.75	P3 32.75	GE-C 33.63	P9 38.00	GE 40.75
P3-C	30.00		.75	2.75	3.63	8.00**	10.75**
P9-C	30.75			2.00	2.88	7.25**	10.00**
P3	32.75				.88	5.25	8.00**
GE-C	33.63					4.37	7.12*
P9	38.00						2.75

P3-C P9-C P3 GE-C P9 GE

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

** $p < .01$

TABLE 7

Verbal Expression Scaled -
Second Administration

Duncan's New Multiple Range Test

of Differences Between Means

Means		P9-C 33.13	GE-C 33.50	P3-C 34.63	P3 35.25	GE 37.13	P9 39.38
P9-C	33.13		.37	1.50	2.12	4.00	6.25*
GE-C	33.50			1.13	1.75	3.63	5.88
P3-C	34.63				.62	2.50	4.75
P3	35.25					1.88	4.13
GE	37.13						2.25

P9-C GE-C P3-C P3 GE P9

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

TABLE 8

Manual Expression Scaled -
Second Administration

Duncan's New Multiple Range Test

of Differences Between Means

Means		P9-C 32.88	P3-C 32.88	P3 35.25	GE-C 36.63	P9 38.13	GE 40.13
P9-C	32.88		.00	2.37	3.75	5.25	7.25*
P3-C	32.88			2.37	3.75	5.25	7.25*
P3	35.25				1.38	2.88	4.88
GE-C	36.63					1.50	3.50
P9	38.13						2.00

P9-C P3-C P3 GE-C P9 GE

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

TABLE 9

Visual Closure Scaled -
Second Administration

Duncan's New Multiple Range Test

of Differences Between Means

Means		P9-C 39.25	P3-C 40.50	P3 42.37	GE-C 43.75	P9 48.13	GE 61.25
P9-C	39.25		1.25	3.12	4.50	8.88*	22.00**
P3-C	40.50			1.87	3.25	7.63	20.75**
P3	42.37				1.38	5.76	18.88**
GE-C	43.75					4.38	17.50**
P9	48.13						13.12**

P9-C P3-C P3 GE-C P9 GE

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

** $p < .01$

TABLE 10

Auditory Sequential Memory Scaled -
Second Administration

Duncan's New Multiple Range Test

of Differences Between Means

Means		GE 31.13	GE-C 31.75	P9-C 34.75	P3-C 35.38	P3 36.63	P9 42.50
GE	31.13		.62	3.62	4.25	5.50*	11.37**
GE-C	31.75			3.00	3.63	4.88	10.75**
P9-C	34.75				.63	1.88	7.75**
P3-C	35.38					1.25	7.12*
Pe	36.63						5.87

GE GE-C P9-C P3-C P3 P9

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

** $p < .01$

In the Visual Sequential Memory analysis effects of both treatment and curriculum were significant. The GE group ($\bar{X}$ = 40.13) scored significantly higher than both Peabody control groups and the P3 group (Table 11). There was no significant difference between the GE group and its control group or the P9 group.

Type of curriculum produced significantly different scores on the Auditory Closure subtest. The GE group ($\bar{X}$ = 35.88) scored significantly higher than all groups except its control group (Table 12).

There were no significant differences among the groups on the Sound Blending subtest.

To summarize the analysis of the ITPA:

1. The GE group scored significantly higher than its control group on:
   A. Visual Closure
   B. Visual Association
   C. Total Score

2. The P9 group scored significantly higher than its control group on:
   A. Visual Association
   B. Verbal Expression
   C. Visual Closure
   D. Auditory Sequential Memory
   E. Visual Sequential Memory

3. The P3 group did not score significantly higher than its control group.

4. The GE group scored significantly higher than the P9 treatment and control groups on:
   A. Visual Closure
   B. Auditory Closure
   C. Total Score

   The P9 group scored significantly higher than the GE group on Auditory Sequential Memory.

5. The GE group scored significantly higher than the P3 treatment and control groups on:
   A. Auditory Reception
   B. Auditory Association
   C. Visual Association
   D. Visual Closure
   E. Visual Sequential Memory
   F. Auditory Closure
   G. Total Score

6. There was no significant difference between the GE control children and the P9 group except for Auditory Sequential Memory, on which the P9 group scored significantly higher.

TABLE 11

Visual Sequential Memory Scaled -
Second Administration

Duncan's New Multiple Range Test

of Differences Between Means

		P9-C	P3	P3-C	GE-C	P9	GE
Means		27.63	28.75	31.00	34.50	35.38	40.13
P9-C	27.63		1.12	3.37	6.87*	7.75**	12.50**
P3	28.75			2.25	5.75*	6.63*	11.38**
P3-C	31.00				3.50	4.38	9.13**
GE-C	34.50					.88	5.63
P9	35.38						4.75

P9-C P3 P3-C GE-C P9 GE

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

** $p < .01$

TABLE 12

Auditory Closure Scaled -
Second Administration

Duncan's New Multiple Range Test

of Differences Between Means

Means		P3-C 29.63	P9-C 30.25	P3 30.25	P9 30.37	GE-C 32.88	GE 35.88
P3-C	29.63		.62	.62	.74	3.25	6.25*
P9-C	30.25			.00	.12	2.63	5.63*
P3	30.25				.12	2.63	5.63*
P9	30.37					2.51	5.51*
GE-C	32.88						3.00

P3-C P9-C P3 P9 GE-C GE

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

Factor Analysis of the ITPA

The raw scores of the 12 subtests of the Illinois Test of Psycholinguistic Abilities were subjected to a principle components analysis which was followed by a varimax rotation. A value of 1.0 was chosen to be the commonality estimate and was the value placed on the diagonals of the correlation matrix. The number of factors subjected to the varimax rotation consisted of those factors whose eigenvalues were equal to 1.0 or greater. This resulted in a three-factor solution to the problem. Factor 1 accounts for 47% of the variance and can be identified as a visual factor. Factor 2 raises the cumulative proportion of the total variance to 57% and can be identified as an auditory factor. The third factor again raises the cumulative proportion of the total variance to 66% and can be identified as a closure factor.

Table 13 gives the factor loadings as a result of the varimax rotation for the three-factor solution. This table contains only those factors whose loadings were .60 or greater as these represented the variables which were relatively independent of the other two factors. Loadings below this .60 value tended to be distributed among two and also among three factors.

