DOCUMENT RESUME

ED 140 527 EC 100 865

AUTHOR Leiss, Robert H.; Proger, Barton B.
TITLE Language Training for Trainable Mentally Retarded Children: ITPA, Peabody, and Distar Techniques.
PUB DATE Apr 77
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, New York, April 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Exceptional Child Research; *Language Development; *Language Programs; Language Skills; Learning Activities; Mentally Handicapped; *Teaching Methods; *Trainable Mentally Handicapped
IDENTIFIERS *Distar Language I; *Peabody Language Development Kit

ABSTRACT

To determine the most effective language training activities for trainable mentally retarded (TMR) children, the variables of degree of previous language training, IQ, Peabody language treatment program versus Distar language treatment program, pretest versus posttest, and sex were examined with 122 TMR Ss (7 to 14 years old). Results of the Peabody Picture Vocabulary Test, the Illinois Test of Psycholinguistic Abilities, and the Mechann Verbal Language Development Scale indicated that Ss were significantly better aided by the Distar program than by the Peabody program, and that only the children previously exposed to language training showed any significant growth in language. A second study examined in depth the nature of previous language experience on Distar performance. (IM)

LANGUAGE TRAINING FOR TRAINABLE MENTALLY RETARDED CHILDREN: ITPA, PEABODY, AND DISTAR TECHNIQUES

Robert H. Leiss and Barton B. Proger

Montgomery County Intermediate Unit No. 23

Norristown, Pennsylvania

Paper Presented at Annual Meeting of

American Educational Research Association,

New York City, April 7, 1977

INTRODUCTION

This research project was designed to investigate the efficacy of providing intensified language stimulation programs to the trainable mentally retarded. Since there is a frequently reported and often observable deficiency in the language skills of this particular population of children, it would appear to be of significance to determine what effect, if any, a concerted program of language stimulation will have. The outcomes of this experiment have a direct bearing upon the caseload composition of speech and language clinicians in the public schools, have implications for the types of testing and assessment procedures utilized with the trainable mentally retarded, and may definitely influence the degree and manner in which language services are provided to these children. In effect, the findings of the study can suggest whether the trainable mentally retarded can benefit from intensified language stimulation, and the amount of language stimulation required to achieve a significant degree of improvement.

Improvements in language were assessed on sensitive instruments and a sub-population of the subjects in this project provided information concerning specific amounts of improvement in language abilities as a result of therapy sessions. This information is of value in providing for the evidence of the effectiveness of intensified language programs with this type of children. Such information is of value to other researchers and experimenters and allows for the development of further important research hypotheses.

PAST RESEARCH

Despite the fact that a large percentage of the trainable mentally retarded children are enrolled in special classes in the schools and in spite of the research evidence which has consistently reported such children to have speech and language problems (Bangs, 1961; Brandfon, 1951; Daum, 1953; Donovan, 1957; Everhart, 1953; Gens, 1950, 1951; Goodwin, 1955; Gottsleben, 1955; Harrison, 1958; Irvin, 1942; Karlin and Kennedy, 1936; Karlin and Strazzula, 1952; Kennedy, 1930; Kolstoe, 1958; Lewald, 1932; Lubman, 1955; Lyle, 1960; Masket, 1958; Matthews, 1957; Meader, 1940; Sachs, 1955; Schiefelbusch, 1963; Schlanger, 1953b, 1953c; Schlanger and Gottsleben, 1957; Schneider and Vallon, 1954; Sheridan, 1948; Sirkin and Lyons, 1941; Tarjan, et. al., 1961; Town, 1913; Wood, 1957; Wolfensberger, et. al., 1963) there exists a paucity of important research with regards to the efficacy of speech and language programs with the trainable mentally retarded.

Among children in special classes, Matthews (1957) estimated an incidence of speech problems of 79 percent. Lubman (1955) studied subjects with IQs below 50 and noted that 95 percent had speech defects. Johnson et. al., (1960) reported an incidence of about 57 percent in a study of trainable mentally retarded children. Wood (1957) noted about 21 percent of a sample studied at a speech and hearing center to have language deficiencies associated with mental retardation. This does not, however, indicate any estimate of the number of mentally retarded who have language problems.

The estimates of the incidence of language deficiencies among the mentally retarded varies from less than 10 percent to almost 80 percent. This variance

is primarily due to the differences in the groups studied and the definitions of what constitutes a language problem.

One of the major theoretical questions is whether lack of language development among mentally retarded children is an inevitable consequence of mental retardation or whether intensive training can improve the rate of language development. The studies of language training programs for the retarded are few. Since 1955 therapy with the mentally retarded has emphasized more than articulatory proficiency; it has demonstrated the necessity for providing appropriate language development programs.

Schneider and Vallon (1954) emphasize the necessity for therapy with the severely retarded and challenge the view of West, Kennedy, and Carr (1947), who thought that therapy with the severely retarded was useless, as being too pessimistic. They state that the simple ability to express the wants or needs of oneself in a socially approved manner, along with the ability to merely express one's wants or needs, is an undeniable asset to the child intellectually, emotionally, and socially.

In 1955, Schneider and Vallon reported on a therapy program for trainable retarded children in a day school class. The children were categorized into three groups: (1) Delayed language development, (2) Insufficient language development, and (3) Disturbances of articulation. Appropriate therapy activities were presented to each group for one year. The resultant data revealed gains for all groups. These judgments were, however, subjective, and no control group had been used.

Johnson and Capobianco (1957) studied a group of severely retarded children following a year of language training; they reported no significant improvement. This study was noteworthy as one of the first experimental assessments of a

language program for the retarded in which the results were contradictory to preceding reports.

Kilstoe (1958) observed the effect of a language training program with a small group of mongoloid children. On five subtests of the Illinois Language Scale, the experimental group gained significantly over the controls during a five and one-half month's period. Rittmanic (1958) set up a pilot program in group oral language with institutionalized retardates. Despite the lack of statistical evidence, he claimed that the program was successful.

Smith (1962) conducted a language program for sixteen educable retarded children; he assessed the progress by using the ITPA. The experimental group showed a 6.75 month gain in Language Age during a three-month's period; the controls declined .4 months in Language Age. Smith did not attempt to remediate any specific disabilities. Improvement was, however, noted on all the language abilities as measured by the ITPA. Blue (1963) supervised a language program for trainable retardates similar to the previously described program by Smith. The program was conducted for an eleven-week period and utilized the ITPA for pre- and post- measurement. The experimental group showed a Language Age gain of 5.67 months as compared to the control group's 3.67 months. The difference was not statistically significant. This is considered one of the more prominent studies on the efficacy of language therapy for trainable retardates.

