5-1985

A Comparison of Evaluation Models for Handicap Intervention in a Head Start Program

Carin Niebuhr
Utah State University

A COMPARISON OF EVALUATION MODELS FOR HANDICAP INTERVENTION IN A HEAD START PROGRAM

by

Carin Niebuhr

A thesis submitted in partial fulfillment

of the requirements for the degree

of

MASTER OF SCIENCE

in

Psychology

Approved:

Major Professor

Committee Member

Committee Member

Committee Member

UTAH STATE UNIVERSITY
Logan, Utah

1985

ACKNOWLEDGEMENTS

The Head Start Handicap Intervention Research Project was done in conjunction with the Multi-Agency Project for Pre-Schoolers developed by the Exceptional Child Center at Utah State University. Grateful acknowledgement is given to John and Nancy Chambers, Phyllis Cole, Kathy Gordon, and Adrienne Peterson for assistance in the child assessments and intervention planning done for this project.

This project was also dependent on the cooperation and assistance of the children, parents, and staff at Jackson County Head Start. Ada Pollach, Program Education Coordinator, was responsible for the logistic organization for child assessments of 200 children at seven centers. Nancy Nordyke and Kathy Johnson, Handicapped Specialists, carried the major responsibility for handicap program implementation.

The major credit for this project must be given to Dr. Glendon Casto whose initial interest in finding new statistical techniques for preschool programs first let to the design of this project. Without Dr. Casto's expertise, support, and encouragement, the hurdles due to the complexity and logistic difficulty of the research project would never have been overcome. I believe this project is an effective example of the value of combining academic research expertise with field program needs and am much in debt to Dr. Casto for his skills and commitment in this area.

Carin Niebuhr

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

ABSTRACT

A Comparison of Evaluation Models for Handicap
Intervention in a Head Start Program

by

Carin Niebuhr, Master of Science
Utah State University, 1985

Major Professor: Glendon Casto, Ph.D.
Department: Psychology

The Model A and Model C Title I evaluation options were compared by using both options to measure the effectiveness of handicap intervention in a Head Start program. Two hundred three children in Jackson County (Oregon) were pretested with the Developmental Indicators of Learning Test (DIAL), the Peabody Picture Vocabulary Test (PPVT), the Visual-Motor Integration Scale (VMI), and the Carrow Elicited Language Inventory (CELI). The 43 children who scored below the predetermined cut-off level were placed in a six-month intervention program. One hundred forty-nine children remaining in the Head Start program in May were posttested with the same tests. Model A analysis of mean scores of the intervention group indicated significant score change on all three testing instruments. Model C analysis indicated no positive score change. It was posited that the Model A effect in this project was large because it combined a positive intervention effect with a positive general program effect. The Model C option showed no effect because the estimated nonintervention scores were very large due to the large positive score change in the nonintervention group.

(85 pages)

CHAPTER I

INTRODUCTION

Head Start was created in 1965 by Congressional mandate as one of the weapons in the "War on Poverty." The purpose was to prevent school failure in disadvantaged children. The writings of Hunt (1961) and Bloom (1964) had suggested that all children had critical periods of brain development, "magic years" when environmental intervention was crucial for later cognitive development. Remedial efforts beginning in elementary school were too late; if society wanted low-income children to succeed, it needed to educate them in the critical, early years of development. Congress responded to these arguments for early intervention by approving and funding the Head Start program, a nationwide effort to educate and nurture young low-income children.

Data from studies with several groups of handicapped children (Bronfenbrenner, 1975) have indicated that there may be even more need for early intervention with handicapped children than with disadvantaged children since many handicapping conditions appear to have sensitive periods of early development when environmental or sensory intervention may help to maximize the child's potential. Even without concern for a "sensitive period," handicapped children require additional early adaptive education to adjust to or overcome their deficiencies. There are three general reasons given for providing early intervention for handicapped children: (a) to enhance general cognitive development; (b) to enhance specific cognitive or sensory development related to the handicap; and (c) to accelerate the attainment of adaptive or compensatory skills to overcome deficiencies related to the handicap.

In 1972, Congress mandated that 10 percent of all enrollees in Head Start be handicapped children. Congress was concerned because until 1972, only 25,000 of the estimated one million preschool handicapped children were served by programs receiving federal funds, and there were few state or local handicapped programs for young children (Wynne, Ulfelder, & Dakoff, 1975). Head Start was the only nationwide preschool program available, and Congress reasoned that a program designed to educate disadvantaged children would also be effective in educating handicapped children.

Although enrollment of handicapped children was officially mandated in 1972, it was not until 1974 that the Head Start national office provided guidelines for defining what constituted a handicap (see Apendix A). Additional legislation was added when Congress passed Public Law 94-142 which required all education programs for handicapped children to include parental consent for evaluation and treatment, individualized education plans, and education for all children in the least restrictive setting. In 1976, Congress passed a bill which supplemented Head Start programs with additional funding to assist with its handicapped programs. Over the past 9 years, over 350,000 handicapped children have been mainstreamed into Head Start programs (Status of Handicapped Children in Head Start Programs, 1980).

There has been little research done on the effects of intervention with handicapped children. Research done on other early intervention programs with disadvantaged children indicates some general trends:

(a) there is usually an immediate positive effect on IQ, achievement, and developmental test scores (Berieter & Englemann, 1966; Karnes, Hodgine, & Kirk, 1969; Weikert, 1970); (b) many of these increases "wash-out" by third to fifth grade (Weikert, 1970); (c) there are fewer placements of children in special education classes and fewer school dropouts (Weikert, Bond, & McNeil 1978); (d) differences in the effectiveness of various curriculum models are not great, although there is some preference for structured programs, particularly with more disadvantaged children (Stanley, 1972); and, (e) parental involvement in the education programs, particularly in those models which use parents as teachers, may prevent some of the wash-out of educational gains which happens with center programs (Bronfenbrenner, 1975).

Given increasing legislative concern with cost-effectiveness and program effectiveness, early education programs need to provide data which addresses these concerns and assists them in making program improvements. It is, however, difficult to do meaningful evaluation research in the field. With the exception of heavily funded demonstration programs, ongoing service programs must conduct evaluation with existing personnel and facilities. They must also, as May (1979) pointed out, adhere to existing criteria for pupil selection and service delivery. Another difficulty in undertaking evaluation research is that in the wake of Public Law 94-142, the option of placing children in a "no-service" control group is becoming increasingly unacceptable (May, 1979).

There are alternatives to control group designs developed for Title I evaluation which may be applicable to other remedial education programs such as Head Start. The Title I Outcome Evaluation (Tallmadge & Wood, 1978) was developed under a 1974 mandate by Congress requiring Title I to create a program evaluation and reporting system that would adequately inform Congress on the effectiveness of Title I programs. The Outcome Evaluation presents three evaluation options: (a) Model A, which uses norm-referenced comparisons; (b) Model B, which uses control group comparisons; and, (c) Model C, which uses regression line comparisons. If Model A and C are valid measures of program effectiveness, they could be important evaluation tools for Head Start and other intervention programs. One way to determine the validity of supposedly similar evaluation options is to use both and compare results. A research project designed to carry out this comparison in a standard Head Start program using regular program resources and personnel would also answer many questions about using the evaluation models with a Head Start population in a normal field setting. The lack of such research is the problem underlying the research for this thesis.

Objectives

The purpose of this project was to determine whether the norm-test (Model A) and the regression-line (Model C) options of the Title I evaluation models would produce similar results if applied to an intervention program done with educationally-handicapped children in Jackson County (Oregon) Head Start.

Hypothesis I - Norm-Group Hypothesis (Model A)

The mean posttest status will be significantly higher than the mean pretest status for the intervention group children as measured by: (a) scores on the Peabody Picture Vocabulary Test (PPVT), (b) scores on the Visual-Motor Integration Scale (VMI), and (c) scores on the Carrow Elicited Language Inventory (CELI).

The status of the children in the norm-group model is measured in normal curve equivalent units, a normalized percentile scoring system with a mean of 50 and a standard deviation (SD) of 21.06. The assumption is that, without intervention, the students would remain at the same NCE status in relation to the rest of the norm group.

Hypothesis II - Regression-Line Hypothesis (Model C)

The actual mean posttest score obtained by the intervention group children on the (a) PPVT, (b) VMI, and (c) CELI will be significantly higher than the mean posttest PPVT, VMI, and CELI scores estimated from the comparison group regression line.

Hypothesis III - Comparison of Results Using Model A and Model C

1. Model A and Model C Analysis of (a) PPVT, (b), VMI, and (c) CELI scores of the treatment and nontreatment groups in this project will give the same results in terms of educational significance using a gain criteria of six NCE as defined by Tallmadge and Wood (1980a).

2. Model A and Model C Analysis of (a) PPVT, (b) VMI, and (c) CELI scores of the treatment and nontreatment groups in this project will give the same results in terms of statistical significance using a t-score analysis at a .05 level of confidence.

CHAPTER II

REVIEW OF LITERATURE

Rationale for Early Intervention

In the early part of the 20th century, it was largely assumed that a person's capabilities were a result of genetic inheritance and, thus, not subject to remediation. However, in the 1950's, psychologists began to examine this assumption. Hunt's (1961) review led him to postulate that there was often a critical period of learning which, if passed through without appropriate stimulation, impeded or prevented later learning. Benjamin Bloom's (1964) demographic studies indicated that the early environment of children was crucial to later intellectual performance. He also believed that rate of brain growth in early years depended on amounts of environmental stimulation during periods of critical growth. Although there has been much criticism in recent years (Stanley, 1972; Zigler & Valentine, 1979) of the critical period of learning concept, it was instrumental to the establishment of early intervention programs, particularly Head Start.

