AN EMPIRICAL EXAMINATION OF ANALYSIS OF COVARIANCE WITH AND WITHOUT PORTER'S ADJUSTMENT FOR A FALLIBLE COVARIATE

By

JAMES EDWIN McLEAN

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1974

DEDICATION

I dedicate this study to the late Dr. Charles M. Bridges, Jr., my teacher, advisor and friend.

ACKNOWLEDGMENTS

I wish to express my grateful appreciation to my committee members for their guidance in this study. Dr. William B. Ware, my chairman, suggested the topic and was a major influence in its development. Dr. James T. McClave was a constant source of assistance, both theoretically and editorially. Drs. Vynce Hines, William Mendenhall, and P. V. Rao all provided direction along with many helpful suggestions. This study could not have been completed without the spirit of cooperation which I encountered between members of the two departments involved.

The late Dr. Charles M. Bridges, Jr. first encouraged me to enroll in the program and served as my chairman until his death. His counseling and guidance were a major reason for my successful completion of graduate study.

The searching questions of my fellow students led to several worthwhile modifications and I wish to express my appreciation to them.

I also wish to thank my wife, Sharon, who for the last five years, has held two jobs (wife and medical technologist), so that I might further my education. She has also been most understanding about the many hours spent away from home.

TABLE OF CONTENTS

LIST OF TABLES

Abstract of Dissertation Presented to the
Graduate Council of the University of Florida in Partial
Fulfillment of the Requirements for the Degree of Doctor of Philosophy

AN EMPIRICAL EXAMINATION OF
ANALYSIS OF COVARIANCE WITH AND WITHOUT
PORTER'S ADJUSTMENT FOR A FALLIBLE COVARIATE

by

James Edwin McLean

August, 1974

Chairman: William B. Ware

Major Department: Foundations of Education

The purpose of this study was to determine if analysis of covariance or analysis of covariance with Porter's true covariate substitution adjustment or neither is appropriate for analyzing pretest-posttest educational experiments with less than perfectly reliable measures.

Monte Carlo techniques were employed to generate two thousand samples for each of forty-eight sets of conditions. These conditions included six combinations of reliability, two levels of sample size, two levels of gain, and the equality or inequality of pretest means. Each sample was analyzed by both of the tested methods. The proportion of rejections for each method of analysis and the number of times the F statistic for standard analysis of covariance exceeded that for analysis of covariance with Porter's adjustment were recorded.

The hypothesis, "The sampling distributions of the test statistics from analysis of covariance with and without Porter's adjustments are the same," was rejected at the .01 level of significance using the sign test for each of the forty-eight sets of conditions.

To gain further insight into how the two methods of analysis differed, four factorial experiments were conducted using either the computer generated alphas or powers as the criterion variables. The first factorial experiment examined analysis of covariance when the mean gain was zero in both groups, thus, the computer generated alphas were used as criterion variables. The second experiment examined analysis of covariance with Porter's adjustment when the mean gain was zero in each group. Again, computer generated alphas were used as criterion variables. The third and fourth factorial experiments examined standard analysis of covariance and analysis of covariance with Porter's adjustment when the mean gain in only one group was positive. The four factorial experiments included three factors. These were reliability at six levels, sample size at two levels, and the equality of pretest means at two levels (pretest means equal and pretest means not equal). In each experiment, it was found that unequal pretest means combined with low reliability were significant sources of variation.

Additional study showed that unequal pretest (covariate) means combined with low reliability produced misleading results. More specifically, when either method of analysis produced a significant $\underline{F}$ statistic and there was a mean gain of zero in both groups but the pretest mean of one group was less than that of the other, the adjusted posttest means indicated that the group with the higher pretest mean had the larger gain. Likewise, when there was a positive mean gain in the group with the lower pretest mean, the adjusted posttest means indicated the group with a mean gain of zero had the larger gain.

The results of this study combined with the results of others concerned with analysis of covariance where the covariates are fallible point to the inadequacy of the technique as it now exists. These results should be kept in mind if the technique is to be used.

CHAPTER I
INTRODUCTION

Regardless of the theory or hypothesis being investigated, educational researchers must, at some point, relate it to learning. One common element of most definitions of learning is that a change must take place in the learner (DeCecco, 1968, p. 243; Hilgard, 1956, p. 3; Hill, 1971, p. 1-2). Thus most investigations in education are concerned with identifying, measuring, and comparing these changes.

In the past decade, analysis of covariance has become a standard procedure for comparing groups with different levels of ability. The basic derivation of this procedure requires the assumption that the covariable does not contain errors of measurement (Cochran, 1957). In recent years, several researchers have recognized that this assumption is violated when the covariate is a mental test score (Lord, 1960; Porter, 1967; Campbell and Erlebacher, 1970). Lord (1960) and Porter (1967) have proposed adjustments to the analysis of covariance procedure to use when the covariable is fallible, that is, when measurement error is present in the covariable. The focus of this study is to determine if analysis of covariance and analysis of covariance with Porter's adjustment are different for a fallible covariate and if either is appropriate under varying conditions of reliability, sample size, gain, and pretest mean.

The Problem of Comparing Groups of Differing Abilities

The problem of comparing groups based on change manifests itself prominently when the variables are mental measurements. Most physical science measurements are made with the aid of some type of physical instrument. If an engineer is interested in the weight gain of a metal before and after a galvanizing process, instruments are available to measure the weight of the metal to within at least one microgram. On the other hand, if an educator is interested in the change in a student's I.Q. before and after taking part in an experimental program, an error of 5 points or more would not be uncommon. Based on the Wechsler Intelligence Scale for Children with a standard error of measurement of 5.0 (Cronbach, 1970, p. 222), an error of 5.0 I.Q. points or more would occur in approximately 32 percent of the measurements, assuming a normal distribution. If the actual change in I.Q. were near zero for an individual, it is very likely that the error of measurement would exceed the change itself.

The problem is compounded further as the groups being compared often are not of the same ability level. Many of the recent programs featuring innovative teaching practices for compensatory education are available only for the most needy and the comparison group is then sampled from the general population of students (Campbell and Erlebacher, 1970). Programs such as Head Start and Follow Through have probably been the victim of tragically misleading analyses, such as in the Westinghouse/Ohio University study (Campbell and Erlebacher, 1970). Personal contact with the evaluation of Follow Through projects has emphasized the magnitude of the problem. Warnings about the inappropriate uses of existing modes of analysis have come from several sources (Lord, 1960, 1967, 1969; Campbell and Erlebacher, 1970).

Remedies have been offered (Lord, 1960; Porter, 1967), but at this time there is no conclusive evidence to indicate these remedies are appropriate. Studies concerning the robustness of the analyses to violated assumptions have also been in conflict (Peckham, 1970).

Methods of Comparing Groups of Differing Abilities

Two general approaches are available to compare groups with differing abilities. The first is to compare the average change in one group with the average change in the other group by means of a *t* test or analysis of variance. The second is to use analysis of covariance, where the pretest score is the covariate.

There are at least three methods of computing change scores to be considered if the first approach is used. One method is to obtain raw difference scores by simply subtracting each pretest score from its corresponding posttest score. Another method is Lord's true gain (Lord, 1956) which was further developed by McNemar (1958). Basically, this method requires the use of a regression equation to estimate the "true" gain or change from pretest to posttest. A third method of measuring gain, the method of residual gain, was proposed by Manning and DuBois (1962). This method involves regressing the posttest scores on the pretest scores and obtaining a predicted posttest score for each subject. The indicator of change is the observed posttest score minus the predicted posttest score. This method is mathematically equivalent to the standard analysis of covariance (Neel, 1970, p. 30-31), the second approach.

The second general approach is the standard analysis of covariance procedure found in many statistical texts (e.g., Kirk, 1968; Winer, 1971). One of the reported functions of a covariate is to adjust the treatment means of the dependent variable for differences in the values of corresponding independent variables (Hicks, 1965). Logically, it seems that by using the pretest scores of groups differing in ability as a covariate, the treatment means can be statistically adjusted for the differences in pretest means. A deficiency of this method is the fallibility of the covariate. A fallible variable is one which is not measured without error or with perfect reliability (Lord, 1960). Lord (1960), recognizing this deficiency in the use of analysis of covariance, derived an adjustment to compensate for the fact that the pretest scores were fallible. Porter (1967) modified Lord's procedure for more than two groups and empirically investigated the sampling distribution of Lord's statistic. Porter's solution has been suggested as an alternative to the standard analysis of covariance when the covariate is fallible.

Limitations Used in this Study

This study was designed to examine the characteristics of the two covariance methods within the framework of simulated situations which may occur in the analysis of a compensatory education project. Realistically, several limitations were imposed. Even with these limitations, the two methods of analysis were investigated under forty-eight combinations of differing reliability, sample size, mean gain, and pretest means.

The study was limited to two groups, henceforth called the gain and no-gain groups. Investigating more than two groups would not only increase the number of possibilities appreciably but would not be in line with the objective stated previously, that is, comparing a compensatory education group with a "comparison" group. The levels of the reliabilities being investigated are also limited to the range .50 to .90. Seldom do mental measurement instruments have reliabilities above .90 and instruments with reliabilities below .50 are generally considered inadequate.

The levels of sample size were 10 and 100 subjects per group. These quantities are representative of small and large sample sizes encountered in educational experiments.

Furthermore, the true score and true gain variances will be fixed a priori with the true gain variance always four percent of the true score variance. These figures are empirically based on Follow Through achievement data in keeping with the aforementioned objective.

Procedures

The study compared the two methods of analysis under forty-eight sets of conditions. The two methods were analysis of covariance and analysis of covariance with Porter's adjustment for a fallible covariate.

Three levels of reliability and two levels of sample size were used. The three levels of reliability will include situations where the reliability of the pretests are assumed equal and situations where they differ to more closely simulate compensatory education project evaluation. Each of these combinations of conditions was repeated for four separate cases. Case I is where there is no-gain in either group and both groups have equal pretest means. Case II is where there is gain in only one group (the gain group) and both groups have equal pretest means. Cases III and IV are with and without gain where the pretest means are different.