TABLE 13

Factor Analysis of ITPA Raw Scores
(N=48)

Factor 1	Factor 2	Factor 3
.82 Vis. Assoc.	.87 Aud. Seq. Memory	.84 Aud. Closure
.73 Vis. Recept.	.72 Sound Blending	.68 Aud. Assoc.
.65 Manual Expr.		.63 Gramm. Closure
.60 Vis. Closure		.60 Vis. Seq. Mem.
Visual	Auditory	Closure

Caldwell Preschool Inventory

An analysis of variance of the total Caldwell score revealed main effects of treatment but not of curriculum. The mean of the experimental group was 53.88, thus exceeding the control group mean of 45.54. The Duncan's New Multiple Range Test was applied to the differences between the means of the total scores and is summarized in Table 14. The P9-C group ($\bar{X}$ = 41.88) scored significantly lower than the P9 group ($\bar{X}$ =56.63) and the GE group ($\bar{X}$ = 56.75). There were no significant differences between curricula groups on the total score.

TABLE 14

Total Caldwell - Second Administration

Duncan's New Multiple Range Test

of Differences Between Means

		P9-C	GE-C	P3	P3-C	P9	GE
Means		41.88	44.63	48.25	50.13	56.63	56.75
P9-C	41.88		2.75	6.37	8.25	14.75*	14.87*
GE-C	44.63			3.62	5.50	12.00	12.12
P3	48.25				1.88	8.38	8.50
P3-C	50.13					6.50	6.62
P9	56.63						.12

P9-C GE-C P3 P3-C P9 GE

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

Each of the four subtests of the Caldwell was analyzed individually. An analysis of Subtest 1, Personal-Social Responsiveness, indicated a main effect of treatment. A Duncan's Test is summarized in Table 15. The P9 group ($\bar{X}$ = 22.00) scored significantly higher than the P9-C group ($\bar{X}$ = 15.63) and the GE-C group ($\bar{X}$ = 16.63). The P3-C group ($\bar{X}$ = 19.63) scored significantly higher than the P9-C. There was no significant difference between the three curricula groups.

On Subtest 2, Associative Vocabulary, and Subtest 3, Concept Activation-Numerical, an analysis of variance revealed no main effects of treatment or curriculum.

Subtest 4 is Concept Activation-Sensory. An analysis revealed main effects of treatment but not curriculum. The mean for the treatment group was 13.46 as compared with a mean of 10.29 for the control group. The Duncan's test is summarized in Table 16. The P9 group ($\bar{X}$ = 14.25) and the GE group ($\bar{X}$ = 14.00) scored significantly higher than any of the control groups ($\bar{X}$'s: GE-C=10.25, P3-C=10.38, P9-C=10.25). There were no significant differences between the three curricula groups or between the three control groups. The GE and P9 curricula are seemingly effective in increasing sensory concept activation as measured by this subtest.

Englemann Concept Inventory Scale

There were no significant differences revealed on the analysis of variance of the total score of the Englemann test. The scores of Subtest 2 indicated that all the groups scored significantly higher than the P9-C group (Table 17). On Subtest 3, the GE group scored significantly higher than either of the Peabody control groups (Table 18). Generally speaking, all the children had a low rate of correct responses on this test; therefore, the test did not differentiate between groups.

Metropolitan Reading Readiness Test

The analysis of variance of the Metropolitan scores showed that there was a main effect of curriculum but not treatment. The results of the Duncan's test (Table 19) indicated that the P3 and P3-C groups scored significantly higher than the P9 group, the GE-C group, and the P9-C group. The mean of the GE group was below the means of the P3 and P3-C group and above the means of the other groups. It should be noted that all of the scores were very low, with a difference of only 6.12 points between the highest and lowest means.

TABLE 15

Caldwell - Second Administration
Subtest 1: Personal-Social Responsiveness

Duncan's New Multiple Range Test

of Differences Between Means

Means		P9-C 15.63	GE-C 16.63	GE 18.63	P3 19.50	P3-C 19.63	P9 22.00
P9-C	15.63		1.00	3.00	3.87*	4.00*	6.37**
GE-C	16.63			2.00	2.87	3.00	5.37**
GE	18.63				.87	1.00	3.37
P3	19.50					.13	2.50
P3-C	19.63						2.37

P9-C GE-C GE P3 P3-C P9

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

** $p < .01$

TABLE 16

Caldwell - Second Administration
Subtest 4: Concept Activation-Sensory

Duncan's New Multiple Range Test

of Differences Between Means

Means		P9-C 10.25	GE-C 10.25	P3-C 10.38	P3 12.13	GE 14.00	P9 14.25
P9-C	10.25		.00	.13	1.88	3.75*	4.00*
GE-C	10.25			.13	1.88	3.75*	4.00*
P3-C	10.38				1.75	3.62*	3.87*
P3	12.13					1.87	2.12
GE	14.00						.25

P9-C GE-C P3-C P3 GE P9

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

TABLE 17

Englemann - Subtest 2

Duncan's New Multiple Range Test

of Differences Between Means

Means		P9-C 13.75	P3-C 20.88	GE-C 21.50	P9 21.63	P3 22.75	GE 23.75
P9-C	13.75		7.13*	7.75*	7.88*	9.00**	10.00**
P3-C	20.88			.62	.75	1.87	2.87
GE-C	21.50				.13	1.25	2.25
P9	21.63					1.12	2.12
P3	22.75						1.00

	P9-C	P3-C	GE-C	P9	P3	GE

Any two means not underscored by the same line are siguificantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

** $p < .01$

TABLE 18

Englemann - Subtest 3

Duncan's New Multiple Range Test

of Differences Between Means

		P3-C	P9-C	GE-C	P9	P3	GE
Means		7.38	7.50	8.00	9.00	9.00	11.25
P3-C	7.38		.12	.62	1.62	1.62	3.87*
P9-C	7.50			.50	1.50	1.50	3.75*
GE-C	8.00				1.00	1.00	2.25
P9	9.00					1.00	2.25
P3	9.00						2.25

P3-C P9-C GE-C P9 P3 GE

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

TABLE 19

Total Metropolitan - First Administration

Duncan's New Multiple Range Test

of Differences Between Means

Means		GE-C 20.13	P9 21.50	P9-C 21.88	GE 22.63	P3-C 25.75	P3 26.25
GE-C	20.13		1.37	1.75	2.50	5.62*	6.12**
P9	21.50			.38	1.13	4.25*	4.75*
P9-C	21.88				.75	3.87*	4.37*
GE	22.63					*3.12	3.62
P3-C	25.75						.50

GE-C P9 P9-C GE P3-C P3

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

** $p < .01$

A main effect of curriculum was evident in the analysis of Subtest 1, the only subtest which revealed any significant differences. The Duncan's test (Table 20) indicated that both P3 and P3-C groups scored significantly higher than the GE, GE-C, and P9 groups. In addition, the P3 group scored significantly higher than the P9-C group.