Effects of Early Intervention

Bronfenbrenner (1975) reported that the first well-designed experimental programs in early childhood intervention were those of Kirk (1958), Klaus and Gray (1968), and Weikert, Deloria, and Lawsor (1974). Children in these studies showed dramatic initial gains of up to 15 points in IQ scores. Later follow-ups (Stanley, 1972) indicated that while several structured and semistructured cognitive programs could produce such gains, much of it disappeared after the child had been in elementary school for several years. Stanley concluded by seriously questioning the long-term effectiveness of early intervention programs. Jensen (1969) criticized large-scale compensatory education programs designed for children of any age, stating that such programs failed to permanently increase IQ scores or scholastic performance because individual differences are determined more by genetics than by the environment. He stated that deprived environments can stunt but enrichment cannot go beyond prevention of stunting. He concluded that the genetic factor rather than the stunting factor was the major cause of poor achievement in low SES students. Thus, to Jensen, most of a child's achievement level is predetermined at birth and is not subject to environmental remediation.

In a critique of 64 reviews of early intervention involving 1,027 articles, researchers at Utah State University (Bush, White, Casto, & Shearer, 1982) found that the two conclusions drawn most often by reviewers about outcomes in relation to early intervention were that the earlier the age of intervention, the greater the developmental gains (14 reviewers) and that early intervention is effective if developmentally appropriate (8 reviewers). However, in looking at measurable outcomes, the critiques reported that 15 reviews concluded that early intervention gains eroded rapidly, while only five reviews concluded that they could be maintained.

In an effort to find out if the deterioration of intervention effects could be avoided, Lazar, et al. (1977) spearheaded a consortium to study the longitudinal effects of early childhood programs with low-income children. Lazar and his associates studied children ages 9 to 19 who had been enrolled in 11 preschool intervention projects of proven short-term effectiveness. They found that graduates of these programs as they grew up were retained in grades less often, were less likely to be assigned to special education classes, scored higher on math achievement tests in upper elementary grades, and had higher vocational aspirations than the control group of low-income children not enrolled in preschool intervention. In another study of long-term effectiveness of early childhood programs, Bronfenbrenner (1975), reviewed seven programs for low-income families. He concluded that early childhood intervention which involved parents directly in the actual education process had long-term effects in score-gain retention and in positive parental attitudes toward the child. These gains were not maintained in preschool poverty programs which worked directly with children without parent education. In this study, Bronfenbrenner also urged economic and social support for disadvantaged families, stating that families that had to worry about basic survival would not have the psychological or physical resources to readily become involved in the education of their children. According to a review of literature reviews by Bush and White (1983), more reviewers cited degree of parent involvement as a key treatment variable (23 reviewers) than any other variable listed.

Overall, there is strong agreement about the short-term effects of preschool intervention but disputed findings on how much these gains deteriorate over time. Involvement of parents in the educational process appears to prevent some of the long-term loss.

Effects of General Head Start Intervention

Head Start is one kind of early childhood intervention. It differs from many other preschool education programs by having a developmental framework of health, nutrition, and social services as well as an educational component. It also includes a commitment to locally-designed options and parent involvement in policy making which other intervention models generally do not have.

The first large-scale longitudinal research conducted on Head Start was done by Westinghouse (Cicirelli et al., 1969). It was almost the downfall of the program (Zigler & Valentine, 1979). Published at the same time as Jensen's (1969) controversial articles, the Westinghouse study concluded that Head Start programs produced initial test score gains for children which disappeared once a child entered elementary school.

Later research provided more encouraging results. In a review of 59 research projects done between 1967 and 1977 (Mann, Harrell, & Hurt, 1977), the overall results for Head Start programs were shown to be similar to those cited above for other early intervention programs. These included findings that Head Start has positive initial impact in cognitive growth, that Head Start students were less likely than non-Head Start peers to be placed in special classes or held back in elementary or high school, that Head Start has positive effects on children's health and social development, and that it improved parental attitude about their children and increased the amount of time parents spend with children.

In a second review (Aitken, Hubbell, & Jones, 1982) found that "almost all studies showed significant gains over the operating year for children in Head Start on intelligence and achievement measures" (p. 6). The researchers also found that while Head Start students improved, they did not catch-up to middle-class performance on cognitive levels; that most studies show maintenance of cognitive gains through elementary school, but not through high school; and that, Head Start students were less likely to be retained, to dropout, or to be placed in special education classes. Thus, the pattern of program effectiveness in Head Start is similar to that in general for early childhood intervention programs.

Effects of Mainstreaming and Early Childhood Intervention

Public Law 94-142 states that each child must be educated in the least restrictive environment. This is usually interpreted as requiring mainstreaming, or integration into the normal group, except where the specific handicapping condition requires a more restricted environment. Research on mainstreaming with older children indicates that one problem can be an increase in peer rejection and negative socialization effects for handicapped children integrated into regular classrooms (Wynne et al., 1975). There has also been research indicating that teacher intervention to promote social acceptance in mainstreaming can be effective in increasing positive social interactions (Wynne et al., 1975). Research done with parents of kindergarten children (Turnbull, Winton, Blacher, & Salkine, 1982) indicated acceptance of the concept of mainstreaming, but parents of of both handicapped and nonhandicapped children expressed concerns about the effective implementation of mainstreaming in a typical elementary classroom if no special training or intervention program is planned.

In looking at the effectiveness of mainstreaming programs, Wynne, Ulfelder, & Dakoff (1975) found that preschool mainstreaming had a more positive effect on retarded children with environmental deprivation than those with organic impairment. Using a control-group evaluation model, Cooke, Ruskus, Apolloni, and Peck (1981) tested the effect of mainstreaming on both handicapped and nonhandicapped children. They found that integration was effective only if there was: (a) an intensive planned intervention focused on social interaction between handicapped and nonhandicapped children, and (b) an educational program geared to meeting the varying cognitive abilities of all children in the class. Casual, nonplanned mainstreaming resulted in some losses by both handicapped and nonhandicapped children when compared to the control groups in a non-mainstreamed program. DeWeerd and Cole (1976), in a study of 688 graduates of the Handicapped Children's Early Intervention Program, found signs of longitudinal success, particularly in the high placement rate of children in regular classrooms. There was not, however, an adequate control group with which to compare the handicapped children.

Thus, mainstreaming handicapped children can result in an effective education program for both handicapped and nonhandicapped children, but requires more individualized planning and focused educational and social intervention by the teacher than non-mainstreamed classroms.

Effects of Head Start

Handicapped Intervention

The Status of Handicapped Children in Head Start Programs (1980) reported that Head Start programs have made considerable efforts to comply with the mandate to find and serve handicapped children. They also found that handicapped children were successfully mainstreamed into the Head Start programs. Many of the mainstreamed handicapped children in Head Start programs showed increases in playful and positive peer interaction, and there were gains in most developmental skills. Small class size, lower handicapped/nonhandicapped ratios, the experience of teachers with handicapped children, and amount of time in mainstreaming situations were all positively related to developmental gains and increased social interaction.

The Review of Head Start Research Since 1970 (Hubbell, 1983) supports these findings regarding handicapped children in Head Start. The review cites the second volume of the Applied Management Study which surveyed handicapped children in 59 Head Start programs and found 66 percent to be socially integrated, 23 percent somewhat integrated, and 10 percent socially isolated. The Applied Management Study also described a comparison study between handicapped children in Head Start and handicapped children in other preschool programs. The major differences in the content of the Head Start and non-Head Start programs were that Head Start involved mainstreaming handicapped children into the general Head Start population whereas the other preschool handicapped programs involved a more individualized program which enrolled only handicapped children. The two groups (Head Start and non-Head Start) were matched for amount and kind of handicap. This study found that Head Start children with speech problems scored significantly higher than their non-Head Start peers on all subscales of the Alpern-Boll Developmental Tests, except social skills. It also found that Head Start children with learning disabilities or emotional disturbances scored higher than matched non-Head Start handicapped children on academic skills. There were no significant differences between the two groups on the other subscales of the Alpern-Boll. The findings generally indicate that Head Start handicapped children can do as well as comparably handicapped peers placed in non-mainstreamed handicapped programs.

Research Techniques

The numerous statistical and logistical problems facing any research project involving human behavior are multiplied many times over when developmental research with young children who are also handicapped is attempted. Because of the expense and problems associated with control-group research with needy young children, most large-scale research in early childhood has used a combination of quasi-experimental control groups and normed tests. Campbell and Boruch (1975) pointed out some of the distortion caused by the attempts to use nonrandomly-selected control groups and the kinds of influences this has had on evaluations of compensatory programs. Campbell and Boruch were particularly concerned that many of the statistical assumptions in quasi-experimental research designs tended to overestimate expected scores in the low ranges and thus resulted in undervaluing results of compensatory programs which worked with disadvantaged children.

Another limiting factor in research with preschool handicapped children was the passing of Public Law 94-142 which in effect eliminated the option of nontreatment of handicapped children and limited control group research projects to comparison between kinds of treatment rather than between treatment and nontreatment. May (1979) pointed out this factor and several others in describing the increasing difficulty of doing research on handicap intervention projects which measure change in more than one subject. He also pointed out that few normed test developers have handicapped children included in their normative populations and that test scores for handicapped or deprived populations may not follow the patterns exhibited by the rest of the population. After listing difficulties in assessing the relative progress of handicapped children with any of the available statistical techniques, he advised program directors to be aware of their research options and to make optimal but imperfect choices based on their program's resources and needs.