Blessing (1964) reported on an experimental program which was designed to improve the vocal encoding of mentally retarded children. After a period of three-month's training the ITPA was used to note progress. The results revealed only a tendency toward improvement by the experimental group.

Harvey, Yep, and Sellin (1966) reported on a two-year program for trainable mentally retarded children. Their program emphasized the areas of: (1) self-

concept development, (2) Social competence, (3) Motor coordination, and (4) Language development. Their results indicated highly significant improvements in the four areas. All scores, with the exception of social competence, declined over the summer of the first year. This was interpreted to mean: (1) that there are differences between home and school environments, and (2) it is essential to maintain minimal programs during the summer for these children. The second year revealed significant increases in all areas. They concluded that evaluation of programs should be allowed to occur over longer periods of time, particularly with individuals with low IQs.

Richardson (1967) describes a language training program for retarded children at the University of Oklahoma Child Study Center. It indicates that early sensory-motor training, beginning at the pre-verbal experience level is of utmost importance to the language development of these children. Methods used in the program are related to research evidence on the development of language and thinking which indicates that: (1) Early exposure to a variety of looking and listening experiences is important in language development, (2) Primary learning requires perceptual and pre-verbal experiences, (3) There is a close relationship between motor movements and perceptual development, (4) Language development requires the development of both motor and perceptual patterns, (5) The major source of internal mediators is the orienting response, (6) Linguistic labels serve to mediate learning processes, and (7) Language development is both a part of and a result of primary learning.

Jordan (1967) reports that speech therapy outcome studies with the mentally retarded reveal that special psycholinguistic instruction can significantly increase psycholinguistic attainment. He suggests that programmed learning and operant conditioning be utilized to teach language to the mentally retarded.

Potter and Mattson (1968) also indicate that the educable mentally retarded are capable of manifesting and sustaining improvement in speech and language performance after therapy. Ensminger and Smith (1965) state, "knowing that specific language skills can be improved and that retardates display a rather distinctive profile of their own, group language programs should be developed with this pattern of abilities and disabilities as the focal point." (p. 104).

Early attempts at therapy for language disabilities were reported with optimism, but were not objectively evaluated. Encouraging progress has been reported with the educable retarded; the trainable child, however, presents some difficulty. Since many of the children involved in these studies were institutionalized and since the size of the group was limited, it becomes difficult to generalize from these findings to the population of trainable mentally retarded children who are enrolled in special classes in the public schools.

A factor of possible significance which may serve to influence the results of research concerning the effectiveness of language stimulation for trainable mentally retarded children may be the amount of treatment which is provided. There is a lack of research information indicating, for example, how many periods of language training are necessary during the period of a year in order for such children to achieve significant improvements in language.

TREATMENTS

First Year: The set of 8 classes which received no special language training went through their typical daily routines. However, the 8 classes which received language stimulation 4 times a week and the 8 classes which received the stimulation 8 times a week, were given specially structured lessons. The lessons were based primarily on the tasks embedded among the twelve subtests of the Illinois Test of Psycholinguistic Abilities (ITPA). There were 8 lessons for each subtest, or a total of 96 lessons. Each lesson lasted about 25 minutes. The training was spread out from the start of November to about the middle of May. Four experienced speech clinicians carried out the program. An important point is that the 8 lessons associated with any given ITPA subtest were not sequenced one after the other; rather, the set of lessons were distributed over the course of the total program so that the children would have more of a chance to retain and put into practice whatever benefits they had received.

Second Year: Each child received language training stimulation 4 times a week. A total of 96 different lessons were available for either the Distar Language Program (Level 1) or the Peabody Language Development Kit (Levels P and 1). A day's session in either program lasted 25 minutes. Both language training programs lasted from the beginning of November to about the middle of May. Certificated speech clinicians carried out the program.

Third Year: During the third year of the project, each TMR child was assigned to one of three program offerings. The assignments were made not on a random basis, but instead upon the basis of what degree of language facility each child had. In this manner, the treatments were assigned in a rational, real-life fashion. It must be remembered that the primary goal was the appropriate education of the children, not basic research. However, in spite of what might be termed

various types of self-selection biases at work (related to language facility, which might be termed a developmental/organismic variable, as opposed to sex, which would be the classical type of organismic variable), the reader will see later in the design itself that the final data actually analyzed arose from representative, randomly selected children.

The three treatments consisted of the Peabody Program (Level P), the Distar Program, or a combination of the two. Those children who were essentially non-verbal received the Peabody Level P offering. Those children who had some language facility (knowledge of some words and phrases) were given the combination of Peabody Level P and Distar. Finally, if a child could use at least short sentences, he was given the regular Distar program.

The third year was devoted to implementing the Peabody Language Development Kit program (Level P) with the young, non-verbal trainable mentally retarded children, using this same program with a more structured format as the children gain verbal ability at the word level. The Distar program will be used for those children who have attained phrase level or higher. In addition to these programs, specific lessons which were intended to facilitate and generalize verbal expression were also carried out. This continuum of language programs was implemented in a small group by the classroom teacher. During the third year, specific training was provided to each of the classroom teachers involved in the project. This training was carried out during the regular school year.

SAMPLE

First Year: The first year's sample consisted of 157 children located in 24 classes for the trainable mentally retarded. The children were between 7 and 14 years of age and possessed IQs between 25 and 50. From this population of 157 children, 120 were randomly selected. That is, 10 children were randomly selected from the 12 research design cell combinations formed by the factors of treatments

(3 levels) by IQ (2 levels) by sex (2 levels).

Second Year: The second year's population consisted of two groups: continuees (those who were in the first year of the study) and new entries (those who were brought into the study only during the second year of the project). In particular, there were 116 continuees (out of the original 157) and 114 new entries.