This test proved to be inappropriate for our population. The children were not able to perform any of the test items if the test was given in groups with general instructions for each subject. For this reason, the test was administered to each child individually, and directions for each test item were given. Thus, the validity of the results is questionable.

TABLE 20

Metropolitan Reading Readiness Test - Subtest 1

Duncan's New Multiple Range Test

of Differences Between Means

		GE-C	P9	GE	P9-C	P3-C	P3
Means		4.88	5.25	6.00	6.38	8.25	9.63
GE-C	4.88		.37	1.12	1.50	3.37**	4.75**
P9	5.25			.75	1.13	3.00**	4.38**
GE	6.00				.38	2.25*	3.63**
P9-C	6.38					1.87	3.25**
P3-C	8.25						1.38

GE-C P9 GE P9-C P3-C P3

Any two means not underscored by the same line are significantly different.

Any two means underscored by the same line are not significantly different.

* $p < .05$

** $p < .01$

DISCUSSION

It will be helpful to precede the discussion of the external evaluation with some comments on the use of paraprofessionals, the Peabody materials, and the internal evaluation.

After three months of careful observation and feedback to our paraprofessional teachers, they were performing admirably. They quickly understood the principles of behavior modification and the importance of precise recording of a child's responses to particular tasks. We were lucky to have teachers who were bright, flexible, and appreciated constructive criticisms. It is probably more difficult to work with some professionally trained "traditional" early childhood educators who would actively resist the use of structured learning materials and behavior modification techniques. It was, however, very costly in terms of time for either observers or the project director to monitor the daily performance of the teachers and hold daily conferences with them concerning how they could improve their teaching skills. The Southeastern Educational Laboratory is presently using a sophisticated preservice training program for paraprofessional teachers rather than relying exclusively on an in-service training program.

In general, the Peabody materials accompanying Level #P possess two strengths: (1) they are very easy for paraprofessionals to use, and (2) the children found the lessons interesting. It should be recognized that we did not use the materials as they were designed -- i.e., a maximum of one lesson per day -- but covered as many lessons as possible each day for a concentrated teaching-learning session. This massed practice approach probably decreased the effectiveness of the lessons; obviously it would have been better, for example, to distribute the four hours of structured learning in group P9 across five days but the overall schedule of the Readimobile program made this impossible.

Our criticisms of the Peabody Level #P center around three issues -- (1) lesson objectives, (2) stimuli, and (3) organization of lessons. Since the lesson objectives are not made clear to the teacher it was necessary for us to isolate the specific lesson objectives or goals ourselves. The Southeastern Educational Lab is currently expanding the present recording system to include lesson objectives, a coding scheme, and a performance checklist. This approach will enable the teacher to keep accurate records herself on each child's progress through the Peabody lessons.

In order to devise a compact instructionsl "kit" the developers of the Peabody #P made some mistakes in the stimuli they selected. Only two examples are required to illustrate the problem. First, the same cards are used to teach color, size, and number rather than separate stimuli that would not be confounded on each of the other dimensions. Our children were very confused until we resorted to "homemade" stimuli. Second, the records that accompany the materials have a major fault -- the recordings are very brief and an individual song is recorded only once. Since numerous exposures were required for the children to learn the songs, the teacher had to leave the group frequently to reset the recording. It would have been far better if each song had been recorded about five times.

The organization of the Peabody has two "flaws" which could be easily corrected. First, we are not convinced that enough consideration (or research) has been given to the sequencing of the lessons. Second, a frequency count of the type of lessons (e.g., classification, following directiona, etc.) reveals that far more considerat on is given to some activities at the expense of others. In general, we would recommend more activities for each "goal" and a more equitable distribution of activities across "goals."

The internal evaluation was designed to establish specific instructional goals or objectives for each lesson and to record every child's verbal and nonverbal responses as related to a particular instructional goal. (A sample performance recording sheet is included in the appendix.) To accomplish this task, two observers were present each day for the P9 group. After satisfactory (r= .95) interobserver reliabilities were obtained, each observer recorded the verbal and nonverbal behavior of four children. These responses were coded as either correct or incorrect so that the teacher could tell after any lesson how well the children had mastered the inatructional goals. This careful recording and feedback to the teacher was probably one of the more valuable accomplishments of the project. If the children as a group did miserably, we would carefully examine the lesson or method of presentation and modify our approach. If an average child had not reached the criterion for successful performance, we could examine the lesson, method of presentation and/or repeat the lesson at a later date. Unfortunately, this evaluation does not approach the ideal of completely individualizing instruction; nevertheless, it assures that most of the children will succeed, and it guarantees that the teachers will have accurate records of the children's behavior.

Recognizing the previously mentioned problems of subject selection and experimental design, an examination of the results section dealing with the external evaluation reveals three primary findings and several secondary results. The primary results may be briefly summarized as follows: (1) the increase in test score performance of all groups on the second administration of three dependent measures; (2) the superiority of the P9 and GE experimental groups over their control groups on some of the dependent measures; and, (3) the effectiveness of the P9 curriculum in eliminating some of the well-documented differences between black lower class and white middle class children.

The increase in test score performance of all groups on the second administration of the Binet, ITPA and Caldwell can be interpreted as the result of increased familiarity with the instrument and examiner. Zigler and Butterfield (1968) have cautioned against possible erroneous interpretations of changes in IQ scores. It may be that many reported IQ changes in preschool programs could be most parsimoniously interpreted as motivational and attitudinal changes rather than substantive cognitive changes. These data then add a new note of caution in interpreting change scores when using the ITPA and Caldwell. The interaction of treatment group and administration reported on the Caldwell highlights another problem -- different groups may experience differential profit from repeated testing on particular instruments. Frank Palmer of the Harlem Research Center in New York has exercised caution in his testing procedures to guarantee a positive rapport between the child and the examiner before any assessment begins. Dr. Palmer discovered two very interest facts while working with two- and three-year-old children. First, it may take 10-15 hours of contact between the examiner and child before testing can begin. Second, the statistically significant correlation of a child's score on the Binet and the length of time before the child could be separated from his mother and testing could begin is high and negative. We should be concerned with controlling these noncognitive factors that may adversely affect a child's performance in a testing situation.