Title I Evaluation Models

Partially because of the research and evaluation difficulties listed above, reported results from national educational intervention programs were often subjective, inconsistent, and sometimes nonexistent. In 1974, an increasingly frustrated Congress mandated the Department of Education to develop evaluation models which could be used nationwide as a program evaluation and reporting system to adequately inform Congress as to the effectiveness of individual and aggregate Title I programs. Title I programs (now called Chapter I) are educational intervention programs federally funded through Public Law 93-380 which are designed to provide disadvantaged nonachieving children with compensatory educational services. The Title I evaluation models were developed under contract by the RMC Research Corporation and then reviewed by a Policy Advisory panel (made up of two Title I parents, several state Title I evaluators, a representative from the National Council on the Education of Disadvantaged Children and a representative from the Chief State School Offices), and a Technical Advisory Panel (make up of five nationally known authorities on measurement and evaluation) (Tallmadge & Wood, 1980b). The Title I Evaluation System includes three acceptable evaluation models: (a) Model A, a norm-referenced model; (b) Model B, a control group model; and (c) Model C, a regression model. Since research designs using control groups have become increasingly difficult with handicapped children, the existence of two acceptable evaluation alternatives would be important and relevant to programs who wish to evaluate intervention effectiveness.

"The focus of all the models is to obtain as clear and unambiguous an answer as possible to the question, 'How much more did pupils learn by participation in the Title I project than they would have learned without it?'" (Tallmadge & Wood, 1980b, p. 2). In Model A, the norm-referenced model, students are selected for the treatment group through any approved selection technique (test scores, teacher selection, grade selection), and are given standardized or locally-normed tests. Scores of individual students who have completed pre and posttests are converted to Normal Curve Equivalents (NCE's) and a group pre- and posttest mean calculated. The treatment posttest mean score is then compared with that of students in the norm group who scored at the same percentile. Any change in NCE for the treatment group between pretest and posttest is assumed to be due to the effects of the Title I program.

Because of the tendency for high and low scores to regress to the mean on subsequent testing, one assumption of Model A is that the test used to select the population not be the same as the test used to document outcomes. However, in cases where this dual usage is unavoidable, Tallmadge and Wood (1980b) developed a correction formula to be used for computing adjusted pretest means where the same test is used for both selection and pretesting.

Model B, the control group model, requires that Title I students be compared to a locally-selected group of comparable students. While randomized selection is seldom feasible in Title I projects, most school districts contain non-Title I schools from which comparison subjects matched on relevant factors can be selected. Tallmadge and Wood (1980b) stated that project impact is to be measured by the difference between the intervention group and control group's mean posttest scores after statistical treatment to control for pretest differences. They also stated that in situations where the control and intervention group's pretest mean scores differ by more than four NCE's it would be advisable not to use Model B for program evaluation.

In Model C, the regression model, all students in the Title I eligible school are given a pretest. The intervention group consists of all students, and only those students who score below the selected cut-off score. The mean posttest score of this treatment group is compared to an estimated mean posttest score derived from the regression line formula developed by Tallmadge and Wood (1980b) given below.

Regression Line Formula

No-Intervention Expectation $= \bar{Y}_c - (\bar{X}_c - \bar{X}_p)\, r_{xyc} \frac{s_{yc}}{s_{xc}}$

	Intervention	Non-Intervention
Mean pretest score	$\bar{X}_p$	$\bar{X}_c$
Mean posttest score	$\bar{Y}_p$	$\bar{Y}_c$
Selection/pretest standard deviation		s_{xc}
Posttest standard deviation		s_{yc}
Pretest-posttest correlation of nonintervention group		r_{xyc}

Tallmadge and Wood (1980a) stated that project impact is measured by the difference between the actual mean posttest score of the treatment group and the estimated mean posttest score.

Research on Title I

Evaluation Models

Generally, research done to compare intervention group gains has indicated similar, if not equal, results for the three evaluation models. House (1979) conducted a study of Models A and C using data already collected from the St. Louis School District's Title I program. He found no significant differences between the two models but added that the small or nonexistent score gains made by the various intervention groups in the St. Louis study made it mathematically unlikely that there would be much variation between evaluation models. Tallmadge and Wood (1980a) compared program gains as measured by the three models using after-the-fact analysis of data from three grades in two Title I school districts. Their conclusion was that the three models yielded similar estimates of Title I gains. They found that there were no significant differences between the three methods in any of the three grades studied, although Model A tended to produce a small positive bias on the order of one NCE. In a third study, Hardy (1979) compared school districts who were implementing Model A evaluations with school districts who were implementing Model C evaluations and found that reported intervention groups gains were significantly higher in school districts choosing the norm-referenced model. However, since none of these schools used both models, he was unable to determine if these differences resulted from site differences in programs choosing the two models, in procedure differences connected with present implementation of the two models or in inherent differences in the evaluation models themselves.

Much of the research on the Title I evaluation models has been done after-the-fact, either by placing existing test scores into Title I formulas or by computer simulation with existing data and score distributions. The primary danger in this method of analysis is that it is difficult to ensure and validate that the operational requirements connected with the specific models being analyzed have been met. In House's (1979) research on Model A and C evaluations, he analyzed data already collected from the St. Louis School District's Title I program. There was nothing in the written report indicating that the intervention group for the Model C analysis conformed to the Model C requirement of the exclusionary below cut-off selection criteria; indeed, the description of how the Model C testing scores were obtained (universal school achievement testing program) made this seem problematic given the heavy pressure within most school districts to use teachers' judgments and parental request as supplementary criteria to add or delete students from special programming. Eldred and White (1982) in an analytical review of research on Title I evaluation models questioned the validity of results of several studies which used data after-the-fact, doubting whether these studies would have the consistency or control of test-relevant variables to provide definitive answers regarding the validity of the evaluation models.

Many of the research studies on the Title I evaluation models have stated concerns with the theoretical assumptions underlying Model A. One area of concern is the assumption that the treatment group resembles students in the norming group when, by definition, the treatment group consists of a specifically-defined subpopulation in compensatory education. Linn (1980) and House (1979) both discussed this concern and questioned the validity of results which assume that scoring gains in a specialized population would equal those in a more general norming population.

A second area of concern is with the unproven assumption of equipercentile scoring, or, the assumption that the percentile score of the intervention group on the posttest would equal its percentile score on the pretest if it received no intervention. House (1979), Linn (1980), and Mandeville (1978) have all questioned the validity of results utilizing Model A because of its reliance on this unproven assumption. Wood (1980) examined the issue of the equipercentile assumption in Model A. She found that the NCE status of untreated students from the bottom third of the population did show gains averaging about one NCE in their NCE status over a year's time (attributed to regression) and, thus, projects which selected from the bottom one-third using Model A would tend to overestimate gains by this amount.

The analytical review of research on Title I evaluations done by Eldred and White (1982) reviewed most of the research cited above and concluded that methodological problems in relation to unproven assumptions and failure of most research to demonstrate that it had complied with Model assumptions, made it impossible to ensure definitive answers on the validity or comparability of the evaluation models. They concluded that while Models B and C are superior theoretical models, Model A was more likely to be used because it was less expensive and easier to implement. They also stated that even

Model A, with its unproven assumptions and possible overestimation, was probably better for a nationwide evaluation system than the random and uncontrolled systems used previously.

Overall, the research done thus far on the Title I evaluation models has raised as many questions as it has answered. While there is some evidence that model options demonstrate similar gains, there are also indications that Model A shows a small positive bias. Many researchers have been concerned with the assumptions underlying Model A and there needs to be more research testing the validity of these assumptions. In addition, there is a great need for planned research designed to compare model findings, rather than the present reliance of analysis of after-the-fact data where methodology cannot be validated or controlled.

Summary

The literature reviewed suggested that early intervention programs have had some initial success, but that program evaluation particularly of handicapped children, has many theoretical and practical difficulties.

The literature also suggests that Title I evaluation models may be practical and usable options for compensatory education, but that more research needs to be done that is specifically designed to validate model assumptions and model comparability.

CHAPTER III

METHODS AND PROCEDURES

Population and Sample

The target population for this research project consisted of educationally-handicapped children enrolled in Head Start. Two hundred and three children in the Jackson County, Oregon, Head Start Program were initially tested. Those children who tested below the preselected cut-off scores were placed in the intervention treatment group and received the general Head Start program plus the project intervention programs. Those children who scored above the cut-off scores and who did not meet the criteria in the Head Start Definitions for the Handicapped (see Appendix A) were placed in the nonintervention group and received only the general Head Start program. Those children who met Head Start's criteria for the handicapped, but did not score below the cut-off level, were removed from consideration in either the intervention or nonintervention groups of this project.

One hundred and forty-nine of the initial 203 children were included in the final intervention and nonintervention groups. The 54 dropouts included two chidren who were diagnosed as handicapped (one cystic fibrosis, one hyperactive) but whose test scores were above the intervention group range, two non-English speaking children whose scores on the battery were judged to be invalid, one child who refused to complete the test, one child whose parents asked that he not be tested, and 48 children who moved from the area before the project was completed. One hundred and forty-three children of the 149 in the completed project came from low-income families whose incomes were below the Head Start poverty guidelines (see Appendix B). The six remaining children were handicapped children from middle-income families.

The following two tables give the age and sex breakdown on the pre- and posttests for the intervention and nonintervention group children who completed the project.

Table 1

Pretest Data

	Intervention		Nonintervention	
Age	Male	Female	Male	Female
3	2	3	10	12
4	13	12	43	21
5	8	4	9	11
6	1	-	-	-
Total	24	19	62	44

Table 2

Posttest Data

	Intervention		Nonintervention	
Age	Male	Female	Male	Female
3	-	-	4	5
4	8	10	28	20
5	10	6	24	14
6	5	3	6	5
7	1	-	-	-
Total	24	19	62	44

The intervention group consisted of 43 children. Thirty-seven of these 43 children were diagnosed as handicapped according to one of the ten defined federal categories. Six of the 43 were children who scored below the preselected cut-off levels but who did not meet Head Start criteria for the definition of handicapped. Table 3 gives the specific handicapping categories for children in the intervention group. All 43 of the children in the intervention group participated in both the intervention program and the regular Head Start program.