Third Year: Students who were in the Intermediate Unit's TMR classes for the trainable mentally retarded were involved in this project. During the first and second years of the project, students were involved in varying ways with ITPA training, Peabody training, and Distar training. As with any large public school operation, the Intermediate Unit's TMR classes during the third year had some children who were continued from the first project year, some who had entered during the second year, and continued into the third year, and still others who were new entries for the third year itself. Further, not all children from the Intermediate Unit's classes were involved in any given year of the project in the activities in question for that year. Thus, over the three years of the project, one has several possibilities for a student's involvement. Table 1 summarizes the possibilities of a student's continuation status in third-year activities. The variable of continuation status becomes crucial in the actual design of this third-year study, as the reader will see later.

Before any purging of data occurred due to continuation status or missing data, there was a total population of 233 children.

For reasons that should be clear to the reader, the investigators decided to eliminate any child with a continuation status of 2, 3, 4, or 5. Apart from the data arising from such continuation statuses being of doubtful "cleanliness," the numbers of children in these categories did not warrant further exploration.

Finally, another area of elimination from the formal concerns of this report were those children who received the combined Peabody/Distar programs.

For the Peabody-versus-Distar comparison, 5 students were randomly selected from each design cell to yield a total of 40 children. For the Distar intensive analyses, 4 children were randomly selected from each design cell to yield a total of 48 children.

INSTRUMENTATION

First and Second Years: The first year's operation of the Language Training Project yielded very limited treatment effects as gauged by the standardized tests used. Several staff members felt that this poor showing was due, to some extent, to the fact that the tests in question (which are among the best recognized instruments currently available) do not adequately tap the language functioning of interest to the study. The specific low level of language functioning given by trainable retarded children may require instrumentation not currently available.

During the first year of project operation, one very involved form of testing was that of Myklebust's Picture Story Language Test, as modified for this study (see Leiss, 1974). Myklebust (1965) used an action-packed picture to elicit samples of a student's written language. In contrast, the second year used an adaptation to the extent that a student's language was elicited in oral rather than written form; these oral language samples were tape recorded to preserve them exactly for later scoring. During the first year of the project, three pictures were used. Each picture was measured for "Productivity" by means of three criteria: total words, total sentences, and words per sentence. Further, each of the three pictures was evaluated for "Meaning/Content" by means of Myklebust's "Abstract-Concrete Scale."

Because of the meager testing results and because of the large amount of work involved in deriving the total of four different scores for each of the three pictures, the modified Myklebust Picture Story Language Test (dubbed "Language Sample" for this study) was largely omitted from the second project year design. One notable exception was to give the Language Sample to students who had continued from the first project year into the second project year. The main reason for this exception was to assess the longitudinal summer-lag forgetting phenomenon in trainable retarded children. To project staff knowledge, such data have never before

been reported in the literature. Thus, no post-testing was given at the end of the 1973-1974 year in terms of the Myklebust Language Sample. It was felt the saving in time was more than justified.

With the above reduction in total individual test administration time required for each child, the second project year opted to maintain a minimal battery of pre- and posttesting. Three instruments would be given as the measurement core: Peabody Picture Vocabulary Test, Illinois Test of Psycholinguistic Ability, and the Mecham Verbal Language Development Scale.

Third Year: A variety of measurement approaches were used in this study to tap as many different language-related skills as possible. A series of standardized and/or commonly used devices were employed. Also, a type of criterion-referenced measurement (CRM) was embodied in the intensive study of the Distar program.

With regard to standard measurement technology, several approaches were used. First, selected subtests of the Illinois Test of Psycholinguistic Abilities (ITPA: Revised) were given to all children. In particular, the Auditory Association, Visual Association, Verbal Expression, and Grammatic Closure subtests were selected (scaled scores only). Second, the Peabody Picture Vocabulary Test (Form B) was given (mental age). Third, the Spencer Memory for Sentences Test was used. Apart from these routinely used devices, the Distar Reading Program Placement Test was also given to all children. The Placement Test is divided into Parts A, B, C, D, and E; in addition, a Total score is also yielded.

The second major phase of the third year concerned only the Distar program itself. It was felt that an intensive investigation of the Distar program would reveal interesting facts about how language develops in TMR children subjected to a highly structured, commercially available program. The school year was divided into seven periods. At the end of each period, the lesson number in the books which the children were currently using was recorded and used as the data input.

For each period, there were five possibilities: Book A, Book B, Book C, Story Book, and Coloring Book. Various schemes of reporting in-process Distar progress were explored prior to settling upon this final choice, but enormous problems arose in selecting a uniform, meaningful method of scaling. The lesson numbers seemed to be the most interpretable system of recording progress. In this sense, the data system is akin to a criterion-referenced measurement (CRM) system in which only absolute levels of performance are reported with no relative comparisons being made.

Finally, the Distar Placement Test needs some further elaboration. The Test (located in Book A, Part I, of Distar Language I: An Instructional System, by S. Engelmann, J. Osborn, and T. Engelmann, Chicago: Science Research Associates, 1972) consists of Segments A through E. Segment A contains 5 tasks, each of which deals with "Action and Identity Statements." For instance, a child would be asked to point to an animal and then asked what the animal was doing. The respective number of items for each Task is as follows: Task 1, 2; Task 2, 2; Task 3, 5; Task 4, 2; Task 5, a. Segment B deals with "Can/Is Action Statements." The child merely has to answer "yes" or "no" to such items. There are 6 such items. Segment C deals with "Polars." For instance, the child is asked whether something is long or short and merely has to answer "yes" or "no". There are 5 such items. Segment D deals with "Polars" also, but at a higher conceptual level. The child is shown a big object, for instance, but then is asked to tell about the opposite concept, a similar object. There are 4 items. Segment E deals with "Prepositions." There are 7 items.

DESIGN AND ANALYSES

First Year: Pre-testing and post-testing was carried out with the Peabody Picture Vocabulary Test (PPVT) and the twelve subtests of the Illinois Test of Psycholinguistic Abilities (ITPA). Further, it was later decided to add the Mecham Verbal Language Development Scale (VLDS) to obtain still another outside criterion of language development. The testing was carried out at both the start and conclusion of the lengthy training period (November to May).