The P9 and GE experimental groups were superior to their control groups in several important instances. The P9 group was superior to its control group on all three of the major dependent measures -- Binet, ITPA, and Caldwell. Additionally, the P9 was superior to its control group on five subtests of the ITPA: Visual Association, Verbal Expression, Visual Closure, Auditory Sequential Memory, and Visual Sequential Memory.

The GE group was superior to its control group on the total ITPA scores, two ITPA subtests (Visual Closure and Visual Association) and one Caldwell subtest (Concept Activation-Sensory).

The lack of superiority of the P3 group on any of the dependent measures compared to its control group is easily explained by the brief exposure to an educational curriculum. The actual instructional time was only 48 hours (4 hours per week for 12 weeks) so it is not surprising that their performance did not improve. This group can probably best be viewed as a contact control group rather than as a meaningful treatment group.

The P9 group when compared to the GE or GE-C groups demonstrated the effectiveness of a structured preschool program in decreasing some of the well-documented differences between black and white children of different socio-economic classes. There were no statistically significant differences between the black lower socio-economic P9 children and the two white middle class groups, GE and GE-C, on the total scores of the Binet or Caldwell. On the total ITPA the P9 group was inferior to the GE group but not significantly different from the GE-C group which scored higher. than the P9-C, P3, and P3-C groups. Additionally, the GE group was superior to the P9-C, P3, and P3-C groups on the total scores of the Binet and ITPA. We take these facts as support for the effectiveness of the P9 preschool program.

Three secondary results merit a brief discussion. First, in general, there were no significant differences between or among the groups on two dependent measures -- the Englemann Concept Inventory and the Metropolitan. An examination of the first 90 lessons of the Peabody should have led us to the early conclusion that our curricula were not designed to improve performance on the Englemann and probably not on the Metropolitan. In addition to the content, the directions of the Metropolitan proved too difficult for our population. (Please don't ask the embarrassing question concerning why the P3 and P3-C groups did so well on the Metropolitan.) The original intention was to test each child twice on the Metropolitan but the examiners were convinced that the second administration was a waste of time and money.

A second interesting result was the high positive correlations among the Binet, ITPA, and Caldwell. Obviously, it would be needless duplication for anyone in the future to use the Binet and the Caldwell together as general measures in evaluating a preschool program; they are so highly correlated that knowledge of one score provides enough information. It is time for either instrument developers to come to the aid of preschool research or preschool research to adopt another approach to evaluation. We favor the latter suggestion.

Another set of potentially helpful results are those connected with the ITPA scores. The magnitude of our S's scores was high (e.g., the subtests scores converted to psycholinguistic age for the P9 group ranged from 5 years 6 months to 8 years). Either we are dealing with children who are precocious psycholinguistically or the norms for the 1968 ITPA manual are poor. We suspect the norms need improvement. The results of the factor analysis of the ITPA fell somewhere between the extremes of other research on the ITPA which finds only one factor for black Southern children (see Don Steadman's work from the Educational Improvement Program at Duke) or more factors than our three (studies which have usually used upper middle class suburban white children).

In concluding the discussion section, comments are in order about research needs in preschool education and approaches to evaluation of preschool programs. We strongly believe that research efforts which compare a treatment group and a distal control group using a pretest-posttest design are worthless. First, in a "successful" program, we don't know whether differences which appear are due to attitudinal and/or motivational changes rather than cognitive changes. Second, if these considerations are partially excluded by employing contact control groups or completely excluded using contact control groups and a Solomon Four group design, then it is impossible in these global intervention efforts to identify the antecedent conditions which produced the "success."

It appears that the most promising approach will be for investigators to concentrate on developing and evaluating "components" of an overall preschool "package." One might, for example, take a component like "classification skills" which appear embedded in almost all global preschool efforts (e.g., New Nursery, Deutsch, Weikart, etc.) and follow this simple approach: (1) identify the instructional goals (e.g., classify geometric stimuli according to number, size, and color); (2) develop a classification skills instructional program using Gagne's task analysis approach and insisting on criterion performance at every important step; and, (3) develop several pretests and posttests evaluative instruments. The pretests will be used for psychoeducational diagnosis to pinpoint the "entry" skills of each child and to aid in instruction. The posttests should contain several clusters of items: (1) an alternate form of the pretest designed to measure the terminal instructional objectives; (2) "near" transfer tests (problems which incorporate dimensions used in the instruction); (3) "far" transfer tests (problems which require the use of the same logical structure but have different specific content); and, (4) "farthest" transfer tests (problems presented in a different format and varying in content).

This "component" rather than "package" approach has several attractive features: (1) it guarantees an operational statement of the "input" (the Peabody materials are one of the few packages that state clearly what the preschool teacher is to do); (2) it provides for a careful, empirical evaluation of each component with instruments that accurately pinpoint a child's achievement before, during, and after instruction; and, (3) it provides the preschool teacher with the freedom to select components that are meaningful and important to her (ultimately, of course, research will identify the proper sequencing of components to attain a particular outcome). Components can be developed in numerous areas including, for example, number skills, perceptual and auditory discrimination, ordering, problem solving, and social skills. After seeing so many terrible "lessons" on "the family" or "Mommies" presented in preschool classes, we are convinced that someone must develop as many components as soon as possible if programs like Head Start are to really be more than socilization experiences for the participants. We have completed one component on multiple classification skills through all of the above steps; it is a very difficult, time consuming and costly approach. Its value, however, is that it is scientifically sound and may have a positive impact on preschool education.

APPENDIX A

TABLE 21

Analysis of Variance of Binet I.Q.