Table 3

Primary Handicaps of the Intervention Group

Primary handicap	Number of children
Speech and language	18
Learning disabled	7
Retarded	5
Health impaired	1
Vision impaired	1
Hearing impaired	1
Emotionally disturbed	3
Physically handicapped	1
Total handicapped	37
Borderline-language Disabled but not handicapped	6
Total intervention group	43

The 106 children in the nonintervention group participated in the pre- and posttesting and the regular Head Start program, but not the project-planned intervention. The nonintervention group contained no handicapped children.

This research project applied a procedure designed to estimate gains made by Title I students to measure gains made by a group of educationally-handicapped children in a Head Start program. One assumption in making this comparison is that the Head Start population will not be different from the Title I school population in any significant way which would effect evaluation outcomes. There are several crucial variables in comparing Title I students and the Head Start intervention group. Both come from lower socio-economic levels although the mean family income for Head Start students tends to be slightly lower than that for Title I students. Both groups involve some degree of functional delay. The criterion of one year delay which was used for selection of children for the intervention group in this project is within the nine month to two-year delay range used by local Title I projects in Jackson County. However, since the Head Start children are younger, the one-year delay represents a higher percentage of delayed functioning than that exhibited by children in the Title I population. If Head Start intervention students are functioning at a relatively lower level than Title I students, that may mean that Head Start scores would demonstrate a higher regression factor in the Model A analysis than those of the Title I population. Wood (1980) stated that for populations functioning below the 30% level gains in Model A evaluations will be overestimated by one NCE because of the regression factor. The mean percentile on pretests for the intervention group in this project varied between 13 and 22%, putting them well below the 30% cut-off on all three pretest mean scores. Thus, if Wood's calculations are correct, gains by the intervention group in Model A analysis may be overestimated by as much as one NCE.

Instrumentation

The following instruments were utilized for the assessment of children in this project.

1. The Peabody Picture Vocabulary Test (PPVT). The PPVT (Dunn, 1965) is a measure of receptive vocabulary skills. It has been standardized on 4,012 children aged 2 years, 3 months to 18 years, 5 months (p. 27). A child's raw score can be converted to a mental age and to a percentile ranking. Two forms of the test are available. Reliability coefficients for the PPVT using raw scores for Forms A and B have been obtained at each age level (Dunn, 1965). Alternative form reliability coefficients for children in the age range of this project varied from .73 at age 5 to .81 at age 3 years, 6 months (p. 30). Interestingly, the reliability coefficients for preschool children (notoriously, the elast reliable age group in test-related correlations) were slightly higher than those scores reported for children ages 6 through 10. The National Day Care Study (Ruopp, Travers, Glantz, & Corden, 1979) found the PPVT was one of the most reliable measures for assessing preschool children.

Considering the PPVT as a measure of a hearing vocabulary, content validity was built into the test by pooling all words in Webster's New

Collegiate Dictionary whose meaning could be illustrated and selecting those items which, statistically, were the best differentiators when given children through the different age ranges (Dunn, 1965). When the inference is made that the PPVT is also a measure of verbal intelligence, then validation is based on the correlation between the PPVT and other standardized intelligence tests (.82 to .86 between the PPVT and the Stanford-Binet) and on the correlation between the PPVT and other standardized intelligence tests. (The PPVT correlates with the Wechsler Intelligence Scale for Children, Verbal Scale [.67].)

The PPVT is easy to administer and to score. It is an appropriate test to give to preschool children because it is quick to give and maintains subject interest.

2. Developmental Test of Visual-Motor Integration (VMI). The VMI (Berry & Buktencia, 1967) is a geometric form-copying task designed for children 2 to 15 years of age. The test stimuli, a series of geometric forms presented in order of increasing difficulty, are contained within the same test booklet in which the child's responses are entered. Scoring criteria for each form are represented on a "pass/fail" basis in a separate manual. Reliability and validity information, norms for converting raw scores into developmental age equivalencies and percentiles, and suggestions for remediation are also contained within this manual.

The VMI yields information on perceptual-motor development and is designed to predict reading and writing readiness. The 1981 norm scores were standardized on 3,090 children (p. 15) and the test manual reports test-retest test score reliability of .83 for boys and .87 for girls. In terms of validity data, there is a reported .89 correlation between VMI scores and chronological age. An additional study found a .50 correlation between VMI scores and first-grade reading achievement (Chissom, 1972).

The VMI presents several advantages over other pencil and paper copying tests. It requires no special qualifications for administering or scoring. It is attractively and interestingly presented so as to maintain children's interest. The scoring is objective. Finally, developmentally sequenced suggestions for remediation steps are presented in the manual, allowing for ease of remedial programming based on test performance.

3. Carrow Elicited Language Inventory (CELI). The CELI (Carrow, 1974) is designed to assess auditory comprehension of various linguistic categories. The CELI was standardized on a population of 475 children (p. 8) in Texas. Sentence items were originally administered to 65 children ages 3.0 to 7.11. Responses were analyzed for information about the grammatical functioning capability of young children. Reliability of the CELI was computed on a population of 25 children on a test/retest basis. A coefficient of .98 was obtained. A validity study was reported by Cornelius (1974). She found that CELI scores discriminated between previously-diagnosed language disorder children and normal children. The CELI scores reflected a significant difference in total language score between two groups.

The CELI consists of 52 sentences which children are asked to repeat. A child's score consists of the number of errors, substitutions, and omissions which he makes. Developmental and percentile scores can be computed. Scores can also be broken down by grammatical groups to assist in remediation.

4. The Developmental Indicators for Assessment of Learning (DIAL). The DIAL is a preschool test of gross motor, fine motor, concepts, and communication skills (Mardell & Goldberg, 1972). The test was standardized on 4,356 children using a stratified sample to balance children on sex, demographic setting, race, and socio-economic status. The technical manual reports test/retest coefficients in the .90's for children ages 3 to 5. In The Eighth Mental Measurements Yearbook (1978), J. Jeffrey Grill cited as strengths of the DIAL its criteria for item selection and its clear instructions. The review article cited a study in which 249 children previously tested by the DIAL were retested several years later in the four subtest skill areas. The DIAL subtests demonstrated .45 to .73 correlations with later subtest performance; thus, demonstrating acceptable long-range predictable validity.

The following instrument was utilized in the development of the treatment program for the intervention group.

5. The Curriculum and Monitoring System (CAMS) Expressive Language, Receptive Language, and Motor Tests. The CAMS (Casto, 1979) system was developed as part of a Bureau of Education for the Handicapped Preschool Demonstration Project at the Exceptional Child Center at Utah State University. The process followed in developing the materials included several steps. The critical skills in each curriculum area were first identified through an exhaustive literature search. They were then critically viewed by curriculum experts who are

knowledgeable in the specific skill areas and who were able to describe skills in behavioral terms. Next, the skills were stated as behavioral objectives and were placed in hierarchial order. Then, criterion-referenced placement tests were developed to assess the specific skills identified in each curriculum area. These were the tests used with the intervention group to develop educational treatment plans.

Local Reliability and Validity Information on Instrumentation

In addition to the published reliability and validity information given in the preceding sections, several analyses were also done on test reliability and validity within the population studied.

1. Reliability. Tables 4, 5, and 6 show pretest/posttest correlations in the intervention, nonintervention, and total project population.

Table 4

Test Correlation Matrix--
Entire Project Population

Pre/post Test	Pre PPVT	Pre VMI	Pre CELI	Pre CAMS
Post PPVT	.77	.54	.54	.62
Post VMI	.58	.78	.55	.65
Post CELI	.49	.43	.78	.49
Post CAMS	.50	.59	.54	.77

N = 149

Table 5

Test Correlation Matrix--

Intervention Group

Pre/post Test	Pre PPVT	Pre VMI	Pre CELI	Pre CAMS
Post PPVT	.85	.61	.44	.65
Post VMI	.64	.84	.44	.68
Post CELI	.49	.28	.70	.36
Post CAMS	.53	.72	.42	.85

n = 43

Table 6

Test Correlation Matrix--

Nonintervention Group

Pre/post Test	Pre PPVT	Pre VMI	Pre CELI	Pre CAMS
Post PPVT	.70	.47	.55	.57
Post VMI	.54	.76	.63	.62
Post CELI	.41	.45	.75	.52
Post CAMS	.46	.53	.59	.71

n = 106

As the tables indicate, pretest/posttest correlations between the same test were generally between .75 and .78, acceptably high when considering the 7-month time differential.

2. Validity. The testing battery consisted of two paired measures of expressive language (the CELI and the DIAL Motor), two paired measures of receptive language (the PPVT and the DIAL Concept), and two paired measures of fine motor (the VMI and the DIAL Fine Motor). If these tests are capable of measuring the selected skills, then correlation should be higher between the pairs than between any nonpaired tests given in the testing battery.

Table 7

Correlation of DIAL Subtests to Test Battery/Pretests

r of DIAL/ Battery/pretests	CAMS	VMI	PPVT	CELI
DIAL gross motor	.69*	.62	.55	.37
DIAL fine motor	.61	.66*	.64	.41
DIAL concepts	.63	.57	.74*	.60
DIAL communications	.51	.60	.57	.54*

n = 43

Three of the four pairings correlate as predicted. However, the CELI's highest correlation is with DIAL concepts rather than

*Paired Correlations

communication. Examining test content, the DIAL communication test contains five subsections. One of these subsections requires the same behavior which the CELI measures. The other four subsections contain communication skills which may not correlate so highly with the CELI. The other three paired correlations are the highest matching within the group, ranging from .66 to 74. The PPVT/DIAL concept correlation is even higher than the PPVT pretest/posttest correlation for the intervention group.

A check to assess whether the instrumentation measures the appropriate skill is to see if children selected for intervention by the testing battery (PPVT, VMI, CELI, and CAMS-Motor) were the same as children who would have been selected had the DIAL subtests been used as the selection instrument, since the battery and the DIAL subtests were presumably measuring similar skills. Table 8 shows the relationship of selection by the battery and the DIAL

Table 8

Selection Table

A) Children selected by battery and DIAL	B) Children selected by one not by other	C) Children selected by neither
31	20	98

Percentage of children placed in same group by both tests = $\frac{A + C}{A + B + C}$

Thus, 87% of children would have been placed in the same group by both the battery and the DIAL pretests.