Every attempt was made to ensure that the three groups of classes were comparable at the start of the study. Because the children had to be kept in their original classes due to administrative and logistical reasons, randomization could be used only at the class level. Thus, the 24 classes were randomly distributed among the three group settings: (a) no stimulation, (b) four times a week, or (c) eight times a week. Further, initial comparability of the three sets of classes was achieved by analyzing pre-test differences on the PPVT, ITPA, and VLDS. Besides using the total raw scores from each of the three criteria, the twelve subtest raw scores from the ITPA were also analyzed. The BMD02V computer program for analysis of variance for factorial design (version of July 22, 1965) from University of California at Los Angeles was used. Each analysis of pre-test differences embodied three factors: (a) treatments, (b) IQ, and (c) sex. However, because IQ and sex differences were not of immediate interest for establishing initial equivalence of groups, only the factor

of treatments will be considered here. No significant pre-test differences with regard to treatments were found on any of the 15 analyses. Thus, for all intents and purposes of the evaluation design, the three sets of groups can be considered initially equivalent. (It should be noted that unequal cell frequencies were present in the original three-factor data matrix involving 157 children. Several chronic absentee children were among the 157 children. After the decision was made to remove these absentees from the initial data matrix, the new unequal cell matrix comprised 148 children. To achieve final equal cell frequencies, a cell size of ten was decided upon and children were randomly deleted from the appropriate cells. The resulting matrix, also used in later analyses, contained 120 children.)

Once initial equivalence of the three sets of groups was ascertained, a formal program evaluation design was selected. In particular, besides the three factors of treatments, IQ, and sex, a fourth factor of measures (pre-test versus post-test) was added. The resulting four-factor design was of repeated-measures type. As with the 15 pre-test analyses, 15 analyses were run in the repeated-measures framework. The computer program used was BMD08V of the UCLA Biomedical package (version of September 1, 1965).

The reader should note that in every one of the 15 analyses, a mixed effects model was derived. That is, the factors of treatments and sex were considered fixed, but the factor of IQ (high and low, as determined by an approximate median split) was taken to be random. (Of course, replications or subjects were considered random in the data matrix wherein 120 children were left after removing unequal cell frequencies.)

Apart from the four-factor, repeated-measures design used in the 15 gain analyses, descriptive analyses were also undertaken of the variable

of socioeconomic status (SES). While IQ has remained the main control variable of interest used in the above-mentioned 15 gain analyses, SES was also of interest. SES could not be included as a fifth factor in the design for the above 15 analyses because the distribution of frequencies among factors was too uneven. Thus, it was decided to analyze separately in a descriptive way the effect of SES on the three treatments. The SES measure was the Minnesota Scale for Paternal Occupations; categories I to IV were considered High SES, while V to VII were Low SES.

Second Year: The primary concern of this study was the treatment comparison between the Peabody program and the Distar program. Wherever administratively possible, the classes containing both continuees and new entries were randomly assigned evenly between the two treatment conditions. Because of the potency of the IQ factor as a control variable, the second factor included in the design was IQ. A median split was employed so that low IQ represented 21 to 43, while high IQ was 44 to 53. The third factor was sex (males versus females). The fourth factor was measures (pretest versus posttest). Thus, the basic design for several analyses was a four-factor, repeated-measures design: treatments by IQ by sex by measures.

Besides the four-factor design mentioned above, a fifth factor was embodied for certain analyses, namely, entry status. This factor had two levels: new entry versus continuee. Thus, the few analyses that included this fifth factor were of a five-facotr, repeated-measures design: treatments, IQ, sex, entry status, and measures.

One series of analyses dealt with the three criteria of the PPVT, ITPA, and VLDS. The pretest and posttest data from the 1973-1974 year were placed within the four-factor design mentioned above. For each of the criteria, two separate analyses were performed: one for continuees and one for new entries. However, before any analyses were run, 7 children were randomly selected from each of the independent-factor cells (treatments by IQ by sex). Thus, each of the analyses had 56 children drawn at random from either the 116 continuees or the 114 new entries. A total of 6 such analyses were run.

A second set of analyses built in as a factor the comparison of continuee versus new entry. Each of these analyses was again done on 1973-1974 data of pretest-posttest type for the PPVT, ITPA, and VLDS. A total of 3 such analyses

were run.

A third set of analyses used only the data from the 56 continuees. These analyses represented longitudinal studies. This set of analyses involved the four pretest-posttest measures from both 1972-1973 and 1973-1974. Three of these analyses were run: PPVT, ITPA, and VLDS.

A fourth set of longitudinal studies were run on the 56 continuees with regard to the Myklebust Language Sample data. The input consisted of the pretest and posttest of 1972-1973 and the pretest of 1973-1974. Six such analyses were run: total words, total sentences, modified words per sentence, words per sentence as per Myklebust, abstractness-concreteness score, and average abstractness-concreteness score.

In all analyses, the BMD08V program of the UCLA Biomedical series was used. The analyses were run on a CDC 6400 computer at Lehigh University. A mixed design was specified, with treatments, sex, and measures as fixed factors, while IQ and replications were random factors.

Third Year:

The two basic sets of analyses in the third year of this project deal with the Peabody versus Distar comparisons and with the Distar intensive comparisons.

For the Peabody versus Distar phase, there were five factors: treatments (T), sex (S), continuation status (C), measures (M), and replicates (R). Replicates was taken as a random factor, while the other four factors were interpreted as fixed. To achieve equal cell frequencies in this design, five children were randomly sampled from each design cell. Thus, while treatments were not originally assigned in a random way (but rather in the real-life manner of what treatment would be most appropriate with a given child), the final design involved random sampling of children. Replicates were nested under the factors of treatments, sex, and continuation status. With the exception of replicates (which had five levels), all factors had two levels. While the factor of IQ might have been included, the design already was "saturated" with factors that did not allow any further stratifying of subjects if a reasonable number of children were to be kept in each cell. In terms of any differential bias that might exist in favor of a treatment, the pre-level of the test in question, as well as the previous experience of the children as indicated by continuation status, were deemed more crucial factors than IQ itself.

For the Distar intensive phase, three different sets of results were generated: (a) pretest-posttest results on the same criteria as in the Peabody versus Distar phase, but this time replacing treatments by IQ to obtain additional information on the Distar Program itself; (b) intra-year findings for each of the 7 periods (separately) of the school year that the project used for CRM data recording; and (c) intra-year data across all periods. In these regards, all three sets of intensive Distar analyses were loosely looked upon as "longitudinal" in nature. For (a), five factors were embodied: sex, IQ, continuation status, replicates, and measures. IQ was determined on an approximate, median-split basis, which in this study happened to be between 44 and 45; thus, the factor of IQ was random rather than fixed (that is,

it was empirically determined rather than logically/arbitrarily predetermined). Replicates was also taken at random. Sex, continuation status, and measures were all interpreted as fixed. All factors had two levels, with the exception of continuation status, which had three (E.S.1, C.S.6, and C.S.7), and replicates, which had 4 levels (4 randomly selected children from each categorical cell). For (b), measures as a factor was omitted. For (c), the few criteria that had complete data across all seven periods of the year employed the same kind of, and number of levels of, factors as in design (a) with the exception of measures, which now had seven levels.