For each of the forty-eight sets of conditions, two thousand samples were computer generated and analyzed using both methods of analysis and a .05 level of significance. Two thousand samples provide empirical estimates of the fraction of type I and type II errors to within .01 of their true values with ninety-five percent confidence. The samples were generated from an assumed normal population. The generation technique is one proposed by Box and Muller (1958) and modified by Marsaglia and Bray (1964). This procedure generates random normal deviates with a mean and variance of 0 and 1 respectively. These normal variables are then transformed to attain the specified means and variances. The IBM 370/165 computer of the North Florida Regional Data Center was used for the generation and analyses.

Using these analyses, two research questions were examined:

1. Is there any difference between the sampling distributions of the test statistics of standard analysis of covariance and analysis of covariance with Porter's adjustment?
2. What factors affect each type of analysis?

The first research question can be restated in terms of a null hypothesis for each set of conditions:

> The sampling distributions of the test statistics from analysis of covariance with and without Porter's adjustment are the same for each set of conditions.

This hypothesis was tested for each reliability, sample size, gain, and pretest mean combination.

The second research question was examined in four factorial experiments. Each factorial experiment included the factors of reliability at six levels, sample size at two levels, and equality of pretest means at two levels. The first experiment consisted of examining standard analysis of covariance with computer generated alpha values as the criterion variables. The second examined analysis of covariance with Porter's correction also using computer generated alpha values as the criterion variables. The third and fourth factorial experiments examined analysis of covariance with and without Porter's adjustment respectively, using computer generated powers as criterion variables.

Relevance of the Study

This study is designed to compare two recommended methods of analyzing educational experiments under known simulated conditions which are presumably realistic. The results should indicate whether either of the methods is appropriate for its recommended use, thus giving educational researchers some direction when faced with the choice. If both methods were shown to be appropriate, the study could determine which one is superior. The study should also indicate if either method of analysis is appropriate for a set of restricted conditions, e.g. for only certain levels of reliability.

Organization of the Dissertation

A statement of the problem, a description of possible solutions, and an overview of the procedure of this study has been included in Chapter I. A comprehensive review of the related literature is provided in Chapter II. This review includes a historical overview of the problem and a detailed description of the methods being compared. A description of the procedures followed for this study is contained in Chapter III. The data and analyses are presented in Chapter IV and a discussion of the results is provided in Chapter V. A summary of the study is provided in Chapter VI.

CHAPTER II
REVIEW OF THE RELATED LITERATURE

Comparing groups of differing abilities has been a problem for some time. The literature has included discussions of possible solutions for at least the past three decades. Many of these discussions have been in conflict with one another, and there is still no general agreement. The evaluation of federal compensatory education projects in the last decade has intensified the debate and the need for a theoretically sound and practically useful solution. Some of the solutions proposed over the years are matching subjects, gain scores, and analysis of covariance. A discussion of these solutions and why they may not be considered satisfactory is provided in this chapter. The major properties of analysis of covariance are also considered. Because Porter's adjustment for a fallible covariable in analysis of covariance is not readily available in the literature, its derivation is provided in this chapter.

Historical Review of the Problem

A paper by Thorndike (1942) examined in detail the fallacies of comparing groups of differing abilities by matching subjects. The crux of his argument was that the regression effects were systematically different "whenever matched groups are drawn from populations which differ with regard to the characteristics being studied," (p. 85).

During the late 1950's and early 1960's the literature was inundated with papers on how to or how not to measure gain or change. One of the leaders during this period was Lord (1956, 1958, 1959, 1963). His proposal for estimating the "true" gain was originally set forth in 1956 with further developments in 1958 and 1959. McNemar (1958) extended his results for the case of unequal variances among the groups. Garside (1956) also proposed a method for estimating gain scores and Manning and DuBois (1962) presented their derivation of residual gain scores.

The debate over what techniques, if any, should be used for measuring gain or change was at its peak in the early 1970's. Cronbach and Furby (1970) concluded that one should generally rephrase his questions about gain in other ways. Marks and Martin (1973) underscored the Cronbach and Furby (1970) conclusion. O'Connor (1972) reviewed developments of gain scores in terms of classical test theory. Neel (1970) employed Monte Carlo techniques to compare four identified methods for measuring gain. The compared methods were raw difference, Lord's true gain, residual gain, and analysis of covariance. Under equivalent conditions, he found that Lord's true gain tended to produce a greater significance level than the user would intend, that is, a higher fraction of type I errors than alpha.

The analysis of covariance method has been widely recommended by a number of authorities in the field (Thorndike, 1942; Campbell and Stanley, 1963; O'Connor, 1972). Campbell and Erlebacher (1970) point out that the Westinghouse/Ohio University Study was evaluated, possibly incorrectly, using the analysis of covariance. Lord (1960, 1967, 1969) has sounded a warning about its use that has been echoed by others (Werts and Linn, 1970; Campbell and Erlebacher, 1970; Winer, 1971). The warning stressed that the analysis of covariance requires the assumption that the covariate is measured without error. Lord (1960) has proposed an adjustment which was later generalized by Porter (1967). The adjustment is based on the substitution of the true score estimate for the observed value of the covariate.

Analysis of Covariance

The analysis of covariance procedure was originally introduced by Sir Ronald A. Fisher (1932, 1935). According to Fisher (1946, p. 281), the analysis of covariance "combines the advantages and reconciles the requirements of two widely applicable procedures known as regression and analysis of variance." The procedure is well documented in contemporary texts (Snedecor and Cochran, 1967, p. 419-446; Kirk, 1968, p. 455-489; Winer, 1971, p. 752-812). Analysis of covariance is a popular technique in both the physical and social sciences.

Among the principal uses of analysis of covariance pointed out by Cochran (1957), p. 264) is "to remove the effects of disturbing variables in observational studies." It was thought that by using the pretest score as the covariate and comparing two groups with analysis of covariance, the effects of different pretest scores could be eliminated. Lord (1960) pointed out that this was not necessarily the case when the covariate was fallible. The assumptions necessary for the analysis of covariance are the same as those for analysis of variance with the addition of the following:

1. The covariates are measured without error.
2. The regression coefficient is constant across all treatment groups (Peckham, 1970).

Violation of the first assumption induces a bias in the analysis of covariance because of "the presence of 'error' and 'uniqueness' in the covariate, i.e. variance not shared by the dependent variable. If the proportion of such variance can be correctly estimated, it can be corrected for," (Campbell and Erlebacher, 1970, p. 199). This position was upheld by Glass et al. (1972). The basic algebraic derivation of this correction was presented by Lord (1960) and expanded by Porter (1967).

Analysis of Covariance with Porter's Adjustment for a Fallible Covariate

Lord's (1960) derivation of the analysis of covariance adjustment was limited to two treatment groups and is slightly more difficult than is Porter's (1967) procedure. Thus, Porter's procedure will be derived here and used in the analysis. Porter's adjustment is based on the substitution of the estimated true value for the covariate.

Let X denote a fallible variable (e.g. a pretest score), r an estimate of the reliability of X, and T_i the true value of the variable X_i. Let $\hat{T}_i$ denote the estimated true score of X_i. The definition of T_i for each individual is

(1) $$\hat{T}_i = \overline{X} + r(X_i - \overline{X}),$$

or $$\hat{T}_i = rX_i + \overline{X}(1-r).$$

Porter (1967) derived the mean and variance of T_i as follows. By definition, the mean of all $\hat{T}_i$'s is:

(2) $$\bar{\hat{T}} = \frac{\Sigma \hat{T}_i}{N},$$

$$= \frac{\Sigma[rX_i + \overline{X}(1-r)]}{N}$$

by substitution. Thus

$$\bar{\hat{T}} = r\overline{X} + \frac{N\overline{X} - Nr\overline{X}}{N},$$

$$= \overline{X},$$

where N denotes the sample size and it is understood that i is summed over the values 1, 2, ---, N.

Also, by definition, the variance of $\hat{T}$ is

(3) $$S^2_{\hat{T}} = \frac{\Sigma(\hat{T}_i-\overline{X})^2}{N-1},$$

$$= r^2S^2_X,$$

and

(4) $$S_{\hat{T}Y} = \frac{\Sigma(\hat{T}_i-\overline{X})(Y_i-\overline{Y})}{N-1},$$

$$= rS_{xy},$$

where Y is the dependent variable.

Based on these results, Porter (1967) derived an analysis of covariance procedure replacing the fallible covariate, X_i, with the estimated true score, T_i. The analysis of covariance requires the computation of analysis of variance sums of squares for the dependent variable, the covariable, and on the cross-products of the dependent variable and the covariable. Porter (1967) showed that the use of estimated true scores for the covariable did not affect the analysis of variance of the dependent variable, Y. The changes found by Porter (1967) in the other two cases are as follows:

For the analysis of variance of the estimated true scores, $\hat{T}$,

(5) $$SS_{W_{\hat{T}}} = \Sigma\Sigma(\hat{T}_{ij}-\bar{\hat{T}}_{.j})^2$$

$$= r^2\Sigma\Sigma(X_{ij}-\overline{X}_{.j})^2,$$

where $SS_{W_{\hat{T}}}$ denotes the within groups sum of squares,

(6) $$SS_{B_{\hat{T}}} = n\Sigma(\bar{\hat{T}}._{j}-\bar{\hat{T}}..)^2$$

$$= n\Sigma(\bar{X}._{j}-\bar{X}..)^2$$

where $SS_{B_{\hat{T}}}$ denotes the between groups sum of squares,

(7) $$SS_{T_{\hat{T}}} = \Sigma\Sigma(\hat{T}_{ij}-\bar{\hat{T}}..)^2$$

$$= r^2\Sigma\Sigma X_{ij}^2+(1-r)\Sigma\left[\frac{(\ X_{ij})^2}{n}\right]-\frac{(\Sigma\Sigma X_{ij})^2}{N},$$

where $SS_{T_{\hat{T}}}$ denotes the total sum of squares, n denotes the number of $(\hat{T},Y)$ pairs per treatment group, and N denotes the total number of $(\hat{T},Y)$ pairs. In a similar manner, the cross-products sums of squares are:

(8) $$SS_{W_{\hat{T}Y}} = r\Sigma\Sigma(X_{ij}-\bar{X}._{j})\ (Y_{ij}-\bar{Y}._{j}),$$

(9) $$SS_{B_{\hat{T}Y}} = n\Sigma(\bar{X}._{j}-\bar{X}..)(\bar{Y}._{j}-\bar{Y}..),$$

(10) $$SS_{T_{\hat{T}Y}} = r\Sigma\Sigma X_{ij}Y_{ij}+(1-r)\Sigma\left[\frac{(\Sigma X_{ij})(\Sigma Y_{ij})}{n}\right]-\frac{(\Sigma\Sigma X_{ij}Y_{ij})^2}{N}.$$

Thus, the adjusted sums of squares are

(11) $$SS_W' = SS_{W_Y} - \frac{(rSS_{W_{XY}})^2}{r^2SS_{W_X}},$$

$$= SS_{W_Y} - \frac{(SS_{W_{XY}})^2}{SS_{W_X}};$$

(12) $$SS'_T = SS_{T_Y} - \frac{(rSS_{W_{XY}} + SS_{B_{XY}})^2}{r^2SS_{W_X} + SS_{B_X}},$$

(13) $$SS'_B = SS'_T - SS'_W.$$

Note that the adjusted within groups sum of squares remains unchanged by the substitution of $\hat{T}$ for X but the substitution does alter the adjusted total sum of squares and consequently, the adjusted between groups sum of squares.