Another way of examining this relationship is to find the phi coefficient for the correlation between selection by the battery and selection by the DIAL subtests. Table 9 illustrates the table used to assign numbers to the variables in phi.

Table 9

Selection Table

		Children selected by battery for intervention		
		Yes	No	
Children who would have been selected by DIAL for intervention.	Yes	A 31	B 12	43
	No	C 8	D 98	106
PHI =.67		39	110	149

The phi correlation between the two testing batteries is .67 or a moderately high positive correlation.

The third measure of local validation is to determine if the testing battery selected the appropriate children in Head Start for the intervention program. Using the predetermined test battery score cut-off level, 37 of the 39 handicapped children in the program were assigned to the intervention group. Both the two handicapped children who scored above the cut-off level had handicapping conditions which effected noncognitive areas and thus were not appropriate candidates for the intervention group. Ten of the 39 in the intervention group were diagnosed as handicapped because of their scores on the project's testing batteries. The other 29 children in the intervention group all had handicap diagnosis which were done separately and independently from all project testing and which indicated significant cognitive delay (one year or more). Thus, the testing battery was successful in selecting all 29 of the children who were selected independently by other evaluation systems.

Design

Children in the project were attending Head Start at six different sites scattered throughout Jackson County. Each site designated a team made up of administrators, teachers, and aides to administer the Assessment of Learning Test (DIAL). Each member of the testing team received six hours of training in administering the DIAL. The DIAL is divided into four subsections; each member of the testing team was assigned to the one or two subsections most related to the tester's background and expertise. Children were given the DIAL in October and May. In most instances, children completed all four DIAL subtests in one session, although scheduling logistics required two sessions for about one-fifth of the children.

The second group of tests, the PPVT, the VMI, the CELI, and the CAMS Motor Test, was given to about 90 percent of the children in a one-week period in October and again in May at the six different center sites. Ten percent of the children were absent on one of the testing dates and were tested the following week. The tests were administered by a testing team which included one speech therapist, one psychologist, and two program supervisors. Each team included two testers from outside the program and two program staff. It took approximately 50 minutes for each child to complete the entire testing battery. Group assignment was made after the pretest. All but one of the data collectors in the posttest were unaware of children's group assignment.

Intervention

The CAMS curriculum package used in this project includes 15 objectives in receptive language, 41 objectives in expressive language, and 98 objectives in motor development. Each of the curriculum packages is printed in easy-to-use block style design and bound in a notebook. Each objective is printed on an individual data sheet. Printed at the top of each sheet are the name of the program, the objective number and its name, and the materials needed in teaching that objective. There is also space for entering the student's name and the date on which the activities on the form were begun.

Each sheet is then divided into the following four sections or steps.

1. The Step Statement (SS) tells exactly what skill the student will learn at this step of the program.

2. The Teaching Procedures (TP) tells exactly what the teacher must do to teach the skills described in the Step Statement. It may also state what to do if the student makes a mistake.

3. The Trial Criterion (TC) tells exactly what the student must do to receive a "yes" on a trial.

4. The Step Criterion (SC) tells how many "yeses" a student must get before going to the next page of the objective.

Data about a child's performance on the program including the percentage of correct and incorrect responses, the response rate, the total number of trials, and the total number of sessions are recorded by the teacher on a data summary sheet which is used in deciding when the child should progress to the next task or skill. As the child moves through the developmental sequence from skill-to-skill, it is always possible to know exactly which task is being taught and what progress is being made.

The 43 children selected for the intervention program were given the CAMS criterion-referenced test which matched their area or areas of delay. Those showing delay on the PPVT took the CAMS Receptive Language Test, those showing delay on the CELI took the CAMS Expressive Language Test, and those showing delay on the VMI took the CAMS Motor Test. An individualized plan was developed for each child using information from the following sources: (a) the scores derived from the CAMS criterion-referenced tests, (b) the education deficits indicated by the PPVT, the CELI, the VMI, and the DIAL scores, (c) general behavioral observations by teachers and parents, and (d) evaluation information and recommendations included in the professional diagnosis for the 27 independently-diagnosed handicapped children. Table 10 presents the intervention group categories by types of handicap and types of intervention. As may be noted, children with multiple areas of delay received more than one type of intervention.

Table 10

Program Planned Intervention

Primary Handicap		Receptive Language Interven-vention	Expressive Language Interven-vention	Motor Inter-vention	Other Inter-vention
Speech (including the 6 borderline handicapped)	24	15	17	-	1
Learning disabled	7	-	1	7	1
Retarded	5	5	2	4	3
Health impaired	1	1	-	-	-
Vision impaired	1	1	-	1	1
Hearing impaired	1	1	1	-	-
Emotionally disturbed	3	3	-	1	3
Physical handicap	1	-	-	1	1
Total		26	21	14	10 = 71

The intervention program was started in mid-November and continued through the end of April. While individual intervention time varied depending on the needs of the child, the average planned program had children receiving one hour and fifty minutes of intervention a week.

The minimum planned intervention time for any child was 45 minutes per week; the maximum, 6 hours per week. The interventions included individual and small group classes, home visits, and, less often, intervention in the classroom. The program was administered by the Handicap Coordinator, the Speech Therapist, two aides, the child's parents, the child's classroom teacher, and, occasionally, outside resource professionals.

The individual and small-group classes were directed by the Speech Therapist and the Handicap Coordinator and implemented by themselves and the two handicap aides. Aside from the activities of the intervention program itself, the children in the intervention group were mainstreamed into the 13 classrooms of the Jackson County Head Start program and participated in all usual Head Start activities.

Data Collection

The sequence of the steps outlined in the User's Guide for implementation of the Title I evaluation models (Tallmadge & Wood, 1980b) was followed in the data collection and analysis steps for comparing Model A and Model C.

Results in Title I evaluations are expressed in NCE (Normal Curve Equivalent) units. The NCE is a standard scoring system with a mean of 50 and a standard deviation of 21.06. It was purposely created to resemble percentile scoring with a standard scoring correction factor. Scores are determined by converting from raw scores to percentiles to NCE s.

Sequence of steps for data analysis as outlined in Tallmadge and Wood (1980b).

Model A

1. Select the project participants.

(a) All 203 children in Jackson County Head Start were given a testing battery in October consisting of the PPVT, the CELI, the VMI, and the CAMS Motor.

(b) Children who scored functionally one year below their chronological age on the PPVT, the CELI, or the VMI and CAMS were placed in the intervention group.

2. Select a nationally-normed achievement test which is an adequate measure of the functional level of students in the project.

(a) All tests used in the testing battery were nationally normed.

(b) Tests for preschool children are usually labelled developmental rather than achievement because preschool children tend to gain skills in a developmental sequence rather than the academically-based achievement learning or older children involved in formal education. Thus, the tests used in this study were developmental tests.

(c) Tallmadge and Wood (1980b) suggest that the mean raw score of the group should fall between a third and three-quarters of the highest possible test score. Most of the tests given in this project were individual developmental tests with varying starting and ending points, depending on the age and skill level of the student. If we assume that a fair equivalent of the highest possible score would be the highest score of any student in the total testing group of 143, then scores on the PPVT, the CELI, and the VMI all met this 1/3 to 3/4 range.

3. Pretest the participants within two weeks of normative testing.

(a) All 203 children in Jackson County Head Start were given the DIAL tests in mid-October, and the PPVT, the CELI, and VMI in late October.

(b) There is no practical way of meeting the norm-test date criteria in a research project with preschool children which involve more than one test because most preschool tests are not normed during any particular time in the calendar year, and norming dates for some preschool tests are simply not available. Norm date testing is important for older children because of the enormous significance of the context of the school year. However, for preschool children, chronological age is a much more crucial variable and developers of tests for preschool children respond to this by providing age equivalents (mean age at which children attain set score levels) and by establishing short chronological age norm intervals (two-month intervals for the PPVT at the preschool level).

4. Score the tests and record the pretest scores.

All test scores from tests given in this research project were scored within a week and recorded on a master list. Raw scores and age equivalency scores were both records.

5. After the project, posttest the project participants.

(a) The project intervention was done between November and April.

(b) All children remaining in the project (143) were posttested with the VMI, PPVT, CAMS, and CELI in early May.

f. Score the tests and record the posttest scores.

(a) All posttest scores were scored within a week and were recorded on a master list. Raw scores and age equivalency scores were both recorded.

(b) Thirty-two of the posttest DIAL scores of the intervention group (15%) were ceiling scores. The User's Guide (Tallmadge & Wood, 1980b) states that significant numbers of ceiling scores will result in incorrect estimates of project gains in all three project models. Because of the unsuspected ceiling scores, the DIAL could not be used for a measurement instrument for Model A.

(c) Fortunately, Tallmadge and Wood (1980b) have developed an adjustment formula which can be used in those situations in which the pretest is also the selection test. Table 11 gives this formula.

Table 11

Tallmadge and Wood (1980b)

Adjustment Formula

$$\overline{X}_p' = \overline{X}_p + (1-r_{xx})(\overline{X}_t - \overline{X}_p)$$

$\overline{X}_p$ = Mean score of intervention group on selection pretest.

$\overline{X}_t$ = Mean score of total group (from which the intervention students were selected) on the selection pretest.

r_{xx} = The test-retest reliability for the total group.

$\overline{X}_p'$ = Adjusted mean score to be used where selection measure is also pretest measure.

The pretest scores on the VMI, the PPVT, and the CELI were adjusted according to this formula and used as measurements of pretest status. The unadjusted mean pretest scores were used as selection instruments.

7. Convert scores to NCE's and compare pretest and posttest means of the intervention group.

(a) Convert pretest and posttest scores on the VMI, the PPVT, and the CELI to NCE.