From the above "Design" discussion, the reader should be able to visualize the nature of the three different designs used in this study: (a) Peabody versus Distar, (b) Distar Intensive Analyses (Single Criteria), and (c) Distar Intensive Analyses (Multiple Criteria).

In all analyses, the BMD08V program of the UCLA Biomedical series was used. The analyses were run on a CDC 6400 computer at Lehigh University, Bethlehem, Pennsylvania.

So that the reader understand in what manner the statistical tests of significance were carried out, it is necessary to describe briefly the error terms. In this discussion, several abbreviations will be used: T (treatments), S (sex), C (continuation status), M (Measures), and R (replicates). For design (a), the Peabody versus Distar phase of the study, the variance source to be tested for significance and its appropriate error term, are given respectively in pairs as follows: Mean, R (TSC); T, R (TSC); S, R (TSC); C, R (TSC); M, RM (TSC); TS, R (TSC); TC, R (TSC); SC, R (TSC); TM, RM (TSC); SM, RM (TSC); CM, RM (TSC) TSC, R (TSC); TSM, RM (TSC); TCM, RM (TSC); TSCM, RM (TSC); and RM (TSC), not tested:

For design (b), the Distar intensive analyses of single criteria, the variance source to be tested for significance and its appropriate error term, are given respectively in pairs as follows (Q denotes IQ) Mean, Q; S, SQ; Q, R(SQC); C, QC; SQ, R(SQC); SC, SQC; QC, R(SQC)1 SQC, R (SQC); and R(SQC), not tested.

For design (c), the Distar intensive analyses of multiple criteria, the variance source to be tested for significance and its appropriate error term, are given respectively in pairs as follows: Mean, Q; S, SQ; Q, R(SQC); C, QC; M, QM; SQ, RM(SQC); CM, QCM; SQC, R(SQC); SQM, RM(SQC); SCM, SQCM; QCM, RM(SQC); R(SQC), not tested; SQCM, RM(SWC); and RM(SWC), not tested.

RESULTS AND DISCUSSION

First Year:

The Language Training Project has many important implications for realistic school practice and for future research. The main variable of interest was the intensity of application of a single training approach based around the twelve subtest tasks of the ITPA.

There are two findings that have a huge impact on realistic school functioning. First, there were no significant differences among levels of intensity (the "treatments") in any of the 15 analyses. Second, only two out of the 15 analyses yielded any significant changes (one was a gain and the other was a loss); in general, there appeared to be little improvement of the children. From these two results, there seems to be only one conclusion possible: specific, prolonged language training based upon the ITPA is ineffective no matter what the intensity of application is. However, it should be noted that this does not mean that other types of language training with the trainable mentally retarded would be similarly ineffective. Nonetheless, this general conclusion must be tempered by the presence of some significant interactions.

The treatment-by-IQ interactions which occurred (seven significant ones out of 15 total) showed two situations. In three of the seven interactions, the logically expected superiority of the High-IQ group over the Low-IQ group did not materialize for the four-times-a-week group. In the other four interactions, the expected superiority situation did not occur for either the four-times-a-week group or the eight-times-a-week group. From these findings and insepctions of tables not included in this

because of their excessive detail, it appears that the specific language training actually impeded the High-IQ groups.

In summary, this research report of the first year's results showed that specific language training based upon the ITPA has no effect on trainable mentally retarded children and in fact seriously hinders the upper level IQ group in this population.

Second Year: The basic evaluative emphasis during the second year of the project was on the global pretest-posttest assessments via the PPVT, ITPA, and VLDS. Because of the many univariate analyses performed in this annual project evaluation, some words of interpretative caution should be attached to the results. Primary weight should be attached to the findings from the analyses that embody the most all-emcompassing comparisons among both continuees and new entries for the three primary criteria, and for the analyses that embody the most all-encompassing comparisons among just continuees for the longitudinal (two-year) data for the three primary criteria. With these precautions as a preface, the basic findings will be discussed.

For the analyses containing both continuees and new entries, there are no generalizable treatment effects in favor of either Peabody or Distar. This is to be expected because human language behavior is so complex that one would hardly expect one program to be effective for all levels of disability or functioning within the TMR population. Thus, one looks to the interactions with treatments to provide the qualifications on lack of general findings that say in specific levels of TMR functioning, certain programs may nonetheless be effective. In the analysis of VLDS the treatment-by-IQ interaction was significant. Not only did the Distar groups surpass the Peabody groups, but the low-IQ group did not lag so far behind the high-IQ group with the Distar program as they did in the Peabody program. In the analysis of PPVT, no two-way interactions with treatments were significant. In the analysis of ITPA, again no significant two-way interactions with treatments were found.

In terms of gain during just the second year, none of the three primary criteria showed significant movement for the analyses that included both continuees and new entries. Further, none of the interactions with gain were significant on VLDS. However, for PPVT, the IQ-by-gain interaction was significant.

Regardless of language program, low-IQ children actually lost over time, while high-IQ children gained over time. On the ITPA, again no significant two-way interactions with gain occurred.

Focusing just on the continuees from the first year of the project, one can detect some interesting trends. Here, longitudinal data was used from both project years. On the ITPA, the main effect for treatments was not significant, nor were any of the two-way interactions with treatments. On the PPVT, a different picture emerged. The treatment-by-IQ interaction ($p < .05$) showed that while no overall difference between Peabody and Distar existed, the low-IQ children in Peabody were greatly hindered in comparison to the other three treatment-by-IQ combination groups. Also on the PPVT, the treatment-by-gain interaction ($p < .05$) showed the continuees had significantly higher performance in the Distar groups than in the Peabody groups, with the greatest gain occurring during the second year. On the VLDS, no general treatment effect occurred, but two interactions with treatments are worthy of discussion. The treatment-by-IQ interaction ($p < .05$) showed that for continuees, the low-IQ Peabody group performed significantly worse than the high-IQ Peabody group, while the corresponding difference in the Distar groups was in the same direction but less pronounced. The treatment-by-sex interaction was also highly significant ($p < .01$); in the Peabody groups, males were significantly lower than females, while the reverse was true in the Distar groups.