A question arises concerning what value to use as the estimate of reliability in the formulae. A solution to this problem was proposed by Campbell and Erlebacher (1970).

> In a pretest-posttest situation one may find it reasonable to make two assumptions that would generate appropriate common-factor coefficients. First, if one has only the pretest-posttest correlations, one may assume that the correlation in the experimental group was unaffected by the treatment. (We need a survey of experience in true experiments to check on this.) Second, one may assume that the common-factor coefficient is the same for both pretest and posttest. Under these assumptions, the pretest-posttest correlation coefficient itself becomes the relevant common-factor coefficient for the pretest or covariate, the "reliability" to be used in Lord's and Porter's formulas. (p. 200)

This recommendation will be followed in this study, thus the correlation between the pretest scores and posttest scores will be used as the reliability estimate in the formulae.

Summary

As noted, analysis of covariance and analysis of covariance with Porter's adjustment have been recommended by several authors. Campbell and Erlebacher (1970) computer generated data for two overlapping groups with no true treatment effect and concluded that the analysis of covariance method was inappropriate and that analysis of covariance with Porter's adjustment should only be undertaken with great tentativeness. Porter (1967) computer generated data to compare the *F* sampling distribution with the theoretical *F* distribution using his adjustment. He concluded that samples of 20 or larger were needed to have a useful approximation to the theoretical *F* distribution. He also found that the estimation degenerated further when the reliability was less than .7. No study was found which compared both techniques under similar conditions for both gain and no-gain groups.

CHAPTER III
PROCEDURES

The analysis of covariance and the analysis of covariance with Porter's adjustment were compared under forty-eight sets of conditions on the basis of computer generated data. Six combinations of reliability, two sample sizes, two levels of gain, and two different sets of pretest means were used. Both equal and unequal reliabilities were used in the comparisons. A random sample of two thousand observations was generated under each combination of conditions and analyzed by analysis of covariance with and without Porter's adjustment.

The sampling distributions of the two analyses were statistically compared with a sign test for each of the forth-eight sets of conditions. The computer generated alpha values and powers were then used as the criterion variables in four factorial experiments. Subsequent *a posteriori* analyses were performed where warranted.

The Model

A standard model was used to represent the pretest and posttest scores of one subject. The model follows the traditional measurement approach as found in Gulliksen (1950) or Lord and Novick (1968) and extended to gain score theory by O'Connor (1972).

(14) $$X = T + E_1$$

and

(15) $$Y = T + G + E_2$$

where

X = observed pretest score,

Y = observed posttest score,

T = true pretest score

G = true gain,

E_1= random measurement error in pretest score,

E_2= random measurement error in posttest score.

The following properties about E_1 and E_2 are assumed to exist:

The errors E_1 and E_2

i) have zero means in the group tested,

ii) have the same variances for both groups,

iii) are independent of each other and of the true parts of each test.

It is further assumed that T and G are independent (across subjects) and that all components follow a normal probability distribution. These assumptions parallel Lord's (1956) stated and implied assumptions.

Selecting the Reliabilities

Reliability is related by definition to the variances of the observed scores, the true scores, and error. This section shows how that relationship can be used to establish desired reliabilities. The basic definition of reliability (Helmstadter, 1964, p. 62) is given by equation (16) where the symbol, ρ_{XX}, denotes reliability.

$$\rho_{XX} = \frac{\sigma_X^2 - \sigma_E^2}{\sigma_X^2} . \tag{16}$$

Then,

(17) $$\sigma^2_{E_1} = \sigma^2_X(1-\rho_{XX}).$$

Choosing the variance of X <u>a priori</u> to be 100, the variance of E_1 can be found in the following manner as the reliability of X assumes different values:

(18) $$\sigma^2_{E_1} = 100(1-\rho_{XX}).$$

The independence of T and E_1 in (14) implies

(19) $$\sigma^2_X = \sigma^2_T + \sigma^2_{E_1}.$$

Combining (18) and (19) and solving for σ^2_T yields

(20) $$\sigma^2_T = 100\rho_{XX}.$$

The posttest variances can be selected in a similar manner. Based on the assumptions,

(21) $$\sigma^2_Y = \sigma^2_T + \sigma^2_G + \sigma^2_{E_2}.$$

The variance of the gain scores is chosen in the manner prescribed in Chapter I. Then, combining (16) and (21),

(22) $$\sigma^2_{E_2} = \frac{\sigma^2_T + \sigma^2_G}{\rho_{YY}} - \sigma^2_T - \sigma^2_G$$

where ρ_{YY} denotes the established reliability of the posttest scores. Thus it can be seen that the effect of selecting specified reliabilities can be obtained by selecting the variances of E_1, E_2, and T in accordance with (18), (20), and (22) respectively.

<u>The Regression of Y on X</u>

Recall that one of the assumptions necessary for analysis of covariance (Cochran, 1957) is that the regression slopes of the dependent variable on the covariate must be equal for each treatment group. Let $\beta_{Y \cdot X}$ denote the regression slope for one treatment group.

Then,

(23) $$\beta_{Y \cdot X} = \rho_{XY} \cdot \frac{\sigma_Y}{\sigma_X}, \qquad \text{(Ferguson, 1971, p. 113),}$$

implies

(24) $$\beta_{Y \cdot X} = \frac{\sigma_{XY}}{\sigma_T^2 + \sigma_{E_1}^2} .$$

The covariance between X and Y is equal to the variance of T since it has been assumed that all the components of the pretest and posttest are independent except T with itself. Therefore

(25) $$\beta_{Y \cdot X} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_{E_1}^2},$$

and

(26) $$\beta_{Y \cdot X} = \rho_{XX}$$

by equation (16). Thus, the slope, Y on X, of any group is equal to the reliability of its pretest. If the reliabilities of the pretests were the same for both groups, the assumptions concerning slopes would be satisfied.

From equation (16), it can be seen that the reliability of a test is dependent to a certain extent upon the variability of the sample for which it is given. In equation (16), σ_X^2 is in both the numerator and denominator of the fraction with σ_E^2 being subtracted from the numerator. As σ_X^2 increases both the numerator and denominator increase by the same amount assuming σ_E^2 is not changed thus, ρ_{XX} increases. Two groups of equal ability would likely produce similar variances, hence similar reliabilities. However, the technique is often used for comparing compensatory programs with a comparison group sampled from the general population of untreated children in the same community such as with the Westinghouse/Ohio University study (Campbell and Erlebacher, 1970). These comparison children tend to be higher in ability than the treatment group. The resulting differences in variation tend to produce different reliabilities. In order to simulate such a situation this study used different reliabilities for the gain and the no-gain groups in addition to the comparisons when the reliabilities are equal. The reliabilities for the gain group are decreased by fifteen percent which roughly approximates the reduced variation empirically observed in Project Follow Through achievement data.

Selecting Means

The true score mean was set a priori at 100 when both the gain and no-gain groups have equal means. The situation in which the gain group has a lower mean was also analyzed. In this case, the true score mean of the gain group was set at 80. These values have been empirically chosen based on Project Follow Through data. Based on the assumptions,

(27) $$E(X) = E(T+E_1) = \mu_T,$$

hence the mean of the observed pretest scores is equal to the mean of the true scores. Also

(28) $$E(Y) = E(T+G+E_2) = \mu_T+\mu_G.$$

Thus, the mean of the observed posttest scores is equal to the sum of the means of the true scores and the gain scores.

Clearly in the case of the no-gain group and in both groups where no gain was used, the mean gain was zero. The selected value of μ_G for the gain situation was based on power considerations, that is, μ_G was chosen such that the power of an F test for analysis of covariance was .50.

A linear model representation of the analysis of covariance for two groups is

$$Y = \beta_0 + \beta_1 X + \beta_2 W + \varepsilon \tag{29}$$

where X is the pretest score (covariate), Y is the posttest score, and W is a dummy variable designating group membership (W=1 if gain group, 0 if no-gain group). Testing the hypothesis that β_2 is equal to 0 in equation (29) is equivalent to the F test for treatments in the analysis of covariance procedure. It can be shown that β_2 in equation (29) is equivalent to the mean gain, μ_G. The mean of the posttest for the gain group is

$$E(Y_G) = \beta_0 + \beta_1 \mu_{X_G} + \beta_2 \tag{30}$$

where Y_G is a posttest score for the gain group and μ_{X_G} is the mean pretest score for the gain group. Likewise, the mean of the posttest for the no-gain group is

$$E(Y_{NG}) = \beta_0 + \beta_1 \mu_{X_{NG}} \tag{31}$$

where Y_{NG} is a pretest score for the no-gain group, $\mu_{X_{NG}}$ is the mean pretest score for the no gain group. But it has been assumed that μ_{X_G} and $\mu_{X_{NG}}$ are equal. Furthermore, the gain group has mean gain, μ_G and the no-gain group has mean gain zero, thus

$$\mu_G = E(Y_G) - E(Y_{NG}) = \beta_2 \ . \tag{32}$$

Hence, choosing the value of β_2 that yields a power of .50 is equivalent to choosing a value of μ_G to produce a power of .50 in the analysis of covariance procedure.