Source	df	Sums of Squares	Mean Square	F
Treatment (T)	1	2370.09	2370.09	12.4666**
Curriculum (C)	2	4020.40	2010.20	10.5736***
T X C	2	484.19	242.09	1.2734
Error	42	7984.81	190.11	
Administration (A)	1	356.51	356.51	24.0320***
T X A	1	1.76	1.76	.1186
C X A	2	11.52	5.76	.3883
T X C X A	2	71.65	35.82	2.4147
Error	42	623.06	14.83	

** $p < .01$
*** $p < .001$

TABLE 22

Analysis of Variance of Illinois Test
of Psycholinguistic Abilities Raw Score

Source	df	Sums of Squares	Mean Square	F
Treatment (T)	1	13848.01	13848.01	6.3663*
Curriculum (C)	2	35697.56	17848.78	8.2055**
T X C	2	2887.77	1443.89	.6637
Error	42	91358.31	2175.20	
Administration (A)	1	6550.51	6550.51	86.6438***
T X A	1	68.34	68.34	.9039
C X A	2	223.52	111.76	1.4782
T X C X A	2	321.81	160.91	2.1283
Error	42	3175.31	75.60	

* $p < .05$
** $p < .01$
*** $p < .001$

TABLE 23

Analysis of Variance of The
Caldwell Preschool Inventory

Source	df	Sums of Squares	Mean Square	F
Treatment (T)	1	1633.50	1633.50	6.4870*
Curriculum (C)	2	290.02	145.01	.5758
T X C	2	960.81	480.41	1.9078
Error	42	10576.00	251.81	
Administration (A)	1	504.17	504.17	25.3668***
T X A	1	.17	.17	.0083
C X A	2	138.52	69.26	3.4848*
T X C X A	2	23.40	11.70	.5885
Error	42	834.75	19.88	

* p < .05
** p < .01
*** p < .001

TABLE 24

Analysis of Variance of Total Stanford-Binet IQ Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	1250.521	1250.521	12.42**
Curriculum (C)	2	2173.042	1086.521	10.79**
T X C	2	309.042	154.521	1.53
Error	42	4229.875	100.71	
Total	47	7962.479		

** p < .01

TABLE 25

Analysis of Variance of Total Illinois Test of
Psycholinguistic Abilities Scaled Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	10384.083	10384.083	4.24*
Curriculum (C)	2	2930.125	1465.063	21.01**
T X C	2	555.292	277.646	.11
Error	42	102877.500	2449.464	
Total	47	216747.000		

* p < .05
** p < .01

TABLE 26

Analysis of Variance of
Visual Reception Scaled Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	184.083	184.083	8.57**
Curriculum (C)	2	12.667	6.334	.29
T X C	2	2.167	1.084	.05
Error	42	901.750	21.470	
Total	47	1100.667		

** $p < .01$

TABLE 27

Analysis of Variance of
Auditory Reception Scaled Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	46.021	46.021	1.87
Curriculum (C)	2	144.542	72.271	2.94
T X C	2	96.542	48.271	1.97
Error	42	1030.875	24.545	
Total	47	1317.979		

TABLE 28

Analysis of Variance of

Auditory Association Scaled Scores

(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	157.688	157.688	4.55*
Curriculum (C)	2	382.792	191.396	5.52**
T X C	2	40.875	20.438	.59
Error	42	1455.625	34.658	
Total	47	2036.979		

* $p < .05$

** $p < .01$

TABLE 29

Analysis of Variance of

Visual Association Scaled Scores

(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	52.083	52.083	1.13
Curriculum (C)	2	380.667	190.334	4.11*
T X C	2	11.167	5.584	.12
Error	42	1944.000	46.285	
Total	47	2387.917		

* $p < .05$

TABLE 30

Analysis of Variance of
Verbal Expression Scaled Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	82.688	82.688	2.48
Curriculum (C)	2	113.042	56.521	1.67
T X C	2	82.125	41.063	1.23
Error	42	1399.625	33.324	
Total	47	1677.479		

TABLE 31

Analysis of Variance of
Manual Expression Scaled Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	99.188	99.188	3.05
Curriculum (C)	2	214.542	107.271	3.30*
T X C	2	34.125	17.063	.53
Error	42	1364.625	32.491	
Total	47	1712.479		

* $p < .05$

TABLE 32

Analysis of Variance of
Grammatical Closure Scaled Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	143.521	143.521	4.37*
Curriculum (C)	2	121.542	60.271	1.85
T X C	2	47.042	23.521	.72
Error	42	1377.875	32.806	
Total	47	1689.979		

* $p < .05$

TABLE 33

Analysis of Variance of
Visual Closure Scaled Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	270.800	270.800	4.18*
Curriculum (C)	2	1559.592	779.796	12.05**
T X C	2	52.175	26.088	.40
Error	42	2717.200	64.695	
Total	47	4599.667		

* $p < .05$
** $p < .01$

TABLE 34

Analysis of Variance of
Auditory Sequential Memory Scaled Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	10.083	10.083	.29
Curriculum (C)	2	124.542	62.271	1.76
T X C	2	317.542	158.771	4.50*
Error	42	1483.500	35.321	
Total	47	1935.667		

* $p < .05$

TABLE 35

Analysis of Variance of
Visual Sequential Memory Scaled Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	192.000	192.000	6.17*
Curriculum (C)	2	429.125	214.563	6.90**
T X C	2	9.375	4.688	.15
Error	42	1306.500	31.107	
Total	47	1937.000		

* $p < .05$
** $p < .01$

TABLE 36

Analysis of Variance of
Auditory Closure Scaled Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	7.521	7.521	1.42
Curriculum (C)	2	273.500	136.750	2.42
T X C	2	51.167	25.584	2.42
Error	42	826.125	19.689	
Total	47	1158.313		

TABLE 37

Analysis of Variance of
Sound Blending Scaled Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	46.021	46.021	1.45
Curriculum (C)	2	30.542	15.271	.48
T X C	2	81.792	40.886	1.29
Error	42	1329.625	31.658	
Total	47	1487.979		

TABLE 38

Analysis of Variance of Total Caldwell Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	800.333	800.333	5.99*
Curriculum (C)	2	405.500	202.750	1.52
T X C	2	345.167	172.584	1.29
Error	42	5610.250	133.577	
Total	47	7161.250		

* p <.05

TABLE 39

Analysis of Variance of Caldwell--Subtest 1
Personal-Social Responsiveness
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	82.68750	82.68750	6.12365*
Curriculum (C)	2	.79167	.39583	.02931
T X C	2	21.37500	10.68750	.79149
Error	42	567.12500	13.50297	
Total	47	671.97917		