(b) Calculate the intervention group's mean NCE on the VMI, the PPVT, and the CELI.

(c) Convert the pretest mean NCE score of the intervention group on the VMI, PPVT, and CELI using the adjustment formula in Table 11

(d) Compare the pre- and posttest status by comparing mean pretest NCE scores with mean posttest NCE scores. Tallmadge and Wood (1980b) state that a difference of six NCE represents "educational significance."

(e) Test the statistical significance of the difference using a t-test at a .05 level of confidence.

<u>Model C</u>

1. Select the nationally normed test(s) to be used on the evaluation instrument.

The VMI, the PPVT, and the CELI were selected as evaluation instruments.

2. Administer the test(s) selected to the group of students from which the intervention group is to be selected.

(a) One assumption crucial to Model C is that the (nonintervention) and (intervention) group's mean scores will be relatively stable. Tallmadge and Wood (1980b) state that this stability requires a sample size of at least 100 and a nonintervention and intervention group containing at least 30 students each. The total sample size in this project was 149; the intervention group numbered 43 and the nonintervention group numbered 106. Thus, the project met this criteria.

(b) The evaluation tests were administered to all children in Jackson County Head Start in October.

3. Score the tests and record the scores.

Scores were recorded within one week of test administration.

4. Establish a cut-off score. Assign all students scoring below this value to the intervention group.

Cut-off score was established as one year below chronological age level on the PPVT, the CELI or the VMI. All children scoring below this level were assigned to the intervention group.

5. After the project, administer posttest to all children in the intervention group and nonintervention group.

All children in Jackson County Head Start who had been given the pretest and were still enrolled in the program were given the posttests in early May.

6. Score the test and record the scores.

Scores were recorded within one week of posttesting.

7. Convert raw scores to NCE's.

Pre- and posttest scores were converted to NCE.

8. Calculate the no-project expectation using the formula on page of this thesis.

9. Compare the observed posttest mean NCE with the no-intervention expectation mean NCE.

(a) Determine the mean posttest NCE.

(b) Compare the no-intervention expected posttest mean with the actual posttest mean. Tallmadge and Wood (1980a) stated that a difference of six NCE represents educational significance.

10. Test the statistical significance of the difference between the two means using a t-test at a .05 level of confidence.

Results

In this section, each hypothesis is tested using the data results from the project.

Hypothesis I

The mean posttest status will be significantly higher than the adjusted mean pretest status for the intervention group children on the PPVT, the VMI, and the CELI.

The NCE score is a normalized percentile scoring system. The assumption is that without intervention the students would remain at the same NCE status in relation to the rest of the norm group.

Table 12 displays data relating to the comparison of the pretest and posttest status of scores using Model A analysis. The differences between pre- and posttest means were tested for statistical significance using the t-test for correlated means and an alpha level of .05 (see Appendix D for more detail on the analysis).

Table 12

Model A Analysis

	PPVT $\bar{X}$	VMI $\bar{X}$	CELI $\bar{X}$
Pretest status			
(Adjusted mean)	31.35	32.59	27.54
Posttest status	41.16	36.97	35.30
Posttest minus pretest status	9.76	4.38	7.76
Educationally significant			
(over 6 NCE)	Yes	No	Yes
t-ratio	4.16	2.126	2.82
Statistically significant			
(over t .05 = 2.02)	Yes	Yes	Yes

NOTE: Expressed in normal curve equivalent (NCE) units.

The analysis indicates that intervention group's mean gains on the PPVT, the VMI, and the CELI all were statistically significant, and that intervention group scores on the PPVT and the CELI also meet criteria set by Tallmadge and Wood (1980b) for demonstrating educationally significant gains.

Hypothesis II

The actual mean posttest score obtained by the intervention group children on the (a) the PPVT, (b) the VMI, and (c) the CELI will be significantly higher than the posttest PPVT, VMI, and CELI mean score estimated from the regression line equation.

Table 13 displays data related to the comparison of estimated and actual mean scores using Model C Analysis. Because the research hypothesis was directional, tests of statistical significance would be called for only if the differences were in the direction predicted. They were not. Actual mean intervention group scores on the PPVT, the VMI, and the CELI were all lower than the estimated mean scores, indicating no intervention group gains on any of the three measurement instruments. Hypothesis II is rejected.

Table 13

Model C Analysis

	PPVT $\bar{X}$	VMI $\bar{X}$	CELI $\bar{X}$
Estimated posttest score	43.44	38.11	38.79
Actual posttest score	41.11	36.97	35.30
Mean estimated score minus mean actual score	-2.4	-1.1	-3.48
Educationally significant (over 6 NCE)	No	No	No
t-score	Below 0	Below 0	Below 0
Statistically significant (over 2.02)	No	No	No

NOTE: Means expressed in normal curve equivalent (NCE) units.

Hypothesis III

1. Model A (Hypothesis I) and Model C (Hypothesis II) will have similar findings for educational significance for the treatment group

in this project as indicated by scores on the PPVT, the VMI, and the CELI.

2. Model A (Hypothesis I) and Model C (Hypothesis II) will have similar findings for statistical significance for the treatment group in this research project as indicated by scores on the PPVT, the VMI, and the CELI.

Displayed in Table 14 are data relating to Hypothesis III(c).

Table 14

Comparison of Model A and Model C

	Model A	Model C
Mean gains PPVT	+9.76	-2.4
Educationally significant	Yes	No
Statistically significant	Yes	No
Mean gains VMI	+4.38	-1.14
Educationally significant	No	No
Statistically significant	Yes	No
Mean gains CELI	+7.76	-3.48
Educationally significant	Yes	No
Statistically significant	Yes	No

Model A indicates educationally significant gains on the PPVT and the CELI. Model C indicates no gains. Hypothesis III_1, is rejected.

Model A indicates statistically significant gains on the PPVT, the VMI, and the CELI. Model C indicates no gains. Hypothesis III_2 is rejected.

CHAPTER IV

DISCUSSION

The purpose of this research project was twofold: (a) to examine whether an intervention with educationally-handicapped children in a Head Start program made a significant difference on their performance on testing instruments using the Model A and Model C evaluation options developed under Title I; and, (b) to determine whether the Model A and Model C options would give similar results when measuring gains made by this one group of children.

The answer to the second question is clearly negative. Model A analysis indicated substantial gains by the intervention group on all three testing instruments. Model C analysis indicated no gains on any of the three tests. A closer look at the kinds of analytical comparisons which the two different models use may give us an explanation of the different results and assist in determining the answer to the first question, if the handicap intervention used in this project had a significant effect on children's performance.

The results in Model A analysis reflects the change in the intervention group relative to the norm group. There is an assumption here that the intervention group would have no score change without intervention. If this assumption were true, then it would be reasonable to expect the nonintervention children in this project, who received no intervention, to show no status change. The data in Table 15 demonstrate that this is clearly not true.

Table 15

NCE Gains-Model A Option

	Nonintervention Students	Intervention Students	Intervention minus Nonintervention
PPVT $\bar{X}$ gain	6.75	9.76	3.01
VMI $\bar{X}$ gain	3.3	4.38	1.08
CELI $\bar{X}$ gain	5.97	7.76	1.81

NOTE: Gains expressed in normal curve equivalents (NCE), and the means for the intervention subjects are the adjusted means.

The statistics listed in Table 16 indicate that both intervention and nonintervention students in Jackson County Head Start made score gains which increased their percentile and NCE standing in relation to the tests' norm groups. The fact that both groups made substantial gains tends to cast doubts on the assumption in Model A that Title I students' NCE status would remain the same without specialized intervention. If this assumption is not true, the progress made by a subgroup who receive a general educational experience plus a specialized intervention experience would be a result of the summation of these two experiences. If the nonintervention students in Jackson County made Peabody score gains of 6.75, it is probable that the students in the intervention group, who are enrolled in the same general program, would have also made some gains even without intervention. Thus, it is a reasonable assumption that a portion of the 9.76 Peabody score gain by the intervention group is attributable to the intervention itself and a portion attributable to the general education program.

Analyzing the data in Table 15, it appears evident that children in the intervention group made significant progress and that some of this progress was a result of the general program and some a result of the specialized intervention program. Because of the unclear effects of regression, the undetermined but varying learning rates between groups of different capabilities, and the probable interactions between program and intervention effects, it would be difficult to accurately determine at this time how much of the gain of the intervention group was due to the intervention and how much was due to the general program.

Model C analysis does not focus on the general gain of children in the intervention group. Instead, the focus is on estimating the scores intervention students would have made without the intervention by assuming that the relationship between intervention group pretest and posttest scores would be the same as that between the pre- and posttest scores of students in the project population who scored above the cut-off and did not receive intervention.

There appears to be an assumption here that learning rates of low-scoring students have an established and predictable relationship of higher-scoring students so that a pre-post regression equation developed on the latter can be validly applied to the former. This assumption has not been demonstrated or proven. As Campbell and Boruch

(1975) so clearly demonstrated in their article on quasi-experimental models, research designs which treat score changes by subjects as simple variables, and do not deal with the different learning rates of fast and slow-learning students, will usually result in findings of artifically low, and even negative, effects in projects designed, as this one was, to work with slow students.

The Model C analysis in this project showed negative gains for scores on all three testing instruments. If accurate, these results would suggest that intervention students would have made more gains if they had received no intervention at all. While it is possible to conceive of situations where needy students would benefit more from remaining full-time in the general program and not receiving any specialized training, the data displayed in Table 15 and Table 16 suggest that it is unlikely that such negative intervention effects are true in this case. As discussed earlier, the data in Table 15 indicate that both intervention and nonintervention groups in this project made gains and that the intervention group gains were higher than the nonintervention group gains. It is difficult to accept that a mean NCE gain of 9.76 PPVT points by the intervention students indicates a negative intervention effect.