This power can be obtained from the following probability statement:

(33) $\Pr[\underline{t}^{*}>\underline{t}] = .50$

where $\underline{t}^{*}$ is a noncentral $\underline{t}$ statistic. This expression can be approximated by the substitution of a $\underline{z}$ statistic for $\underline{t}^{*}$.

(34) $$\Pr\left[\frac{\hat{\beta}_2-\beta_2}{\sigma_{\hat{\beta}_2}} > \underline{t}\right] = .50 .$$

Thus a value of β_2 can be chosen such that the substituted $\underline{z}$ statistic is equal to $\underline{t}_{.025}$ with the appropriate degrees of freedom. This value of β_2 is the value of the average gain, μ_G, such that the power of analysis of covariance is .50 under the condition of perfect reliability. In order to find this value of μ_G, a numerical expression for $\sigma^2_{\hat{\beta}_2}$ is needed.

The variance of $\hat{\beta}_2$, $\sigma^2_{\hat{\beta}_2}$, can be approximated in the following manner. Karmel and Polasek (1970, p. 245) state that $\sigma^2_{\hat{\beta}_2}$ is

(35) $$\sigma^2_{\hat{\beta}_2} = \sigma^2_Y \frac{\Sigma(X-\overline{X})^2}{[\Sigma(X-\overline{X})^2][\Sigma(W-\overline{W})^2]-[\Sigma(X-\overline{X})(W-\overline{W})]^2} .$$

Dividing both the numerator and denominator by N^2 and substituting population variances for sample variances, $\sigma^2_{\hat{\beta}_2}$ is approximately equal to the following:

(36) $$\sigma^2_{\hat{\beta}_2} \simeq \frac{\sigma^2_Y}{N} \frac{\sigma^2_X}{\sigma^2_X\,\sigma^2_W-(\sigma_{XW})^2} .$$

The quantity, σ_{XW}, the covariance between X and W has been assumed equal to zero thus, in equation (36), the σ^2_X's divide out. Hence

$$\sigma^2_{\hat{\beta}_2} = \frac{\sigma^2_Y}{N} \frac{1}{\sigma^2_W} . \tag{37}$$

Under the conditions assumed for the model, the variance of Y, σ^2_Y, is equal to 4. The variance of W, σ^2_W can be computed to be .25. Thus, by substitution, $\sigma^2_{\hat{\beta}_2}$ equals .80 when N is equal to 20 and $\sigma^2_{\hat{\beta}_2}$ equals .08 when N is equal to 200.

Hence, for N = 20, μ_G can be found by the following expression:

$$\mu_{G_{20}} = \underline{+}t_{.025,17}\sigma_{\hat{\beta}_2} = 2.11\sqrt{.80} = 1.88. \tag{38}$$

Likewise, for N = 200, μ_G can be found by solving equation (39).

$$\mu_{G_{200}} = \underline{+}t_{.025,197}\sigma_{\hat{\beta}_2} = 1.97\sqrt{.08} = .56. \tag{39}$$

The approximated values of μ_G were tested using Monte Carlo generated variables and found to indeed produce a power of .50.

The approximated values of μ_G under each set of conditions are shown in Tables 1 and 2.

TABLE 1

CONDITIONS UNDER WHICH COMPARISONS WERE MADE WHEN THE RELIABILITIES FOR BOTH GROUPS WERE EQUAL

RELIABILITY	GROUP SAMPLE SIZE	GAIN GROUP MEAN GAIN	PRETEST MEAN FOR GAIN GROUP	PRETEST MEAN FOR NO-GAIN GROUP
.90	10	0	100	100
.90	10	0	80	100
.90	100	0	100	100
.90	100	0	80	100
.70	10	0	100	100
.70	10	0	80	100
.70	100	0	100	100
.70	100	0	80	100
.50	10	0	100	100
.50	10	0	80	100
.50	100	0	100	100
.50	100	0	80	100
.90	10	1.88	100	100
.90	10	1.88	80	100
.90	100	.56	100	100
.90	100	.56	80	100
.70	10	1.88	100	100
.70	10	1.88	80	100
.70	100	.56	100	100
.70	100	.56	80	100
.50	10	1.88	100	100
.50	10	1.88	80	100
.50	100	.56	100	100
.50	100	.56	80	100

TABLE 2

CONDITIONS UNDER WHICH COMPARISONS WERE MADE WHEN THE RELIABILITIES FOR BOTH GROUPS WERE UNEQUAL

RELIABILITY FOR GAIN GROUP	RELIABILITY FOR NO-GAIN GROUP	GROUP SAMPLE SIZE	GAIN GROUP MEAN GAIN	PRETEST MEAN FOR GAIN GROUP	PRETEST MEAN FOR NO-GAIN GROUP
.76	.90	10	0	100	100
.76	.90	10	0	80	100
.76	.90	100	0	100	100
.76	.90	100	0	80	100
.60	.70	10	0	100	100
.60	.70	10	0	80	100
.60	.70	100	0	100	100
.60	.70	100	0	80	100
.42	.50	10	0	100	100
.42	.50	10	0	80	100
.42	.50	100	0	100	100
.42	.50	100	0	80	100
.76	.90	10	1.88	100	100
.76	.90	10	1.88	80	100
.76	.90	100	.56	100	100
.76	.90	100	.56	80	100
.60	.70	10	1.88	100	100
.60	.70	10	1.88	80	100
.60	.70	100	.56	100	100
.60	.70	100	.56	80	100
.42	.50	10	1.88	100	100
.42	.50	10	1.88	80	100
.42	.50	100	.56	100	100
.42	.50	100	.56	80	100

Generation of Random Normal Deviates with Specified Means and Variances

The study required the use of computer generated normally distributed random variables with specified means and variances. Two thousand sets of variables were generated for each of the forty-eight sets of conditions. Muller (1959) identified and compared six methods of generating normal deviates on the computer. A method described by Box and Muller (1958) was judged most attractive from a mathematical standpoint. According to Muller (1959, p. 379), "Mathematically this approach has the attractive advantage that the transformation for going from uniform deviates to normal deviates is exact." This method was endorsed by Marsaglia and Bray (1964). They modified the algorithm to reduce central processing computer time without altering its accuracy.

The method first requires the generation of two independent uniform random variables, U_1 and U_2, over the interval (-1, 1). The variables

$$Z_1 = U_1[-2 \ln(U_1^2+U_2^2) \;/\; (U_1^2 + U_2^2)]^{1/2}$$

and

$$Z_2 = U_2[-2 \ln(U_1^2+U_2^2) \;/\; (U_1^2 + U_2^2)]^{1/2}$$

will be two independent random variables from the same normal distribution with mean zero and unit variance. The variables were then transformed to have the desired means and variances.

Analysis for Comparing the Selected Methods

Each of the two thousand sets of generated data for the forty-eight sets of conditions was analyzed by analysis of covariance and analysis of covariance with Porter's adjustment. The number of times the *F* statistic from analysis of covariance exceeded the *F* statistic from analysis of covariance with Porter's adjustment was noted along with the proportion of rejections by each method of analysis.

A sign test (Siegel, 1956, p. 63-67) was then performed for each set of conditions to test the null hypotheses in Chapter I, that is, "the sampling distributions of the test statistics from analysis of covariance with and without Porter's adjustment will be the same for each set of conditions." These tests were run at the .01 level of significance.

The fraction of rejections noted (alphas and powers) was then used as the dependent variable in four factorial experiments to gain further insight into what factors affected each method of analysis being studied. Each factorial experiment had three factors. These were reliability which included six combinations of reliability used in the study, sample size, which included n = 10 and n = 100 subjects per group, and the equality of pretest means, which included a level where both pretest means were equal and a level where they differed. The layout of the factorial experiments is illustrated in Table 3. The factorial experiment illustrated is the situation for which there was no gain in either group and the standard analysis of covariance was used. Thus, the dependent variables are the alpha values generated by the computer. When significant main effects or interactions occurred, the appropriate _a posteriori_ analytical procedures to locate the sources of the variation were followed.

TABLE 3

FACTORIAL DESIGN FOR STANDARD ANALYSIS OF COVARIANCE WHEN MEAN GAIN OF GAIN GROUP WAS ZERO AND CRITERION VARIABLES ARE MONTE CARLO GENERATED ALPHA VALUES

RELIABILITY (GAIN GROUP AND NO-GAIN GROUP)

	90 90		70 70		50 50		76 90		60 70		42 50	
SAMPLE SIZE	10	100	10	100	10	100	10	100	10	100	10	100
PRETEST MEANS SAME	.048	.050	.048	.050	.056	.044	.056	.055	.050	.046	.054	.042
PRETEST MEANS DIFFERENT	.089	.050	.232	.980	.374	.989	.134	.794	.259	.994	.400	1.000

CHAPTER IV
RESULTS

The number of times the $\underline{F}$ statistic from the standard analysis of covariance exceeded the $\underline{F}$ statistic from analysis of covariance with Porter's adjustment is listed for each set of conditions in Tables 4, 5, 6, and 7. These quantities were used as test statistics for the sign tests used to test the hypothesis,"There is no difference between the sampling distributions of the test statistics from analysis of covariance with and without Porter's adjustment." This hypothesis was tested for each of the forty-eight sets of conditions at the .01 level of significance.

Siegel (1956) stated that a large sample test statistic for the sign test is

(40) $$\underline{z} = \frac{x-.5N}{.5\sqrt{N}}$$

where N is equal to the number of pairs of observations, x is equal to the number of times the first measurement of the pair exceeds the second measurement of the pair, and $\underline{z}$ is the standard normal variate. For a level of .01, the null hypotheses would be rejected when $\underline{z}$ was less than -2.33 or greater than 2.33. This is equivalent to rejecting the null hypotheses when x was less than 949 or greater than 1051 and N equals 2000. Thus, an inspection of the last column of Tables 4, 5, 6, and 7 reveals that the null hypothesis was rejected for each of the forty-eight sets of combinations.