* p < .05

TABLE 40

Analysis of Variance of Caldwell--Subtest 2
Associative Vocabulary
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	126.75000	126.75000	7.80000**
Curriculum (C)	2	72.16667	36.08333	2.22051
T X C	2	144.50000	72.25000	4.44615
Error	42	682.50000	16.25000	
Total	47	1697.89584		

** p < .01

TABLE 41

Analysis of Variance of Caldwell--Subtest 3
Concept Activation- Numerical
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	52.08300	52.08300	.04723
Curriculum (C)	2	66.79167	33.39583	3.02861
T X C	2	14.04167	7.02083	.63671
Error	42	463.12500	11.02678	
Total	47	596.04134		

TABLE 42

Analysis of Variance of Caldwell--Subtest 4
Concept Activation- Sensory
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	93.52083	93.52083	6.88346*
Curriculum (C)	2	28.16667	14.08333	1.03658
T X C	2	22.16667	11.08333	.81577
Error	42	570.62500	13.58630	
Total	47	714.47917		

* p <.05

TABLE 43

Analysis of Variance of Total Englemann's and Bereiter's
Concept Development Scores
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	456.333	456.333	2.80
Curriculum (C)	2	21.125	10.563	.06
T X C	2	260.542	130.271	.80
Error	42	6843.250	162.934	
Total	47	7581.250		

TABLE 44

Analysis of Variance of Englemann's and Bereiter's Concept Inventory Scale--Subtest 1

Source	df	SS	MS	F
Treatment (T)	1	.02083	.02083	.00037
Curriculum (C)	2	235.54167	117.77083	2.08060
T X C	2	88.04167	44.02083	.77770
Error	42	2377.37500	56.60416	
Total	47	2700.97917		

TABLE 45

Analysis of Variance of Englemann's and Bereiter's Concept Inventory Scale--Subtest 2

Source	df	SS	MS	F
Treatment (T)	1	192.00000	192.00000	4.40355*
Curriculum (C)	2	224.29167	112.14583	2.57208
T X C	2	90.37500	45.18750	1.03638
Error	42	1831.25000	43.60119	
Total	47	2337.91667		

*p. <.05

TABLE 46

Analysis of Variance of Englemann's and Bereiter's Concept Inventory Scale--Subtest 3

Source	df	SS	MS	F
Treatment (T)	1	54.18750	54.18750	6.55164*
Curriculum (C)	2	21.12500	10.56250	1.27708
T X C	2	7.62500	3.81250	.46096
Error	42	347.37500	8.27083	
Total	47	430.31250		

* $p < .05$

TABLE 47

Analysis of Variance of Englemann's and Bereiter's Concept Inventory Scale--Subtest 4

Source	df	SS	MS	F
Treatment (T)	1	2.08333	2.08333	.35035
Curriculum (C)	2	18.87500	9.43750	1.58709
T X C	2	9.29167	4.64583	.78128
Error	42	249.75000	5.94642	
Total	47	280.00000		

TABLE 48

Analysis of Variance of
Total Metropolitan Reading Readiness Test
(First Administration)

Source	df	SS	MS	F
Treatment (T)	1	9.188	9.188	.86
Curriculum (C)	2	213.792	106.896	10.05**
T X C	2	17.375	8.688	.82
Error	42	446.625	10.634	
Total	47	686.979		

** p < .01

TABLE 49

Analysis of Variance of
Metropolitan Reading Readiness Test--Subtest 1

Source	df	SS	MS	F
Treatment (T)	1	2.52083	2.52083	.96579
Curriculum (C)	2	118.16667	59.08333	22.63626**
T X C	2	15.16667	7.58333	2.90536
Error	42	109.62500	2.61011	
Total	47	245.47917		

** p < .01

TABLE 50

Analysis of Variance of

Metropolitan Reading Readiness Test--Subtest 2

Source	df	SS	MS	F
Treatment (T)	1	7.52083	7.52083	1.94834
Curriculum (C)	2	15.04167	7.52083	1.94834
T X C	2	8.79167	4.39583	1.13878
Error	42	162.12500		
Total	47	193.47917		

TABLE 51

Analysis of Variance of

Metropolitan Reading Readiness Test--Subtest 3

Source	df	SS	MS	F
Treatment (T)	1	.75000	.75000	.15889
Curriculum (C)	2	23.79167	11.89583	2.52018
T X C	2	9.87500	4.93750	1.04603
Error	42	198.25000	4.72023	
Total	47	232.66667		

TABLE 52

Analysis of Variance of

Metropolitan Reading Readiness Test--Subtest 4

Source	df	SS	MS	F
Treatment (T)	1	.18750	.18750	.12475
Curriculum (C)	2	9.29167	4.64583	3.09109
T X C	2	.87500	.43750	.29109
Error	42	63.12500	1.50297	
Total	47	73.47917		

TABLE 53

Analysis of Variance of

Metropolitan Reading Readiness Test--Subtest 5

Source	df	SS	MS	F
Treatment (T)	1	.02083	.02083	.15556
Curriculum (C)	2	.54167	.27083	2.02222
T X C	2	.29167	.14583	1.08889
Error	42	5.62500	.13392	
Total	47	6.47917		

TABLE 54

Analysis of Variance of Total Stanford-Binet IQ Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	1121.333	1121.333	10.76**
Curriculum (C)	2	1858.875	929.438	8.92**
T X C	2	246.792	123.396	1.18
Error	42	4378.000	104.238	
Error	47	7605.000		

** $p < .01$

TABLE 55

Analysis of Variance of Total Illinois Test of
Psycholinguistic Abilities Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	16023.521	16023.521	6.60*
Curriculum (C)	2	15344.042	7672.021	23.76**
T X C	2	5845.292	2922.646	1.20
Error	42	101932.625	2426.967	
Total	47	239145.479		

* $p < .05$

** $p < .01$

TABLE 56

Analysis of Variance of
Auditory Reception Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	36.750	36.750	1.59
Curriculum (C)	2	265.167	132.583	5.72**
T X C	2	54.000	27.000	1.17
Error	42	972.750	28.269	
Total	47	1328.667		

** p < .01

TABLE 57

Analysis of Variance of
Visual Reception Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	96.333	96.333	3.22
Curriculum (C)	2	27.125	13.562	.45
T X C	2	53.792	26.896	.90
Error	42	1254.750		
Total	47	1432.000		