The final reflections on analyses containing only the continuees deal with the gain phenomenon. On the ITPA, a significant ($p < .01$) gain occurred regardless of treatment. However, strangely enough, while the posttest of each year was higher than the corresponding pretest, the overall performance of the first year was higher than the second year. No overall change occurred in the analysis of the PPVT; however, change did occur depending upon treatment group (a finding already discussed above). On the VLDS, a highly significant ($p < .01$) change over

time occurred regardless of treatment. Here, there was a notable gain during the second year of the project for the continuees, while their first year's performance was more or less static. Also on the VLDS, there was a significant ($p < .01$) IQ-by-measures interaction. The interaction was caused mainly by the low-IQ students losing at a greater rate than the high-IQ students. The low-IQ students on an average lost twice as many points over summer as did the high-IQ students. Also, while the high-IQ students finally got back up at the end of the second year where they had been at the end of the first year (but got no higher!), the low-IQ students did not even get up to the level they were at during the end of the first year.

In summary, then, the above findings are those in which perhaps the greatest degree of confidence could be placed in lieu of actually having a multivariate analysis of variance design. From this brief precis of key findings, it now remains to put a perspective on them.

What can one conclude from the primary set of results? With regard to treatments, while no significant differences emerged for the high-IQ children, the low-IQ children were aided moreso by Distar than by Peabody. Further, the continuees showed greater gains during the second year of the project in Distar than in Peabody. Of course, one must remember that these continuees during the first year were in various types of ITPA-based language stimulation programs. Thus, these children who continued on into the second year of the project (at which time Peabody and Distar were introduced) had the benefit of earlier language stimulation, although the first year's project report indicated such ITPA-based training was of minimal value. (Children who had various degrees of ITPA-based training during the first year were, of course, randomly represented in each of the Peabody and Distar groups so that no differential pre-treatment contamination existed at the start of the second year).

In terms of change over time, two observations are possible. First, because of the poor showing in analyses 7, 8, and 9, gain in the total sample was not marked (i.e., in those analyses where both continuees and new entries were considered). However, when one considers only the continuees, significant gain in language functioning did occur. Second, the summer lag phenomenon did occur for those TMR children who were continuees; that is, in considering the posttest from the first year and the pretest of the second year, a marked decrease in performance occurred.

The final set of observations concern the measurement realm. It is clear that the battery of standardized tests used in both the first and second years of the project have not been specific enough to tap areas of language functioning of concern to this project. That is, the PPVT, ITPA, and VLDS are simply not valid enough reflections of the types of language training used with the TMR children. The sensitivity of these instruments is extremely poor for detecting subtle changes in TMR children's performance. Just what measurement devices might be substituted for the present ones is a question to which the present investigators cannot give a legitimate answer. It would seem desirable to consider implementing a curriculum-based, criterion-referenced measurement system. For example, if one is in the Distar program, then perhaps a recording system could be developed that would reflect developmental mastery changes of the children as they move throughout the various sequential units of Distar. In its crudest form, this CRM system might use only the sequential unit numbers at the end of every week or every two weeks for each child throughout the school year. One could make the CRM system a little more precise if he not only considered developmental unit numbers (which reflect an implicit mastery of the curricular continuum) but also appended percentage mastery scores on some criterion attached to each unit.

The last set of measurement considerations concern the analyses that were considered only subordinate in importance: longitudinal language sample data on continuees. With regard to the Myklebust modified Language Sample procedure, a few general conclusions are possible. First, the procedure is time-consuming both to administer and to score. Second, when all the various scores of the Language Sample are considered, only the Total Words and Total Sentences appeared to be sensitive to the types of language functioning of TMR children.

In summary, then, the second-year results appear to be more positive (mainly in favor of Distar over Peabody) than the first-year results in which different intensities of ITPA-based language training yielded a very bleak picture. Nonetheless, even the second-year findings are relatively mild and contain no stunning revelations.

Third Year (Peabody Versus Distar): The reader must be cautioned on the restrictions that pertain to this design. In brief, these constraints involve the self-selection bias built into the study by default, with the more language-advanced students receiving the Distar program. If there is a significant bias at work that contaminates treatment comparisons in any noticeable way, then this should be easily detectable by examining the interactions between treatments and measures; in particular, the pretest levels pitted against treatments should show the obvious differential biases at work. Out of a total of 12 analyses of Peabody-versus-Distar type, only two showed a significant treatments-by-measures interaction. These two results involved the criteria of Parts D and E of the Distar Placement Test. On all other criteria (Distar Placement Test: Total Score and Parts A through C; ITPA: Grammatic Closure, Auditory Association, Verbal Expression; and Visual Association; Spencer Memory for Sentences Test; and Peabody Picture Vocabulary Test) the treatments-by-measures interactions were insignificant.

Because of this situation, the investigators decided to simply omit any discussion of the contaminated Parts D and E of the Distar Placement Test and to proceed very cautiously with the results of the remaining criteria.

Of primary interest to this study was the main effect of treatments. Of the 10 "uncontaminated criteria," the treatment main effect was significant in all cases except on ITPA Grammatic Closure and ITPA Visual Association. Of the 8 significant treatment effects, the Distar Program was clearly superior to the Peabody Program on all of the criteria except ITPA Auditory Association, on which the opposite finding occurred.

With regard to the other variables in the design, little was found except for measures. Sex yielded no differences in language activity except on ITPA Visual Association, on which boys were superior to girls. No difference at all was found among the 10 uncontaminated criteria for the variable of continuation status; apparently the previous, formal language experiences of the children had little effect. The variable of measures yielded significant main effects on each of the 10 criteria except on the Spencer Memory for Sentence Test. Thus gains uniformly occurred, regardless of treatments.

With regard to interactions on the 10 uncontaminated criteria, the only significant two-way result was for sex by continuation status on ITPA Visual Association; in particular, for C.S.6, girls performed significantly higher than boys, while for C.S.1 no difference occurred.

Fortunately, triple interactions were kept to a minimum on the 10 criteria of "uncontaminated" type. Thus, discussion of results is greatly simplified. The only exceptions were treatments by continuation status by measures on ITPA Verbal Expression, treatments by sex by continuation status on ITPA Visual Association, treatments by sex by measures on Distar Placement Test (Part A), and treatments by continuation status by measures on Distar Placement Test (Part C).