The analysis of variance summary tables are presented in Tables 8, 9, 10, and 11. The analysis of variance summary table for the case when the Monte Carlo generated alpha values from the standard analysis of covariance were used as the criterion variables is presented in Table 8. The analysis of variance summary table for the case when the Monte Carlo generated alpha values from the analysis of covariance with Porter's adjustment were used as the criterion variables is presented in Table 9. The analysis of variance summary tables for the cases when Monte Carlo generated powers were used as the criterion variables for standard analysis of covariance and analysis of covariance with Porter's adjustment are presented in Tables 10 and 11, respectively. Each analysis of variance table includes analyses of the simple effects where they are warranted. Scheffe's S method for testing linear contrasts is included in Table 8. These linear contrasts are defined in Table 12. Each $\underline{F}$ statistic which exceeds the critical value at the .05 level is denoted by an asterisk.

A result of particular interest is applicable to analysis of covariance both with and without Porter's adjustment. That is, when a spuriously high fraction of significant $\underline{F}$ statistics occurred when there was a mean gain of zero in both groups and the pretest means differed, the adjusted posttest means indicated that the gain was in favor of the group having the largest pretest mean. This result was more pronounced for lower reliabilities.

When the gain group had a positive gain, the no-gain group had a mean gain of zero, and the no-gain group also had a larger pretest mean, a similar situation occurred. In this situation, when a significant $\underline{F}$ statistic occurred, the adjusted posttest means usually indicated the no-gain group had recorded the larger gain. Again, these results became more pronounced as the reliability of the scores were reduced.

TABLE 4

FRACTION OF SIGNIFICANT F'S AND NUMBER OF TIMES F_S EXCEEDED F_P WHERE THERE WAS NO GAIN IN EITHER GROUP AND THE PRETEST RELIABILITIES WERE EQUAL

RELIABILITY OF GAIN GROUP SCORES	RELIABILITY OF NO-GAIN GROUP SCORES	SAMPLE SIZE OF EACH GROUP	PRETEST MEAN OF GAIN GROUP	PRETEST MEAN OF NO-GAIN GROUP	FRACTION OF SIGNIFICANT F'S FOR STANDARD ANALYSIS OF COVARIANCE[a]	FRACTION OF SIGNIFICANT F'S FOR ANALYSIS OF COVARIANCE WITH PORTER'S ADJUSTMENT[a]	NUMBER OF TIMES STANDARD F EXCEEDED PORTER'S F[b]
.90	.90	10	100	100	.048	.053	1366
.90	.90	10	80	100	.089	.064	1624
.90	.90	100	100	100	.050	.053	827
.90	.90	100	80	100	.540	.136	1784
.70	.70	10	100	100	.048	.059	796
.70	.70	10	80	100	.232	.097	1803
.70	.70	100	100	100	.050	.066	846
.70	.70	100	80	100	.980	.543	1988
.50	.50	10	100	100	.056	.068	681
.50	.50	10	80	100	.374	.154	1927
.50	.50	100	100	100	.044	.080	795
.50	.50	100	80	100	.999	.888	1992

[a]Monte Carlo derived alpha values.
[b]Sign test is significant if column entry is greater than 1051 or less than 949.

TABLE 5

FRACTION OF SIGNIFICANT *F*'S AND NUMBER OF TIMES F_S EXCEEDED F_P WHERE THERE WAS NO GAIN IN EITHER GROUP AND THE PRETEST RELIABILITIES WERE NOT EQUAL

RELIABILITY OF GAIN GROUP SCORES	RELIABILITY OF NO-GAIN GROUP SCORES	SAMPLE SIZE OF EACH GROUP	PRETEST MEAN OF GAIN GROUP	PRETEST MEAN OF NO-GAIN GROUP	FRACTION OF SIGNIFICANT *F*'S FOR STANDARD ANALYSIS OF COVARIANCE[a]	FRACTION OF SIGNIFICANT *F*'S FOR ANALYSIS OF COVARIANCE WITH PORTER'S ADJUSTMENT[a]	NUMBER OF TIMES STANDARD *F* EXCEEDED PORTER'S *F*[b]
.76	.90	10	100	100	.056	.062	627
.76	.90	10	80	100	.134	.072	1542
.76	.90	100	100	100	.055	.068	1612
.76	.90	100	80	100	.794	.269	1903
.60	.70	10	100	100	.050	.058	784
.60	.70	10	80	100	.259	.102	1139
.60	.70	100	100	100	.046	.072	322
.60	.70	100	80	100	.994	.656	1938
.42	.50	10	100	100	.054	.052	617
.42	.50	10	80	100	.400	.179	1954
.42	.50	100	100	100	.042	.078	1380
.42	.50	100	80	100	1.000	.926	1982

[a]Monte Carlo derived alpha values.
[b]Sign test is significant if column entry is greater than 1051 or less than 949.

TABLE 6

FRACTION OF SIGNIFICANT *F*'S AND NUMBER OF TIMES F_s EXCEEDED F_p WHERE THERE WAS GAIN IN THE GAIN GROUP AND THE PRETEST RELIABILITIES WERE EQUAL

RELIABILITY OF GAIN GROUP SCORES	RELIABILITY OF NO-GAIN GROUP SCORES	SAMPLE SIZE OF EACH GROUP	MEAN GAIN OF GAIN GROUP	PRETEST MEAN OF GAIN GROUP	PRETEST MEAN OF NO-GAIN GROUP	FRACTION OF SIGNIFICANT *F*'S FOR STANDARD ANALYSIS OF COVARIANCE[a]	FRACTION OF SIGNIFICANT *F*'S FOR ANALYSIS OF COVARIANCE WITH PORTER'S ADJUSTMENT[a]	NUMBER OF TIMES STANDARD *F* EXCEEDED PORTER'S *F*[b]
.90	.90	10	1.88	100	100	.123	.128	1114
.90	.90	10	1.88	80	100	.057	.064	843
.90	.90	100	.56	100	100	.126	.130	627
.90	.90	100	.56	80	100	.318	.055	1578
.70	.70	10	1.88	100	100	.074	.094	891
.70	.70	10	1.88	80	100	.130	.060	1643
.70	.70	100	.56	100	100	.087	.106	821
.70	.70	100	.56	80	100	.958	.393	1934
.50	.50	10	1.88	100	100	.074	.090	882
.50	.50	10	1.88	80	100	.277	.104	1975
.50	.50	100	.56	100	100	.073	.116	1389
.50	.50	100	.56	80	100	.998	.802	1972

[a]Monte Carlo derived powers.
[b]Sign test is significant if column entry is greater than 1051 or less than 949.

TABLE 7

FRACTION OF SIGNIFICANT F'S AND NUMBER OF TIMES F_s EXCEEDS F_p WHERE THERE WAS GAIN IN THE GAIN GROUP AND THE PRETEST RELIABILITIES WERE NOT EQUAL

RELIABILITY OF GAIN GROUP SCORES	RELIABILITY OF NO-GAIN GROUP SCORES	SAMPLE SIZE OF EACH GROUP	MEAN GAIN OF GAIN GROUP	PRETEST MEAN OF GAIN GROUP	PRETEST MEAN OF NO-GAIN GROUP	FRACTION OF SIGNIFICANT F'S FOR STANDARD ANALYSIS OF COVARIANCE[a]	FRACTION OF SIGNIFICANT F'S FOR ANALYSIS OF COVARIANCE WITH PORTER'S ADJUSTMENT[a]	NUMBER OF TIMES STANDARD F EXCEEDED PORTER'S F[b]
.76	.90	10	1.88	100	100	.101	.104	783
.76	.90	10	1.88	80	100	.070	.061	1127
.76	.90	100	.56	100	100	.101	.104	816
.76	.90	100	.56	80	100	.668	.135	1901
.60	.70	10	1.88	100	100	.074	.080	1189
.60	.70	10	1.88	80	100	.154	.066	1883
.60	.70	100	.56	100	100	.075	.102	792
.60	.70	100	.56	80	100	.983	.480	1937
.42	.50	10	1.88	100	100	.081	.090	1146
.42	.50	10	1.88	80	100	.280	.094	1554
.42	.50	100	.56	100	100	.063	.102	541
.42	.50	100	.56	80	100	1.000	.874	1987

[a]Monte Carlo derived powers.
[b]Sign test is significant if column entry is greater than 1051 or less than 949.

TABLE 8

ANOVA SUMMARY TABLE USING MONTE CARLO GENERATED ALPHAS FROM THE STANDARD ANALYSIS OF COVARIANCE AS THE CRITERION VARIABLES

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F
Pretest Mean	1.59960	1	1.59960	(544.08)
PM at R_1	.07049	1	.07049	23.98*
PM at R_2	.33582	1	.33582	114.22*
PM at R_3	.40513	1	.40513	137.80*
PM at R_4	.16687	1	.16687	56.76*
PM at R_5	.33466	1	.33466	113.83*
PM at R_6	.42510	1	.42510	144.59*
PM at SS_{10}	.11525	1	.11525	39.20*
PM at SS_{100}	2.10000	1	2.10000	714.29*
Reliability	.10996	5	.02199	(7.48)
R at PM_1	.00008	5	.00001	<1.00
R at PM_2	.22278	5	.04455	15.15*
Ψ_1	.00018	1	.00018	<1.00
Ψ_2	.10306	1	.10306	35.05*
Ψ_3	.73933	1	.73933	251.47*
Ψ_4	.04743	1	.04743	16.13*
Sample Size	.59977	1	.59977	(204.00)
SS at PM_1	.00005	1	.00005	<1.00
SS at PM_2	1.21540	1	1.21540	413.40*
PM x R	.11291	5	.02258	7.68*
PM x SS	.61568	1	.61568	209.41*
R x SS	.01470	5	.00294	1.00
Residual	.01468	5	.00294	
Total	3.06730	23		

*Sample statistic greater than critical value at .05 level.

TABLE 9

ANOVA SUMMARY TABLE USING MONTE CARLO GENERATED ALPHAS FROM ANALYSIS OF COVARIANCE WITH PORTER'S ADJUSTMENT AS THE CRITERION VARIABLES

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F
Pretest Mean	.45844	1	.45844	(25.08)
PM at SS_{10}	.00832	1	.00832	<1.00
PM at SS_{100}	.75050	1	.75050	41.06*
Reliability	.17552	5	.03510	1.92
Sample Size	.33018	1	.33018	(18.06)
SS at PM_1	.00035	1	.00035	<1.00
SS at PM_2	.63021	1	.63021	34.48*
PM x R	.15727	5	.03145	1.72
PM x SS	.30038	1	.30038	16.43*
R x SS	.10189	5	.02038	1.11
Residual	.09139	5	.01828	
Total	1.61507	23		

*Sample statistic greater than critical value at .05 level.