TABLE 58

Analysis of Variance of
Auditory Association Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	67.687	67.687	1.72
Curriculum (C)	2	300.125	150.063	3.82*
T X C	2	58.875	29.438	.75
Error	42	1651.125	39.313	
Total	47	2077.812		

* $p < .05$

TABLE 59

Analysis of Variance of
Visual Association Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	391.021	391.021	7.87**
Curriculum (C)	2	270.375	135.188	2.72
T X C	2	52.542	26.271	.53
Error	42	2086.375		
Total	47	2800.313		

** $p < .01$

TABLE 60

Analysis of Variance of
Verbal Expression Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	147.000	147.000	5.40*
Curriculum (C)	2	14.625	7.313	.27
T X C	2	63.375	21.688	1.16
Error	42	1143.000	32.571	
Total	47	1368.000		

* $p < .05$

TABLE 61

Analysis of Variance of
Manual Expression Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	165.021	165.021	6.63*
Curriculum (C)	2	154.292	77.146	3.10
T X C	2	16.792	8.396	.34
Error	42	1044.875	32.880	
Total	47	1380.979		

* $p < .05$

TABLE 62

Analysis of Variance of
Grammatical Closure Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	120.383	120.383	3.97
Curriculum (C)	2	151.217	75.608	2.49
T X C	2	155.217	77.608	2.56
Error	42	1273.200	30.314	
Total	47	1699.917		

TABLE 63

Analysis of Variance of
Visual Closure Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	1064.133	1064.133	16.12**
Curriculum (C)	2	1093.925	546.863	8.28**
T X C	2	490.092	245.046	3.71*
Error	42	2773.200	66.029	
Total	47	5421.250		

* $p < .05$

** $p < .01$

TABLE 64

Analysis of Variance of
Auditory Sequential Memory Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	93.521	93.521	3.26
Curriculum (C)	2	423.292	211.646	7.39**
T X C	2	154.542	77.271	2.70
Error	42	1203.625	28.658	
Total	47	1874.979		

** $p < .01$

TABLE 65

Analysis of Variance of
Visual Sequential Memory Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	165.021	165.021	5.78*
Curriculum (C)	2	489.292	244.646	8.56**
T X C	2	222.042	111.021	2.89
Error	42	1198.125	28.527	
Total	47	2074.479		

* $p < .05$
** $p < .01$

TABLE 66

Analysis of Variance of
Auditory Closure Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	18.750	18.750	.86
Curriculum (C)	2	193.792	96.896	4.46*
T X C	2	18.875	9.437	.43
Error	42	912.500	21.726	
Total	47	1143.917		

* $p < .05$

TABLE 67

Analysis of Variance of
Sound Blending Scaled Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	22.688	22.688	.63
Curriculum (C)	2	12.875	6.438	.18
T X C	2	45.125	22.563	.63
Error	42	1508.125	35.908	
Total	47	1588.813		

TABLE 68

Analysis of Variance of Total Caldwell Scores
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	833.333	833.333	6.03*
Curriculum (C)	2	23.042	11.521	.08
T X C	2	639.042	319.521	2.31
Error	42	5800.500	138.107	
Total	47	7295.917		

* $p < .05$

TABLE 69

Analysis of Variance of Caldwell--Subtest 1
Personal-Social Responsiveness
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	90.75000	90.75000	6.91115*
Curriculum (C)	2	30.54167	15.27083	1.16296
T X C	2	87.87500	43.93750	3.34610
Error	42	551.50000	13.13095	
Total	47	760.66217		

* $p < .05$

TABLE 70

Analysis of Variance of Caldwell--Subtest 2
Associative Vocabulary
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	46.02083	46.02083	3.75772
Curriculum (C)	2	1.12500	.56250	.04593
T X C	2	78.79167	39.39583	3.21677
Error	42	514.37500	12.24702	
Total	47	640.31250		

TABLE 71

Analysis of Variance of Caldwell--Subtest 3
Concept Activation- Numerical
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	1.68750	1.68750	.15676
Curriculum (C)	2	63.87500	31.93750	2.96682
T X C	2	52.12500	26.06250	2.42107
Error	42	452.12500	10.76488	
Total	47	569.81250		

TABLE 72

Analysis of Variance of Caldwell--Subtest 4
Concept Activation- Sensory
(Second Administration)

Source	df	SS	MS	F
Treatment (T)	1	120.33333	120.33333	12.17098**
Curriculum (C)	2	9.50000	4.75000	.48043
T X C	2	12.16667	6.08333	.61529
Error	42	415.25000	9.88690	
Total	47	557.25000		

** $p < .01$

APPENDIX B

TABLE 73

Correlations of The Subtests of The
Illinois Test of Psycholinguistic Abilities
(Second Administration - Scaled Scores)

	Aud. Rec.	Vis. Rec.	Vis.Seq. Mem.	Aud. Assoc.	Aud.Seq. Mem.	Vis. Assoc.	Vis. Clos.	Verbal Exp.	Gram. Clos.	Man. Exp.	Aud. Clos.	Sound Blendg.	Total
Auditory Reception	1.0000												
Visual Reception	.5118	1.0000											
Visual Sequential Memory	.5360	.3037	1.0000										
Auditory Association	.6972	.4786	.5988	1.0000									
Auditory Sequential Memory	.2447	.3158	.1545	.2278	1.0000								
Visual Association	.6213	.5835	.5591	.5760	.3355	1.0000							
Visual Closure	.5209	.2889	.6675	.5654	.1704	.5836	1.0000						
Verbal Expression	.4643	.2879	.3446	.3829	.4636	.5015	.5769	1.0000					
Grammatical Closure	.6399	.5313	.5877	.8070	.3164	.6448	.6451	.4590	1.0000				
Manual Expression	.6055	.5486	.4643	.5699	.2358	.5570	.5281	.5242	.5375	1.0000			
Auditory Closure	.6730	.2680	.5380	.6464	.0763	.4357	.6065	.4397	.5873	.5590	1.0000		
Sound Blending	.4167	.3497	.3259	.4102	.2382	.3608	.4657	.2750	.4872	.3713	.4366	1.0000	
Total	.7513	.4834	.7443	.7711	.1008	.6709	.7862	.4326	.7526	.6547	.7595	.3938	1.0000

$p < .05$ if $r > .288$
$P < .01$ if $r > .372$

TABLE 74

Correlations of the Subtests of The
Caldwell Preschool Inventory

(Second Administration)

	Per.-Soc. Resp.	Assoc. Vocab.	Con.Act.-Num.	Con.Act.-Sens.
Personal-Social Responsiveness	1.0000			
Associative Vocabulary	.6591	1.0000		
Concept Activation-Numerical	.3721	.6249	1.0000	
Concept Activation-Sensory	.7081	.6594	.5441	1.0000

$p < .05$ if $r \geq .288$
$p < .01$ if $r \geq .372$

APPENDIX C

PERFORMANCE RECORDING SHEET
WAKULLA COUNTY READIMOBILE PRESCHOOL

DATE_____ TEACHER_____
OBSERVER_____

1. LESSON NO. _____ TITLE _____

Specific lesson objective(s):
A.
B.
C.
D.