Third Year: Distar Intensive Analyses (Pretest-Posttest Findings): Basically, this series of analyses was the logical continuation of those presented above and changed only in the sense of replacing the treatments factor with the IQ factor. The purpose of these analyses was to examine more intensively any trends that might be at work within the Distar Program itself (which was the main focus of interest of this third year of the project).

As with the Distar/Peabody analyses above, the pretest-posttest analyses again had 12 criteria. In no case did any sex differences arise, with the exception of a complex triple interaction among sex, IQ, and measures. On all but 4 of the criteria, IQ operated effectively in the expected direction to control some of the variation (the exceptions were ITPA Grammatic Closure, ITPA Auditory Association, ITPA Verbal Expression, and ITPA Visual Association). The latter finding again raises some validity problems on the ITPA, since an IQ difference would be expected on virtually all criteria. Continuation status produced no difference at all on any of the 12 criteria. Measures (pretest and posttest) was significant on only one criterion: Distar Placement Test (Part E).

Apart from the main effects, the only other results that might be taken note of were an IQ-by-measures interaction on four of the 12 criteria (Distar Placement Test: Total Score, Part A, and Part C; ITPA Visual Association). Two triple interactions (sex by continuation status by measures, and sex by IQ by measures) were also significant but are too complex to go into for brief summary purposes here.

Third Year: Distar Intensive Analyses (Intra-Year Findings Separately for Each of 7 Periods): This series of analyses looked at the CRM data that was yielded in day-to-day contacts with the program. The lesson numbers of Distar Book A, Distar Book B, Distar Book C, Distar Story Book, and Distar Coloring Book, were monitored for each child in the Distar Program at the end of each of 7 evenly spaced time periods throughout the school year. During certain periods one will

notice that not all 5 possible lesson numbers were involved in the analyses (e.g., Period 1 has only Book A and Book B); this situation is merely a reflection of the fact that developmentally, the children did not begin using certain of the more advanced components (Book C, Story Book, and Coloring Book) until greater facility in language skills had been achieved at the lower levels (Books A and B).

There were a total of 22 CRM measures (as discussed earlier in the "Instruments" section) processed during this portion of the Distar Intensive Study. Of the three factors in the design (sex, IQ, and continuation status), only IQ resulted in any noticeable pattern of significant differences. Of the 22 CRM measures, only 6 did not yield a significant IQ difference (specifically, Distar Book A, Period 1; Distar Book B, Period 1; Distar Book C, Period 4; Distar Story Book, Period 5; Distar Book C, Period 6; Distar Story Book, Period 6; and Distar Book C, Period 7). There appears to be no real pattern to these nonsignificant differences, and little more need be said here.

The variables of sex and continuation status were consistent in that neither produced any significant differences. Inspection of the cell means tables in Appendix C shows why this situation arose.

Distar Intensive Analyses (Intra-Year Findings Across All 7 Periods): Basically, the purpose was to look at those CRM Distar measures that provided data across all 7 periods of the school year. Only Books A and B yielded such data; this situation was a partial reflection of the developmental problems of the children and the resultant times at which they started Distar Books A, B, or C, or the Story Book, or the Coloring Book.

As one can see from the tables, sex and continuation status were insignificant on both Book A and Book B. The factor of IQ worked effectively to isolate some of the error variance; the significant difference was in the expected direction on both measures. The factor of measures resulted in significant progress being

steadily shown throughout the school; in Book A, the lesson numbers ranged from 33.12 in Period 1 to 77.3 in Period 7, while in Book B, the range was from 21.42 in Book A to 77.23 in Period 7. The only other significant result was the IQ-by-measures interaction, on both criteria; no other interactions were significant on either criterion.

REFERENCES

Bangs, T. E. Evaluating children with language delay. Journal of Speech and Hearing Disorders, 1961, 26, 6-18.

Blessing, K. R. An investigation of a psycholinguistic deficit in educable mentally retarded children: Detection, remediation, and related variables. Unpublished doctoral dissertation, University of Wisconsin, 1964.

Blue, M. C. The effectiveness of a group language program with trainable mental retardates. Unpublished paper, Appalachian State Teachers College, 1963.

Brandfon, W. Speech problems of the mentally retarded child: A study based on case histories. Unpublished master's thesis, Louisiana State University, 1951.

Daum, M. P. A study of the articulation of a group of fifty institutionalized children with intelligent quotients between fifty and seventy. Unpublished master's thesis, Kent State University, 1953.

Donovan, H. Organization and development of a speech program for the mentally retarded children in New York City public schools. American Journal of Mental Deficiency, 1957, 62, 455-459.

Ensminger, E. E., & Smith, J. O. Language development and the ITPA. Training School Bulletin, 1965, 62, 97-107.

Everhart, R. W. The relationship between articulation and other developmental factors in children. Journal of Speech and Hearing Disorders, 1953, 18, 332-338.

Gens, G. W. Speech retardation in the normal and subnormal child. Training School Bulletin, 1950, 47, 32-36.

Gens, G. The speech pathologist looks at the mentally deficient child. Training School Bulletin, 1951, 48, 19-27.

Goodwin, F. G. A consideration of etiologies in 454 cases of speech retardation. Journal of Speech and Hearing Research, 1955, 31, 300-303.

Gottsleben, R. H. The incidence of stuttering in a group of mongoloids. Training School Bulletin, 1955, 51, 209-217.

Harrison, S. A review of research in speech and language development of the mentally retarded child. American Journal of Mental Deficiency, 1958, 63, 236-240.

Harvey, A., Yep, B., & Sellin, D. Developmental achievement of trainable mentally retarded children. Training School Bulletin, 1966, 63, 100-108.

Irwin, O. The developmental status of speech sounds of ten feebleminded children. Child Development, 1942, 13, 29-39.

Johnson, G. O., and Capobianco, R. J. Research project on severely retarded children. Albany, N. Y.: New York Inter-departmental Health Resources Board, 1957, 230 pp.

Johnson, G. O., Capobianco, R., & Miller, D. Speech and language development of mentally deficient children enrolled in training programs. Exceptional Children, 1960, 27, 72-77.