TABLE 10

ANOVA SUMMARY TABLE USING MONTE CARLO GENERATED POWERS FROM THE STANDARD ANALYSIS OF COVARIANCE AS THE CRITERION VARIABLES

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F
Pretest Mean	.97768	1	.97768	(85.54)
PM at SS_{10}	.01620	1	.01620	1.42
PM at SS_{100}	1.61334	1	1.61334	141.15*
Reliability	.12431	5	.02486	2.17
Sample Size	.65076	1	.65076	(56.93)
SS at PM_1	.00001	1	.00001	<1.00
SS at PM_2	1.30482	1	1.30482	114.16*
PM x R	.19910	5	.03982	3.48
PM x SS	.65406	1	.65406	57.22*
R x SS	.05665	5	.01133	<1.00
Residual	.05714	5	.01143	
Total	2.71970	23		

*Sample statistic greater than critical value at .05 level.

TABLE 11

ANOVA SUMMARY TABLE USING MONTE CARLO GENERATED POWERS FROM ANALYSIS OF COVARIANCE WITH PORTER'S ADJUSTMENT AS THE CRITERION VARIABLES

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F
Pretest Mean	.15714	1	.15714	(6.48)
PM at SS_{10}	.00157	1	.00157	<1.00
PM at SS_{100}	.36018	1	.36018	14.85*
Reliability	.14179	5	.02836	1.17
Sample Size	.23285	1	.23285	(9.60)
SS at PM_1	.00046	1	.00046	<1.00
SS at PM_2	.43701	1	.43701	18.01*
PM x R	.16990	5	.03398	1.40
PM x SS	.20461	1	.20461	8.43*
R x SS	.13367	5	.02673	1.10
Residual	.12132	5	.02426	
Total	1.16127	23		

*Sample statistic greater than critical value at .05 level.

TABLE 12

COEFFICIENTS FOR THE LINEAR CONTRASTS USED WHEN RELIABILITY WAS SIGNIFICANT IN THE ANALYSIS OF VARIANCE

		LEVELS OF RELIABILITY					
RELIABILITY OF GAIN GROUP		90	70	50	76	50	42
RELIABILITY OF NO GAIN GROUP		90	70	50	90	70	50
CONTRAST	Ψ_1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$
	Ψ_2	$\frac{1}{2}$	$-\frac{1}{2}$	0	$\frac{1}{2}$	$-\frac{1}{2}$	0
	Ψ_3	$\frac{1}{2}$	0	$-\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
	Ψ_4	0	$\frac{1}{2}$	$-\frac{1}{2}$	0	$\frac{1}{2}$	$-\frac{1}{2}$

CHAPTER V
DISCUSSION

In general, the results of this study support the positions of Lord (1967, 1969), Campbell and Erlebacher (1970), and O'Connor (1972) with respect to their warnings about the implications of analysis of covariance using unreliable test scores. The study shows that even when there is no gain in either group, a much higher fraction of rejections occur than would be expected. The fraction of rejections is even more extreme when the reliability is .70 or lower and the pretest means differ. In addition, Porter's adjustment seems to offer little improvement.

Comparison of the Two Methods of Analysis

The rejection of all forty-eight null hypotheses concerning the equality of the sampling distributions for the two methods of analysis shows that there is a difference in the results obtained from analysis of covariance and analysis of covariance with Porter's adjustment. A closer examination shows that these differences are more extreme when the pretest means of the two groups differ and the reliabilities are low. Further study shows that although the two methods of analysis are different, neither method does an adequate job of modeling reality, that is, both methods tend to produce erroneous proportions of type I and II errors when pretest means differ and the reliabilities are low. When the pretest means are equal, the power of the tests seem to be directly related in a positive manner to the reliabilities when other variables are held constant.

Possibly the most far reaching results were a function of reliability. The data indicated that incorrect decisions about which group had the larger gain could be made using either method of analysis when the pretest means of the two groups differed and the reliabilities were low. When the pretest mean of the gain group was less than that of the no-gain group, the adjusted posttest means showed that the no-gain group was superior both when the mean gain was zero in both groups and when the mean gain was positive only in the gain group. This possibility was pointed out by Lord (1967) and Campbell and Erlebacher (1970).

Factors that Affect Alpha and Beta

The factorial experiments using the computer generated alphas and betas allow one to infer which factors affect the levels of type I and type II errors. Using Monte Carlo generated alphas from the standard analysis of covariance in a factorial experiment indicated that an interaction between pretest means and reliability and an interaction between sample size and pretest means affected the alphas significantly. Further analyses (simple effects) showed that the pretest mean factor was significant at every level of reliability. Both reliability and sample size were significant when the pretest means differed. The tests of linear contrasts showed that there was no significant difference between equal reliabilities and unequal reliabilities when the pretest means differed, but there were significant differences among the levels of reliability when the pretest means differed. These results indicate that a difference in reliabilities between groups, thus a difference in slopes, has no effect, however, the level of reliability does.

The other three factorial experiments indicated that interactions between pretest means and sample size were the major contributors to the differing levels of alpha and power. In all three experiments (See Tables 10, 11, and 12) the sample size was significant when the pretest means differed. Reliability seems to have a somewhat moderate effect in these cases.

Predicting Alpha and Power

Using the results of this study, a regression equation can be set up to predict the experimental probability of a type I error or the experimental probability of rejecting a false null hypotheses when the established probabilities are .05 and .50 respectively. The basic regression equation is

$$(41) \qquad Y = \beta_0+\beta_1 R_G+\beta_2 R_N+\beta_3 S+\beta_4 M+\beta_5 R_G R_N+\beta_6 R_G S +\beta_7 R_G M+\beta_8 R_N S+\beta_9 R_N M+\beta_{10} SM+\varepsilon$$

where

Y = the predicted alpha or power,

R_G = the reliability of the gain group scores,

R_N = the reliability of the no gain group scores,

S = sample size

M = 1 if the pretest means are equal, 0 otherwise, and

β_i = the regression coefficient, i.

Tables 14 and 15 provide confidence intervals for the expected values of alpha and power respectively for each set of conditions noted.

TABLE 13

NINETY-FIVE PERCENT CONFIDENCE INTERVALS FOR THE EXPECTED VALUES OF ALPHAS UNDER SPECIFIED CONDITIONS

RELIABILITY OF GAIN GROUP SCORES	RELIABILITY OF NO GAIN GROUP SCORES	GROUP SAMPLE SIZE	EQUALITY OF PRETEST MEANS*	LOWER CONFIDENCE LIMIT OF ALPHA	UPPER CONFIDENCE LIMIT OF ALPHA
.90	.90	10	1	.000	.137
.90	.90	10	0	.000	.137
.90	.90	100	1	.000	.100
.90	.90	100	0	.556	.741
.70	.70	10	1	.017	.162
.70	.70	10	0	.108	.325
.70	.70	100	1	.000	.141
.70	.70	100	0	.780	.945
.50	.50	10	1	.000	.125
.50	.50	10	0	.297	.451
.50	.50	100	1	.000	.120
.50	.50	100	0	.932	1.000
.76	.90	10	1	.000	.136
.76	.90	10	0	.043	.223
.76	.90	100	1	.036	.143
.76	.90	100	0	.691	.871
.60	.70	10	1	.003	.136
.60	.70	10	0	.228	.361
.60	.70	100	1	.014	.146
.60	.70	100	0	.879	1.000
.42	.50	10	1	.000	.097
.42	.50	10	0	.307	.472
.42	.50	100	1	.000	.117
.42	.50	100	0	.967	1.000

*If 1, the pretest means are equal and if 0, the pretest means are not equal.

TABLE 14

NINETY-FIVE PERCENT CONFIDENCE INTERVALS FOR THE EXPECTED VALUES OF POWER UNDER SPECIFIED CONDITIONS

RELIABILITY OF GAIN GROUP SCORES	RELIABILITY OF NO GAIN GROUP SCORES	GROUP SAMPLE SIZE	EQUALITY OF PRETEST MEANS*	LOWER CONFIDENCE LIMIT OF POWER	UPPER CONFIDENCE LIMIT OF POWER
.90	.90	10	1	.000	.300
.90	.90	10	0	.000	.115
.90	.90	100	1	.000	.191
.90	.90	100	0	.351	.666
.70	.70	10	1	.019	.266
.70	.70	10	0	.048	.295
.70	.70	100	1	.000	.236
.70	.70	100	0	.677	.924
.50	.50	10	1	.000	.176
.50	.50	10	0	.155	.417
.50	.50	100	1	.000	.224
.50	.50	100	0	.863	1.000
.76	.90	10	1	.000	.265
.76	.90	10	0	.000	.199
.76	.90	100	1	.000	.235
.76	.90	100	0	.523	.829
.60	.70	10	1	.000	.210
.60	.70	10	0	.097	.323
.60	.70	100	1	.010	.236
.60	.70	100	0	.782	1.000
.42	.50	10	1	.000	.130
.42	.50	10	0	.156	.438
.42	.50	100	1	.000	1.000
.42	.50	100	0	.909	

*If 1, the pretest means are equal and if 0, the pretest means are not equal.

The predicted value of alpha or power can be found by using the data found in Tables 4, 5, 6, and 7 to fit regression equation (41) to obtain the following estimated regression parameters:

(42)
$$\begin{aligned} \text{alpha} = {} & .228 + .395R_G + .332R_N + .008\ S - .661M \\ & -1.09\ R_G R_N - .004\ R_G S + .619\ R_G M \\ & -.003\ R_N S + .195\ R_N M - .007\ SM \end{aligned}$$

and

(43)
$$\begin{aligned} \text{power} = {} & .045 + .536R_G + .404\ R_N + .010\ S - .700\ M \\ & -1.228\ R_G R_N - .006\ R_G S + .846\ R_G M \\ & + .002\ R_N S + .218\ R_N M - .007\ SM. \end{aligned}$$

A confidence interval for the predicted values of alpha and power could be obtained in the usual manner.