The data in Table 16 look at intervention changes from another direction. Table 16 indicates the magnitude of the gains expressed in normal curve equivalents and percentiles predicted by the regression formula in this project. This formula estimates that Jackson County intervention students would move up twenty-four percentile points on the PPVT even if they received no intervention at all. When you take into account the fact that the intervention groups contain a large number of low-scoring students who have low IQ's and slow-learning rates, this seems a very unlikely score estimation.

Table 16

Mean Score Gains-Model C Analysis

	Estimated Gains by Intervention Group	Actual Gains by Intervention Group	Actual Gains by Noninterven-tion Group
PPVT-NCE	16.9	13.5	6.7
VMI-NCE	7.5	6.4	3.3
CELI-NCE	16.8	13.3	6.0
PPVT- percentile	24%	20%	13.5%
VMI- percentile	11%	8%	8%
CELI- percentile	19%	14%	9%

NOTE: Prescore means of intervention group are not adjusted means.

The large gains by the normal subjects resulted in a pre/post regression line quite different from what would have been obtained had there been no gains. The result was a slope and point of origin quite inappropriate for setting score expectations for the intervention group. The formula seems to be stating, in this case, that low-scoring students should be able to score at least double the gains made by high-scoring students. Unfortunately, this is not an accurate reflection of the real world where many low-scoring students need special intervention just in order to equal gains made by high-scoring students.

Thus, in going back to the question of why results in this project are different using the Model A and Model C options, it appears that Model A combines the positive effects of the general and intervention programs in this research project, while Model C uses the slope and point of origin for the nonintervention Head Start students to set unrealistically high score estimates for the intervention group. Earlier in this paper, it was indicated in the review of the literature that previous research analysis (Hardy, 1979; House, 1979; Tallmadge & Wood, 1980a) involving Model A and Model C options did not find the wide discrepancies found in this project. This may be because gains in the intervention and nonintervention groups in other research projects were not so great, and thus, did not highlight model differences as they appear to have done here.

Would the results of this research project support the use of either of these two evaluation options by program evaluators? If our analysis is correct, then a case can be made for a limited use of the Model A option. The Model A option determines the general gain of students in comparison to the norm group but does not, by itself, give the intervention effect. In school districts such as the one in this project, where students tend to make gains in comparison to the norm group, Model A analysis will tend to overestimate the effects of the intervention because it will lump together the gains due to intervention with the gains due to the general education program. In school districts where students traditionally lose ground compared to the norm group over the years, the effects of a part-time intervention program will tend to be underestimated, since Model A analysis assumes that students without intervention would not lose ground compared to their norm group. Programs can use Model A analysis to measure the overall gain of students receiving intervention compared to the norm group; they would not be able to use Model A analysis, as it is now designed, to measure the effect of the intervention itself.

Given the results in this project, a similar case cannot be made for the present use of the Model C option. Before this option can be used, the accuracy of the regression line formula must be tested in a variety of populations with varying levels of learning gains for both the intervention (below cut-off scores) and nonintervention (above cut-off scores) student groups. It would be simplest to test the formula with a no-intervention design in which neither students above nor below the cut-off received any planned intervention. If the assumption underlying use of the regression line formula were correct, then the estimated scores for the below cut-off group would equal their actual scores since they did not, in this projected case, receive any intervention.

Overall, limitations in both Model A and Model C analysis point out the general need to obtain separate norming scores for disadvantaged and handicapped populations and/or to devise statistical methods of accurately predicting the learning rate of slow-learning students. If neither of these objectives can be met, it will be difficult to justify evaluation method alternatives to control group research for compensatory programs. At the present time, program evaluators must either find ways to create control groups, forego research evaluation, or carefully limit evaluation questions to those which present methods can reasonably answer.

REFERENCES

Aitken, S., Hubbell, R., & Jones G. (1982). A review of Head Start research since 1970. Washington, DC: Office of Human Development.

Berry, K. E., & Buktencia, N. A. (1967). Developmental test of visual-motor integration. Chicago: Follett Publishing Company.

Berieter, C., & Englemann, S. (1966). Teaching disadvantaged children in preschool. Englewood Cliffs, NM: Prentice-Hall.

Bloom, S. B. (1964). Stability and change in human characteristics. New York: Wiley.

Bronfenbrenner, U. (1975). Is early intervention effective? In J. Hellmuth (Ed.), Exceptional Infants, Vol. 3. New York: Bruner/Mazel.

Bush, D. & White, K. (1983). Early Intervention Research Institute--The efficacy of early intervention: What can be learned from previous reviews of the literature. A paper presented at the Rocky Mountain Psychological Association, Snowbird, UT.

Bush, D., White, K., Casto, G., & Shearer, D. (1982). Early Intervention Research Institute--The effectiveness of early intervention: A summary of previous reviewers' conclusions. A paper presented at the Rocky Mountain Psychological Association, Snowbird, UT

Campbell, D. P. & Boruch, R. F. (1975). Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In C. Bennett & A. Lumsdaine (Eds.), _Evaluation and experiment: Some critical issues in assessing social programs._ New York: Academic Press.

Carrow, E. (1974). _The Carrow elicited language inventory._ Austin, TX: Learning Concepts.

Casto, G. (Ed.) (1979). _An early intervention program for the handicapped: Training manual._ New York: Walker Publishing Co.

Chissom, B. (1972). Developmental test of visual-motor integration: A review in _Seventh Mental Measurements Yearbook._ Highland Park, NJ: The Gryphon Press.

Cicirelli, V., et al. (1969, June). _The impact of Head Start: An evaluation of Head Start on children's cognitive and affective development_. Report presented to the Office of Economic Opportunity, pursuant to Contract B89-4536. (Report No. PB 184 328.) Westinghouse Learning Corporation for Federal Scientific and Technical Information, U.S. Institute for Applied Technology.

Cooke, T., Ruskus, J., Apolloni, T., & Peck, S. (1981). Handicapped preschool children in the mainstream: Background, outcomes, and clinical suggestions. _Topics in Early Childhood Special Education, 1_, 73-83.

Cornelius, S. (1974). A comparison of the elicited language inventory with the developmental syntax scoring procedure in assessing language disorders in children. In E. Carrow, _The Carrow elicited language inventory_. Austin, TX: Learning Concepts.

DeWeerd, J., & Cole, A. (1976). Handicapped children's early education program. Exceptional Child, 3, 155-157.

Dunn, L. (1965). The Peabody picture vocabulary test. Circle Pines, MN: American Guidance Services.

Eldred, N. & White, K. (1982). A critical review of the Title I evaluation and reporting systems. Unpublished manuscript, Utah State University, Logan, UT.

Grill, J. (1978). Developmental indicators for the assessment of learning test: A review in Eighth mental measurement yearbook. Highland Park, NJ: The Gryphon Press.

Hardy, R. (1979). A comparison of Model A and Model C: Results of first year implementations in Flordia ETS. Evanston, IL.

House, G. (1979). A comparison of Title I achievement results obtained under UOE Models A1, C1, and a mixed model. Washington, DC: ERDS.

Hubbell, R. (1983). A review of Head Start research since 1970. Washington, DC: Office of Human Development Services, Administration for Children, Youth, and Families, Head Start Bureau.

Hunt, J. M. (1961). Intelligence and experience. New York: Ronald Press.

Jensen, A. (1969). How much can we boost IQ and scholastic achievement? Harvard Education Review. 39, 1-123.

Karnes, M., Hodgine, A., & Kirk, S. (1969). Research and development program on preschool disadvantaged children, Vol. 1. Washington, DC: U.S. Office of Education.

Kirk, S. (1958). Early education of the mentally retarded: An experimental study. Urbana, IL: University of Illinois Press.

Klaus, R., & Gray, S. (1968). The early training project for disadvantaged children: A report after five years. Monographs of the Society for Research in Child Development, 33(4).

Lazar, I., Beller, K., Gordon, I., Deutsch, M., Deutsch, C., Gray, S., Karners, M., Woolman, M., Zigler, E., Weikart, D., Palmer, F., Miller, L., Levenstein, P., & Peterson, J. (1977). The persistence of preschool effects. A long-term follow-up of 14 infant and preschool experiments. Washington, DC: U.S. Office of Human Development.

Linn, R. (1980). Evaluation of Title I via the RMC models: A critical review. Urbana, IL: University of Illinois Press.

Mandeville, G. (1978, March). An evaluation of Title I Model C1: The special regression model. EDRS, A paper presented at the Annual Meeting of the American Educational Research Association, Ontario, Canada.

Mann, A., Harrell, A., & Hurt, M. (1977). A review of Head Start research since 1969 and an annotated bibliography. Washington, DC: Department of Health, Education, and Welfare.

Mardell, C., & Goldberg, D. (1972). Developmental indicators for the assessment of learning. NJ: Childcraft.

May, M. (1979). Evaluating handicapped children's early education programs. Westar conference.

Office of Child Development (1975, September). Transmittal notice #75.11--announcement of diagnostic criteria for reporting handicapped children in Head Start. (DHEW, Publication No. A30-333-4.) Washingtion, DC: U.S. Goverment Printing Office.

Ruopp, R., Travers, J., Glantz, F., & Corden, C. (1979). Children at the center: Final report of the national day care study. Cambridge, MA: Abot Associates.

Stanley, J. (Ed.) (1972). Preschool programs for the disadvantaged five experimental approaches to early childhood education. Baltimore, MD: John Hopkins University Press.

Status of handicapped children in Head Start programs (1980). Washington, DC: Office of Human Development.

Tallmadge, G., & Wood, C. (1978). Title I outcome evaluation. Portland, OR: Northwest Regional Education Laboratory.

Tallmadge, G., & Wood, C. (1980a). Comparability of gains from the three models in the Title I evaluation system. Mountain View, CA: RMC Research Corp.

Tallmadge, G., & Wood, C. (1980b). User's guide, ESEA Title 1 evaluation and report system. Mountain View, CA: RMC Research Corp.