Jordan, T. E. Language and mental retardation: A review of the literature. In R. L. Schiefelbusch, R. H. Copeland, & J. O. Smith (Eds.) Language and mental retardation: Empirical and conceptual considerations. New York: Holt, Rinehart & Winston, 1967, 20-38.

Karlin, I. W., and Kennedy, L. Delay in the development of speech. American Journal of Disabled Children, 1936, 51, 1138-1149.

Karlin, I. W., & Strazzula, M. Speech and language problems of mentally deficient children. Journal of Speech and Hearing Disorders, 1952, 17, 226-294.

Kennedy, L. Studies in the speech of the feebleminded. Unpublished doctoral dissertation, University of Wisconsin, 1930.

Kolstoe, O. P. Language training of low grade mongoloid children. American Journal of Mental Deficiency, 1958, 63, 17-30.

Leiss, R. H. The effect of intensity in a psycholinguistic stimulation program for trainable mentally retarded children. Unpublished doctoral dissertation, Temple University, 1974.

Leiss, R. H., & Proger, B. B. Language training for trainable mentally retarded; Annual project report; First year; Project number F1-720201I-46-I.U.-01-13-C; Title III of the Elementary and Secondary Education Act of 1965; June 15, 1972, to June 30, 1973 ($49,977.00). Norristown, Pa.: Montgomery County Intermediate Unit Number 23, 1973. (Available from Educational Resources Information Center, ERIC, as document number Ed-082-424).

Lewald, J. Speech defects as found in a group of 500 mental defectives. Proc. and Address of American Association for the Study of the Feeble-Minded, 1932, 37.

Lubman, C. C. Speech program for severly retarded children. American Journal of Mental Deficiency, 1955, 60, 297-300.

Lyle, J. G. The effect of an institution environment upon the verbal development of imbecile children: 2. Speech and language. Journal of Mental Deficiency Research, 1960, 4, 1-13. (a)

Lyle, J. G. Some factors affecting the speech development of imbecile children in an institution. Journal of Child Psychology and Psychiatry, 1960, 2, 121-129. (b)

Masket, L. J. A comparative study of articulation problems of non-institutionalized and institutionalized mentally retarded children. Unpublished master's thesis, University of Maryland, 1958.

Matthews, J. Speech problems of the mentally retarded. In L. E. Travis (Ed.) Handbook of speech pathology. New York: Appleton-Century-Crofts, Inc., 1957

Meader, M. H. The effect of disturbances in the developmental processes upon emergent specificity or function. Journal of Speech Disorders, 1940, 5, 211-220.

Potter, R. E., and Mattson, R. H. Some positive findings regarding speech and language therapy with the mentally retarded. Digest of the Mentally Retarded, 1968, 4, 98-102.

Richardson, S. O. Language training for mentally retarded children. In R. L. Schiefelbusch, R. H. Copeland, & J. O. Smith (Eds.) Language and mental retardation: Empirical and conceptual considerations. New York: Holt, Rinehart & Winston, 1967, 146-161.

Rittmanic, P. A. An oral language program for institutionalized educable mentally retarded children. American Journal of Mental Deficiency, 1958, 63, 403-407.

Sachs, M. H. A survey and evaluation of the existing inter-relationships between speech and mental deficiencies. Master's thesis, University of Virginia. (Reported by J. Eisenson. The nature of defective speech. In Psychology of Exceptional Children and Youth. New York: Prentice-Hall, 1955.)

Schiefelbusch, R. L. Children with speech and language impairments. In S. A. Kirk & B. B. Weiner, Behavioral research on exceptional children. Washington, D.C.: Council for Exceptional Children, N. E. A., 1963, 259-290.

Schlanger, B. B. Speech measurements of institutionalized mentally handicapped children. American Journal of Mental Deficiency, 1953, 53, 114-122. (b)

Schlanger, B. B. Speech therapy results with mentally retarded children in special classes. Training School Bulletin, 1953, 50, 179-186. (c)

Schlanger, B. B., & Gottsleben, R. H. Analysis of speech defects among the institutionalized mentally retarded. Journal of Speech and Hearing Disorders, 1957, 22, 98-103.

Schneider, B., and Vallon, J. A speech therapy program for mentally retarded children. American Journal of Mental Deficiency, 1954, 58, 633-639.

Schneider, R., & Vallon, J. The results of a therapy program for mentally retarded children. American Journal of Mental Deficiency, 1955, 59, 417-424.

Sheridan, M. D. The child's hearing for speech. London: Methuen, 1948.

Sirkin, J. and Lyons, W. F. A study of speech defects in mental deficiency. American Journal of Mental Deficiency, 1941, 46, 74-80.

Smith, J. O. Effects of a group language development program upon the psycholinguistic abilities of educable mental retardates. Peabody College Res. Monog. Series in Special Education, 1962, 1, 111-123, George Peabody College for Teachers, Nashville, Tenn. (a)

Smith, J. O. Group language development for educable mental retardates. Journal of Exceptional Children, 1962, 29. (b)

Tarjan, G., Wright, S. W., Dingman, H. F., and Eyman, R. K. Natural History of mental deficiency in a state hospital: III. selected characteristics of first admissions and their environments. Amer. J. Dis. Child., 1961, 101, 195-205.

Town, C. H. Language development in 285 idiots and imbeciles. Psychol. Clin., 1913, 235.

West, R., Kennedy, L., & Carr, A. The rehabilitation of speech. (rev. ed.) New York: Harper & Brothers, 1947.

Wolfensberger, W., Mein, W. R., & O'Connor, N. A study of the oral vocabularies of severely subnormal patients: III. Core vocabulary, verbosity, and repetitiousness. Journal of Mental Deficiency Research, 1963, 7, 38-45.

Wood, N. E. Causal factors of delayed speech and language development. American Journal of Mental Deficiency, 1957, 61, 706-708.

TABLE 1

CONTINUATION STATUS LEVELS

LEVEL	DESCRIPTION
C.S. 1	1974-1975 (1 year of contact)
C.S. 2	case study control group (3 years of contact)
C.S. 3	controls in 1972-1973, then regular program in 1973-1974 and also 1974-1975 (2 years of contact)
C.S. 4	in year 1, out year 2, in year 3 (2 years of contact)
C.S. 5	in all 3 years, but case study in year 1 (3 years of contact)
C.S. 6	in year 2 (1973-1974) and year 3 (1974-1975) (2 years of contact)
C.S. 7	in all 3 years (3 years of contact)