A Direction for Further Research

The results of this study combined with other research (Lord, 1967; Campbell and O'Connor, 1972), indicates a need for further study and the development of a robust method for comparing groups of differing ability when the scores are not perfectly reliable. Any new technique which is proposed should first be investigated under known conditions, either analytically, or by Monte Carlo techniques. This would insure that another inappropriate method is not used for the evaluation of compensatory education projects.

CHAPTER VI
SUMMARY

This study was designed to determine if either analysis of covariance or analysis of covariance with Porter's adjustment is an appropriate analytical procedure for evaluating educational pretest-posttest experiments. In particular, these methods were compared with respect to their use in the analysis of compensatory education projects where the groups may differ in ability.

The study was carried out by computer generating 2000 sets of normal data under forty-eight sets of predetermined conditions of reliability, sample size, gain, and equality of pretest means. Each of the 2000 sets of data for each set of conditions was then analyzed using both standard analysis of covariance and analysis of covariance with Porter's adjustment. A sign test was used to compare the two methods of analysis under each of the forty-eight sets of conditions. It was concluded that the two methods of analysis yielded different results.

Factors affecting the two methods of analysis were then studied separately using the computer generated alphas and powers as criterion variables in four factorial experiments. The factors included reliability at six levels, sample size at two levels, and the equality of pretest means at two levels. From these experiments, it was concluded that pretest means interacting with sample size and sometimes with reliability were significant factors. More specifically, sample size was statistically significant in each case where the pretest means differed. Also, pretest means were significant at every level of reliability for the computer generated alphas produced by the standard analysis of covariance.

It was also learned that when the pretest means differ, both standard analysis of covariance and analysis of covariance with Porter's adjustment produced erroneous results with respect to which group if either had a gain. When both groups had a mean gain of zero and the pretest means differed, significant results usually indicated that the group with the larger pretest mean had the gain. This would correspond to the control group sampled from the general population being credited with the gain in a compensatory education experiment. When there was a gain in only one group and the pretest mean was lower in that group, the analyses still indicated that the other group had the gain.

The results of this study point to the recommendations that analysis of covariance with or without Porter's adjustment should be approached with caution when the reliabilities are below .90 and the pretest means (covariate means) are likely to be different for the groups.

APPENDIX

```
C     FORTRAN PROGRAM WHICH PERFORMED DATA GENERATION AND ANALYSIS
C
C
      DIMENSION E1GN(100),E2GN(100),GAINGN(100),XX(200),YY(200),
     *X(100),E1NG(100),E2NG(100),GAINNG(100),TG(100),TN(100)
      REAL MSB,MSE,MSBP,MSEP
C
      NGROUP=0
C
C     INPUT OF GROUP PARAMETERS WHEREO
C         RELGN REPRESENTS THE RELIABILITY OF THE GAIN GROUP DATA
C         RELNG REPRESENTS THE RELIABILITY OF THE NO GAIN GROUP
C         NPERG REPRESENTS THE NUMBER OF OBSERVATIONS PER GROUP
C         GBAR REPRESENTS THE MEAN GAIN FOR GAIN GROUP
C         PRMNG REPRESENTS THE GAIN GROUP PRETEST MEAN
C         PRMNNG REPRESENTS THE NO GAIN GROUP PRETEST MEAN
C         ISEED REPRESENTS THE SEED FOR RANDOM NUMBER GENERATOR
C
      READ (5,99) ISEED,NGP,NSPL
 99   FORMAT (I8,2X,2I5)
100   READ (5,101) RELGN,RELNG,NPERG,GBAR,PRMNG,PRMNNG
101   FORMAT (2F3.2,I3,F3.2,2F3.0)
      NGROUP=NGROUP+1
      NSAMP=0
      NF10=0
      NF100=0
      NFP10=0
      NFP100=0
C
C     HEADER CARD FOR NEW SET OF PARAMETERS
C
      WRITE (6,102)
102   FORMAT ('1',T43,'F STATISTICS BASED ON THE FOLLOWING PARME'
     *,'TERS')
      WRITE (6,103)
103   FORMAT ('0',T22,'RELGN ',2X,'RELNG ',2X,'NPERG ',2X,' GBAR'
     *,' ',2X,'PRMNG ',2X,'PRMNNG')
      WRITE (6,104) RELGN,RELNG,NPERG,GBAR,PRMNG,PRMNNG
104   FORMAT (1X,T22,2F8.5,I5,3X,3F8.3)
      WRITE (6,105)
105   FORMAT ('0',T45,'SAMPLE NUMBER',5X,'STANDARD F',5X,'PORTER'
     *,'S F')
C
C     COMPUTE VARIANCE COMPONENTS
C
      VARE1G=100*(1-RELGN)
      VARE1N=100*(1-RELNG)
      VARTG=100*RELGN
      VARTN=100*RELNG
      VARGNG=.04*VARTG
      VARGNN=.04*VARTN
```

```
      VARE2G=(VARTG+VARGNG)/RELGN-VARTG-VARGNG
      VARE2N=(VARTN+VARGNN)/RELNG-VARTN-VARGNN
C
106   CONTINUE
      NSAMP=NSAMP+1
C
C     GENERATION OF DATA FOR GAIN GROUP
C     .
C
C     TRUE SCORES
C
      CALL RANGEN(NPERG,ISEED,X)
      DO 300 I=1,NPERG
      TG(I)=X(I)*SQRT(VARTG)+PRMNG
300   CONTINUE
C
C     PRETEST ERROR SCORES
C
      CALL RANGEN(NPERG,ISEED,X)
      DO 301 I=1,NPERG
      E1GN(I)=X(I)*SQRT(VARE1G)
301   CONTINUE
C
C     POSTTEST ERROR SCORES
C
      CALL RANGEN(NPERG,ISEED,X)
      DO 302 I=1,NPERG
      E2GN(I)=X(I)*SQRT(VARE2G)
302   CONTINUE
C
C     GAIN SCORES
C
      CALL RANGEN(NPERG,ISEED,X)
      DO 303 I=1,NPERG
      GAINGN(I)=X(I)*SQRT(VARGNG)+GBAR
303   CONTINUE
C
C     PRETEST AND POSTTEST SCORES
C
      DO 304 I=1,NPERG
      XX(I)=TG(I)+E1GN(I)
      YY(I)=TG(I)+GAINGN(I)+E2GN(I)
304   CONTINUE
C
C
C     GENERATION OF DATA FOR NO GAIN GROUP
C
C     TRUE SCORES
C
      CALL RANGEN(NPERG,ISEED,X)
      DO 400 I=1,NPERG
```

```
      TN(I)=X(I)*SQRT(VARTN)+PRMNNG
400   CONTINUE
C
C     PRETEST ERROR SCORES
C
      CALL RANGEN(NPERG,ISEED,X)
      DO 401 I=1,NPERG
      E1NG(I)=X(I)*SQRT(VARE1N)
401   CONTINUE
C
C     POSTTEST ERROR SCORES
C
      CALL RANGEN(NPERG,ISEED,X)
      DO 402 I=1,NPERG
      E2NG(I)=X(I)*SQRT(VARE2N)
402   CONTINUE
C
C     GAIN SCORES
C
      CALL RANGEN(NPERG,ISEED,X)
      DO 403 I=1,NPERG
      GAINNG(I)=X(I)*SQRT(VARGNN)
403   CONTINUE
C
C     PRETEST AND POSTTEST SCORES
C
      DO 404 I=1,NPERG
      L=NPERG+I
      XX(L)=TN(I)+E1NG(I)
      YY(L)=TN(I)+GAINNG(I)+E2NG(I)
404   CONTINUE
C
C
C     STANDARD ANALYSIS OF COVARIANCE COMPUTATIONS
C
      N=2*NPERG
C
C     INITIALIZATION
C
      SUMXG=0.0
      SUMX2G=0.0
      SUMYG=0.0
      SUMY2G=0.0
      SUMXYG=0.0
      SUMXN=0.0
      SUMX2N=0.0
      SUMYN=0.0
      SUMY2N=0.0
      SUMXYN=0.0
C
C     GROUP SUMS AND SUMS OF SQUARES
```

```
C
      DO 600 I=1,NPERG
C
      SUMXG=SUMXG+XX(I)
      SUMX2G=SUMX2G+XX(I)*XX(I)
      SUMYG=SUMYG+YY(I)
      SUMY2G=SUMY2G+YY(I)*YY(I)
      SUMXYG=SUMXYG+XX(I)*YY(I)
      K=NPERG+I
      SUMXN=SUMXN+XX(K)
      SUMX2N=SUMX2N+XX(K)*XX(K)
      SUMYN=SUMYN+YY(K)
      SUMY2N=SUMY2N+YY(K)*YY(K)
      SUMXYN=SUMXYN+XX(K)*YY(K)
600   CONTINUE
C
C     TOTAL SUMS AND SUMS OF SQUARES
C
      TSUMX=SUMXG+SUMXN
      TSUMX2=SUMX2G+SUMX2N
      TSUMY=SUMYG+SUMYN
      TSUMY2=SUMY2G+SUMY2N
      TSUMXY=SUMXYG+SUMXYN
C
C     COMPUTE TOTAL SUMS OF SQUARES
C
      CFX=TSUMX*TSUMX/N
      CFY=TSUMY*TSUMY/N
      CFXY=TSUMX*TSUMY/N
      TXX=TSUMX2-CFX
      TYY=TSUMY2-CFY
      TXY=TSUMXY-CFXY
C
C     COMPUTE BETWEEN GROUPS SUMS OF SQUARES
C
      BXX=(SUMXG*SUMXG+SUMXN*SUMXN)/NPERG-CFX
      BYY=(SUMYG*SUMYG+SUMYN*SUMYN)/NPERG-CFY
      BXY=(SUMXG*SUMYG+SUMXN*SUMYN)/NPERG-CFXY
C
C     COMPUTE ERROR SUMS OF SQUARES
C
      EXX=TXX-BXX
      EYY=TYY-BYY
      EXY=TXY-BXY
C
C     COMPUTE ADJUSTED SUMS OF SQUARES
C
      TYYADJ=TYY-TXY*TXY/TXX
      EYYADJ=EYY-EXY*EXY/EXX
      BYYADJ=BYY-(TXY*TXY)/TXX+(EXY*EXY)/EXX
C
```

```
C     COMPUTE ADJUSTED MEAN SQUARES
C
       MSB=BYYADJ/1.0
       MSE=EYYADJ/(N-3)
C
C     COMPUTE F STATISTIC
C
       F=MSB/MSE
C
C
C      ANALYSIS OF COVARIANCE WITH PORTER'S ADJUSTMENT
C
C     COMPUTE CORRELATION BETWEEN X AND Y
C
       DENOM=TXX*TYY
       RXY=TXY/SQRT(DENOM)
C
C     COMPUTE SUMS OF SQUARES WITH PORTER'S ADJUSTMENTS
C
       EPORT=EYYADJ
       TPORT=TYY-((RXY*EXY+BXY)**2)/((RXY*RXY*EXX)+BXX)
       BPORT=TPORT-EPORT
C
C     COMPUTE MEAN SQUARES
C
       MSBP=BPORT/1.0
       MSEP=EPORT/(N-3)
C
C     COMPUTE F STATISTIC WITH PORTER'S ADJUSTMENT
C
       FPORT=MSBP/MSEP
C
       IF (F.GT.4.45) NF10=NF10+1
       IF (F.GT.3.89) NF100=NF100+1
       IF (FPORT.GT.4.45) NFP10=NFP10+1
       IF (FPORT.GT.3.89) NFP100=NFP100+1
C     WRITE OUT RESULTS
C
       WRITE (6,800) NSAMP,F,FPORT
800    FORMAT (1X,T50,I4,T64,F8.3,T77,F8.3)
C
       IF (NSAMP.LT.NSPL) GO TO 106
C
       WRITE (6,801)
801    FORMAT ('0',T10,'NF10',T20,'NF100',T30,'NFP10',T40,'NFP100'
      *)
       WRITE (6,802) NF10,NF100,NFP10,NFP100
802    FORMAT (1X,T10,4(I4,6X))
       IF (NGROUP.LT.NGP) GO TO 100
       STOP
       END
```

```
C
C
      SUBRCUTINE RANCEN(M,IR,X)
      CIMENSICN X(M)
      I=1
    1 CALL RANCU(IR,JR,R1)
      IR=JR
      CALL RANCU(IR,JR,R2)
      IR=JR
      R1=2.0*(R1-.5)
      R2=2.C*(R2-.5)
      S2=R1*R1+R2*R2
      IF(S2.GT.1.0) GO TO 1
      Y=SQRT(-2.0*(ALOG(S2)/S2))
      X(I)=R1*Y
      IF(I.EQ.M) GO TO 2
      X(I+1)=R2*Y
      IF(I+1.EQ.M) GO TO 2
      I=I+2
      GC TC 1
    2 RETURN
      END
C
      SUBRCUTINE RANCU(IX,IY,YFL)
      IY=IX*65539
      IF(IY)5,6,6
    5 IY=IY+2147483647+1
    6 YFL=IY
      YFL=YFL*.4656613E-9
      RETURN
      END
```