2. Lesson adequacy: Satisfactory _____ Unsatisfactory_____
Specific suggestions for improvement:

3. Teacher presentation: Satisfactory _____ Unsatisfactory_____
Specific suggestions for improvement:

4. Children's Responses
Code for each response:
Correct Verbalization = V Correct Nonverbal = /
Incorrect Verbalization = V Incorrect Nonverbal = X

Code for overall lesson evaluation:
Child comprehends lesson objective = √
Some comprehension, needs additional work = ?
Very little comp., needs "branching" = 0

Specific Lesson Objectives

Child's Name	A	B	C	D

5. Additional Comments:

REFERENCES

Alpern, G.D. The failure of a nursery school enrichment program for culturally disadvantaged children. Paper presented at the regional meeting of American Association of Psychiatric Clinics for Children, Cincinnati, Ohio, 1966.

American Guidance Service. Peabody Language Development Kits (Levels #P, #1, #2, #3), Publishers' Building, Circle Pines, Minnesota.

Bereiter, C. & Englemann, S. Teaching disadvantaged preschool children. Englewood Cliffs, N.J.: Prentice Hall, 1966.

Berko, J. The child's learning of English morphology. Word, 1958, 44, 150-177.

Deutsch, M., Katz, I., & Jensen, A. Social class, race and psychological development. New York: Holt, Reinhart, & Winston, 1968.

Elkind, D. Cognition in infancy and early childhood. In Y. Brackbill (Ed.), Infancy and early childhood. New York: Free Press, 1967. Pp. 361-394.

Ellis, N.R. (Ed.) Handbook of mental deficiency. New York: McGraw-Hill, 1963.

ERIE, Eastern Regional Institute for Education, Progress and Plans, Syracuse, N.Y., 1968.

Glick, J. Some problems in the evaluation of preschool intervention programs. In R. Hess & R. Bear (Eds.), Early education. New York: Adline Publishing Co., 1968. Pp. 215-221.

Gordon, I. Early child stimulation through parent education. Research proposal to Children's Bureau, February, 1967.

Gray, S.W., & Klaus, R.A. An experimental preschool program for culturally deprived children. Child Development, 1965, 36, 887-898.

Guilford, J.P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Hansen, D. Computer-assisted instruction. Review of Educational Research, 1966, 36, 588-603.

Helmuth, J. (Ed.) Disadvantaged child. Seattle: Special Child Publications, 1967.

Hess, R., & Bear, R. (Eds.) Early education. New York: Adline Publishing Co., 1968.

Hunt, J. McV. Intelligence and experience. New York: Ronald Press, 1961.

Karnes, Merle B., Kirk, Samuel A., Bereiter, C., & Englemann, S. A research program to determine the effects of various preschool intervention programs on the development of disadvantaged children and the strategic age for such intervention. Paper presented at the meeting of the American Educational Research Association, Chicago, February, 1968.

Kirk, S.A., & McCarthy, J.J. The Illinois Test of Psycholinguistic Abilities - an approach to differential diagnosis. American Journal of Mental Deficiency, 1961, 66, 399-412.

Madsen, Charles H., & Madsen, C.K. Discipline/teaching. Boston: Allyn & Bacon, 1969.

McGeoch, J.A., & Irion, A.L. The psychology of human learning. New York: McKay, 1952.

Morris, S.K. Results of a study using the Peabody Language Development Kit: Level P, Experimental Edition. Unpublished master's thesis, Vanderbilt University, 1967.

Osgood, C.E., Suci, G.J. & Tannenbaum. The measurement of meaning. Urbana: University of Illinois Press, 1957.

Painter, G. Home tutorial program for 3 year old children. Unpublished doctoral dissertation, University of Illinois, 1968.

Parker, R.K., & Dunham, R.H. Project Know How: a comprehensive and innovative attack on individual familial poverty. In R.K. Parker (Ed.), Readings in educational psychology. Boston: Allyn and Bacon, 1968. Pp. 385-389.

Reidford, P. Recent developments in preschool education. In D. Brison & J. Hill (Eds.), Psychology and early childhood education, Ontario Institute for Studies in Education, Monograph Series No. 4, 1968. Pp. 5-16.

Schaefer, E. Intellectual stimulation of culturally deprived infants. Proposal to NIMH, No. MH-09224 -01, 1965.

Skinner, B.F. Verbal behavior. New York: Appleton-Century-Crofts, 1957.

Sprigle, H.A., Van de Riet, V. & Van de Riet, Hani. A sequential learning program for preschool children and an evaluation of its effectiveness with culturally disadvantaged children. Paper presented at the American Educational Research Association, New York, 1967.

Swift, Joan W. Effects of early group experiences: the nursery school and day nursery. In M. Hoffman & L. Hoffman (Eds.), Review of child development research, Vol. 1, New York: Russel Sage Foundation, 1964. Pp. 249-288.

Toothman, R. The readimobile program, Southeastern Educational Corp., Atlanta, Georgia, 1968.

Torrance, E.P. Guiding creative talent. Englewood Cliffs, N.J.: Prentice-Hall, 1962.

Vergason, G.A. Retention in retarded and normal subjects as a function of amount of original training. American Journal of Mental Deficiency, 1964, 68, 623-629.

Webster, S.W. (Ed.) The disadvantaged learner. San Francisco: Chandler Publishing Co., 1966.

White, S.H. Some educated guesses about cognitive development in the preschool years. In R. Hess and R. Bear (Eds.), Early education. Chicago: Aldine Publishing Co., 1968. Pp. 203-214.

Zigler, E., & Butterfield, E.C. Motivational aspects of changes in I.Q. tests performance of culturally deprived nursery school children. Child Development, 1968, 39, 1-14.