Turnbull, A., Winton, P., Blacher, J., & Salkine, N. (1982). Mainstreaming in the kindergarten classroom: Perspectives of parents of handicapped and nonhandicapped children. Journal of the Division of Early Childhood, 6, 14-18.

Weikert, D. (1970). A comparative study of three preschool curricula. In J. Front (Ed.), Disadvantaged child. New York: Houghton Mifflin.

Weikert, D. Bond, J. T., & McNeil, J. T. (1978). The Ypsilanti Perry Preschool Project: Preschool years and longitudinal results through fourth grade. Monographs of the High/Scope Educational Research Foundation #3. Michigan: Ypsilanti.

Weikert, D., Deloria, D., & Lawsor, S. (1974). Results of a pre-school intervention project. In S. Ryan (Ed.) A report on longitudinal evaluations of reports and pre-school programs. Washington, DC: U.S. Department of Health, Education, and Welfare.

Wood, C. (1980). The adequacy of the equipercentile assumption in the norm-referenced evaluation model. Mountain View, CA: RMC Research Corp.

Wynne, S., Ulfelder, L., & Dakoff, G. (1975). Mainstreaming and early childhood education for handicapped children: Review and implications of research. Washington, DC: U.S. Office of Education.

Zigler, E., & Valentine, J. (1979). Project Head Start, a legacy of the war on poverty. New York: The Free Press.

APPENDICES

Appendix A. Diagnostic Criteria

APPENDIX A

DIAGNOSTIC CRITERIA

The following diagnostic criteria is taken from Office of Child Development (1975).

Hearing Impairment

A child shall be reported as hearing impaired when any one of the following exists: (a) the child has slightly to severely defective hearing, as determined by his/her ability to use residual hearing in daily life, sometimes with the use of a hearing aid; (b) hearing loss from 26-92 decibels (American National Standard Institute, 1969) in the better ear.

Physical Handicap (Orthopedic Handicap)

A child shall be reported as crippled with an orthopedic handicap who has a condition which prohibits or impedes normal development of gross or fine motor abilities. Such functioning is imparied as a result of conditions associated with congenital anomalies, accidents, or diseases; these conditons include, for example, spina bifida, loss of or deformed limbs, burns which cause contractures, and cerebral palsy.

Speech Impairment (Communication Disorder)

A child shall be reported as speech impaired with such identifiable disorders as receptive and/or expressive language impairement, stuttering, chronic voice disorders, and serious articulation problems affecting social, emotional, and/or educational achievement; and speech and language disorders accompanying conditions of hearing loss, cleft palate, cerebral palsy, mental retardation, emotional disturbance, multiple handicapping conditions, and other sensory and health impairments. This category excludes conditions of a transitional nature consequent to the early developmental processes of the child. When speech and language disorders accompany conditions of hearing loss, cerebral palsy, mental retardation, emotional disturbance, multiple handicappping conditions, and other sensory and health impairments, the child should be reported under the most disabling problem.

Health Impairment

These impairments refer to illness of a chronic nature or with prolonged convalescence including, but not limited to, epilepsy, severe asthma, severe cardiac conditions, severe allergies, blood disorders (e.g., sickle cell disease, hemophilia, leukemia), diabetes, neurological disorders, or autism.

Mental Retardation

A child shall be considered mentally retarded who, during the early developmental period, exhibits significant sub-average intellectual functioning accompanied by impairment in adaptive behavior. In any determination of intellectual functioning using standardized tests that lack adequate norms for all racial/ethnic groups at the preschool age, adequate consideration should be given to cultural influences as well as age and developmental level (i.e., finding of a low I.Q. is never by itself sufficient to make the diagnosis of mental retardation).

Serious Emotional Disturbance

A child shall be considered seriously emotional disturbed who is identified by professionally qualified personnel (psychologist or psychiatrist) as requiring special services. This definition would include, but not limited to, the following conditions: dangerously aggressive towards others, self-destructive, severely withdrawn, and non-communicative, hyperactive to the extent that it affects adaptive behavior, severely anxious, depressed or phobic, or psychotic.

Specific Learning Disabilities

These disabilities refer to a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations. Such disorders include such conditions as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. Not included are learning problems which are primarily the result of visual, hearing, or motor handicaps, of mental retardation, of emotional disturbance, or of environmental disadvantage. For preschool children, precursor functions to understanding and using language spoken or written, and computational or reasoning abilities are included.

NOTE:

The following are considered appropriate diagnostic professionals for each of the handicapping conditions:

Blindness and Visual Impairment: ophthalmologists, optometrists.

Deafness and Hearing Impairment: otolaryngologists, audiologists.

Physical Handicap (Orthopedic): Orthopedists, physiatrists

Speech Impairment: otolaryngologists, speech pathologists.

Health Impairment: pediatricians, general practitioners, psychiatrists, psychologists.

Mental Retardation: pediatricians, psychiatrists, psychologists.

Serious Emotional Disturbance: psychiatrists, psychologists.

Specific Learning Disabilities: psychiatrists, psychologists, educators with at least a Masters Degree in Special Education and with specific training in diagnosis of learning disabilities.

Appendix B. Family Income Guidelines

APPENDIX B

Administration for Children, Youth,

and Families Notice

1981 Family Income Guidelines

Poverty Income Guidelines for
All States Except Alaska and Hawaii

Size of Family Unit	Nonfarm Family	Farm Family
1	$ 4,310	$3,680
2	5,690	4,850
3	7,070	6,020
4	8,450	7,190
5	9,830	8,360
6	11,210	9,530

For family units with more than 6 members, add $1,380 for each additional member in a nonfarm family and $1,170 for each additional member in a farm family.

Poverty Guidelines for Alaska

Size of Family Unit	Nonfarm Famiy	Farm Family
1	$ 5,410	$4,610
2	7.130	6,070
3	8,850	7,530
4	10,570	8,990
5	12,290	10,450
6	14,010	11,910

For family units with more than 6 members, add $1,720 for each additional member in a nonfarm famiilly and $1,460 for each additional member in a farm family.

Poverty Guidelines for Hawaii

Size of Family Unit	Nonfarm Family	Farm Family
1	$ 4,980	$ 4,250
2	6,560	5,590
3	8,140	6,930
4	9,720	8,270
5	11,300	9,610
6	12,880	10,950

For family units with more than 6 members, add $1,580 for each additional member in a nonfarm family and $1,340 for each additional member in a farm family.

Appendix C. Adjustment of Pretest
Mean-PPVT

APPENDIX C

Adjustment of Pretest Mean-PPVT

$\bar{X}_p{}' = \bar{X}_p + [(1-r_{xx}) \quad (\bar{X}_t - \bar{X}_p)]$

$\bar{X}_p{}'$ = Adjusted mean score of intervention group on pretest.

$\bar{X}_p$ = 27.558 NCE=Actual mean score of intervention group on pretest.

R_{xx} = .7726 = Test-retest reliability for total group.

$\bar{X}_t$ = 44.275 NCE=Mean score for the entire group on pretest.

$\bar{X}_p{}'$ = 27.558 + (.2274) (16.717).

$\bar{X}_p{}'$ = 31.35 NCE

Adjustment of Pretest Mean-VMI

$\bar{X}_p{}' = \bar{X}_p = [(1-r_{xx}) \quad (\bar{X}=_t - \bar{X}_p)]$

$\bar{X}_p{}'$ = Adjusted mean score of intervention group on pretest.

$\bar{X}_p$ = 30.605 NCE=Actual mean score of intervention group on pretest.

r_{xx} = .757 = Test-retest reliability for total group.

$\bar{X}_t$ = 38.81 NCE=Mean score for the entire group on pretest.

$\bar{X}_p{}'$ = 30.605 + (.243) (8.205).

$\bar{X}_p{}'$ = 32.59 NCE

Adjustment of Pretest Mean-CELI

$\bar{X}_p{}' = \bar{X}_p + (1-r_{xx}) \ (\bar{X}_t - \bar{X}_p)$

$\bar{X}_p{}'$ = Adjusted mean score of intervention group on pretest.

$\bar{X}_p$ = 21.953 NCE=Actual mean score of intervention group on pretest.

r_{xx} = .7760 = Test-retest reliability for total group.

X_t = 46.893 NCE = Mean score for entire group on pretest.

X_p' = 21.953 NCE + (.2240) (24.94) NCE.

X_p' = 27.539 NCE

Appendix D. t-Calculations--
Model A

APPENDIX D

Table 17

t-Calculations--Model A

	PPVT	VMI	CELI
s^1 =	21.11	19.41	28.1
s^2 =	23.9	19.42	25.2
r_{12} =	.7726	.756	.776
$s_D^2 = s_1^2 + s_2^2 - 2r_{12}\, s_1\, s_2$	237.24	183.18	325.6
$s_{\bar{D}}^2 = \frac{s_{\bar{D}}^2}{N}$	5.51	4.26	7.56
$s_{E\bar{D}} = \sqrt{s_D^2}$	4.16	2.06	2.75
$\bar{D} = \bar{X}_1 - \bar{X}_2$	9.76	4.38	7.76
$t = \frac{\bar{D}}{s_{\bar{D}}}$	4.16	2.126	2.82

NOTE: Formula from Ferguson (1981), p. 180-181.

Appendix E. Formula for Regression

APPENDIX E

Table 18

Formula for Regression $\bar{Y}_p = \bar{Y}_c - (\bar{Y}_c - \bar{X}_p)\, r_{xy_c} \frac{syc}{sxc}$

	PPVT	VMI	CELI
$\bar{Y}_c$	57.8	46.8	62.98
$\bar{X}_c$	51.05	42.1	57.9
$\bar{X}_p$	27.55	30.6	21.9
r_{xyc}	.7003	.7567	.7510
s_{yc}	16.382	17.298	24.159
s_{xc}	18.757	17.308	26.294
$\bar{Y}_p$	43.44	38.11	38.79