BIBLIOGRAPHY

Box, G. E. P. and Muller, Mervin E. "A Note on the Generation of Random Normal Deviates," Annals of Mathematical Statistics, 29 (June, 1958), 610-611.

Campbell, Donald T. and Erlebacher, Albert. "How Regression Artifacts in Quasi-Experimental Evaluations can Mistakenly Make Compensatory Education Look Harmful," in Disadvantaged Child, Vol. 3, Ed. Hellmuth, Jerome. New York: Brunner/Marzel, Inc., 1970.

Campbell, Donald T. and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally and Company, 1963.

Cochran, William G. "Analysis of Covariance: Its Nature and Uses," Biometrics, 13 (September, 1957), 261-281.

Cronbach, Lee J. Essentials of Psychological Testing, Third Edition. New York: Harper and Row, 1970.

Cronbach, Lee J. and Furby, Lita. "How We Should Measure 'Change' or Should We?," Psychological Bulletin, 74 (January, 1970), 68-80.

DeCecco, John P. The Psychology of Learning and Instruction: Educational Psychology. Englewood Cliffs, N. J.: Prentice Hall, Inc., 1968.

Ferguson, George A. Statistical Analysis in Psychology and Education, Third Edition. New York: McGraw-Hill Book Company, 1971.

Fisher, Ronald A. Statistical Methods for Research Workers, Fourth Edition. London: Oliver and Boyd Ltd., 1932.

Fisher, Ronald A. Statistical Methods for Research Workers, Tenth Edition. London: Oliver and Boyd Ltd., 1946.

Fisher, Ronald A. The Design of Experiments. London: Oliver and Boyd Ltd., 1935.

Garside, R. F. "The Regression of Gains Upon Initial Scores," Psychometrika, 21 (March, 1956), 67-77.

Glass, Gene V., Peckham, Percy D., and Sanders, James R. "Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analysis of Variance and Covariance," Review of Educational Research, 42 (Summer, 1972), 237-288.

Gulliksen, Harold. Theory of Mental Tests. New York: John Wiley and Sons, Inc., 1950.

Helmstadter, G. C. Principles of Psychological Measurement. New York: Appleton-Century-Crofts, 1964.

Hicks, Charles R. "The Analysis of Covariance," Industrial Quality Control, 20 (December, 1965), 282-287.

Hilgard, Ernest R. Theories of Learning. New York: Appleton-Century-Crofts, Inc., 1956.

Hill, Winfred F. Learning: A Survey of Psychological Interpretations, Revised Edition. Scranton: Chandler Publishing Company, 1971.

Karmel, P. H. and Polasek, M. Applied Statistics for Economists, Third Edition. Bath, Great Britain: Pitman Publishing, 1970.

Kirk, Roger E. Experimental Design: Procedures for the Behavioral Sciences. Belmont California: Wadsworth Publishing Company, Inc., 1968.

Lord, Frederic M. "A Paradox in the Interpretation of Group Comparisons," Psychological Bulletin, 68 (1967), 304-305.

Lord, Frederic M. "Elementary Models for Measuring Change," in Problems in Measuring Change. Ed. Chester W. Harris. Madison, Wisconsin: The University of Wisconsin Press, 1963.

Lord, Frederic M. "Further Problems in the Measurement of Growth," Educational and Psychological Measurement, XVIII (1958), 437-451.

Lord, Frederic M. "Large-Sample Covariance Analysis When the Control Variable is Fallible," Journal of the American Statistical Association, 55 (1960), 307-321.

Lord, Frederic M. "Statistical Adjustments When Comparing Preexisting Groups," Psychological Bulletin, 72 (1969), 336-337.

Lord, Frederic M. "Statistical Inferences about True Scores," Psychometrika, 24 (March, 1959), 1-17.

Lord, Frederic M. "The Measurement of Growth," Educational and Psychological Measurement, XVI (1956), 421-437.

Lord, Frederic M. and Novick, Melvin R. Statistical Theories of Mental Test Scores. Reading, Massachusetts: Addison-Wesley Publishing Company, 1968.

Manning, Winton H. and DuBois, Philip H. "Correlational Methods in Research on Human Learning," Perceptional and Motor Skills, 15 (1962), 287-321.

Marks, Edmond and Martin, Charles G. "Further Comments Relating to the Measurement of Change," American Educational Research Journal, 10 (Summer, 1973), 179-191.

Marsaglia, G. and Bray, T. A. "A Convenient Method for Generating Normal Variables," SIAM Review, 6 (July, 1964), 260-264.

McNemar, Quinn. "On Growth Measurement," Educational and Psychological Measurement, XVIII (1958), 47-55.

Muller, Mervin E. "A Comparison of Methods for Generating Normal Deviates on Digital Computers," Association for Computing Machinery Journal, 6 (1959), 376-383.

Neel, John Howard. A Comparative Analysis of Some Measures of Change. Dissertation, University of Florida (1970).

O'Connor, Edward F., Jr. "Extending Classical Test Theory to the Measurement of Change," Review of Educational Research, 42 (Winter, 1972), 73-97.

Peckham, Percy D. "The Robustness of the Analysis of Covariance to Heterogeneous Regression Slopes," Paper read at the Annual Meeting of the American Educational Research Association, Minneapolis, Minnesota, (March 5, 1970).

Porter, Andrew Colvin. The Effects of Using Fallible Variables in the Analysis of Covariance. Dissertation, University of Wisconsin, University Microfilms, Ann Arbor, Michigan: (1967).

Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, 1956.

Snedecor, George W. and Cochran, William G. Statistical Methods, Sixth Edition. Ames, Iowa: The Iowa State University Press, 1967.

Thorndike, R. L. "Regression Fallacies in the Matched Groups Experiment," Psychometrika, 7 (June, 1942), 85-102.

Werts, Charles E. and Linn, Robert L. "A General Linear Model for Studying Growth," Psychological Bulletin, 73 (1970), 17-22.

Winer, B. J. Statistical Principles in Experimental Design, Second Edition. New York: McGraw-Hill Book Company, 1971.

BIOGRAPHICAL SKETCH

James Edwin McLean was born January 29, 1945, at Greensboro, North Carolina. He grew up in Orlando, Florida and graduated from Edgewater High School in June, 1963. He graduated from Orlando Junior College in January, 1966, and entered the United States Marine Corps Reserve.

In December, 1968, he received the degree, Bachelor of Science, with a major in mathematics education from the University of Florida. In January, 1969, he enrolled in the Department of Statistics at the University and received the degree, Master of Statistics, in June, 1971. During this period, he worked as a graduate assistant in that department where he taught elementary statistics and probability.

He accepted a teaching assistantship in the College of Education at the University of Florida in September, 1971. He currently holds that position part-time along with the position of research associate for a Project Follow Through evaluation grant.

James Edwin McLean is a member of the American Educational Research Association, National Council on Measurement in Education, the American Statistical Association, and Phi Delta Kappa.

He is married to the former Sharon Elizabeth Robb and they have no children.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William B. Ware

William B. Ware, Chairman
Associate Professor of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Vynce A. Hines

Vynce A. Hines
Professor of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

W. Mendenhall

William Mendenhall
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

P. V. Rao

P. V. Rao
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James T. McClave

James T. McClave
Assistant Professor of Statistics

This dissertation was submitted to the Dean of the College of Education and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1974

B. J. Shaef for M C Baker
Dean, College of Education

Dean, Graduate School