A COMPARATIVE ANALYSIS OF SOME MEASURES OF CHANGE

By

JOHN HOWARD NEEL

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

1970

ACKNOWLEDGEMENTS

The writer wishes to thank his committee members for their guidance in this study, especially Dr. Charles M. Bridges, Jr., who suggested the topic and was a constant source of assistance throughout the study. When this study was begun there was a questioning of change scores in the department which was helpful and encouraging. Those responsible for this atmosphere were Dr. Charles M. Bridges, Jr., Dr. Robert S. Soar, Dr. William B. Ware and Mr. Keith Brown.

Dr. Vynce A. Hines and Dr. P. V. Rao assisted by each detecting an error in the model presented.

Dr. William B. Ware was especially helpful editorially, as was Dr. Wilson H. Guertin.

The writer's wife, Carol, was encouraging, and understanding of the time which was necessarily 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

A COMPARATIVE ANALYSIS
OF SOME MEASURES OF CHANGE

by

John Howard Neel

August, 1970

Chairman: Wilson H. Guertin
Co-Chairman: Charles M. Bridges, Jr.
Major Department: Foundations of Education

The purpose of this study was to determine which of four selected methods was most appropriate for measuring change such as gain in achievement. The four selected methods were raw difference, Lord's true gain, regressed gain, and analysis of covariance procedures.

In order to compare the four methods Monte Carlo techniques were employed to generate samples of pre and post scores for two groups. The reliability, variances, and means of the sampled populations were controlled. One hundred samples were generated at each of 20 combinations of five levels of reliability, two levels of group size, and two levels of gain. Using each of the four methods, a t statistic was calculated for each sample to test the null hypothesis of no difference in amount of gain between the two groups. The number of t's significant at the 0.05 level of significance was recorded for each of the four methods.

At each of the 20 combinations a chi square test was used to test the null hypothesis of equal proportions of significant t's among the four groups. This hypothesis was rejected in each case. It was noted that use of Lord's true gain procedure tended to create a greater significance level than the user would intend. The proportion of significant t's for each of the other three methods of analysis fell reasonably close to the expected values. On this basis use of Lord's true gain procedure was not recommended and since there was no apparent difference among the remaining methods of analysis, none was recommended above the other.

CHAPTER I
INTRODUCTION

Learning has long been a primary focus of investigation for educators. A definition of learning has been offered by Hilgard (1956):

> Learning is the process by which an activity originates or is changed through reacting to an encountered situation, provided that the change in activity cannot be explained on the basis of native response tendencies, maturation, or temporary states of the organism (e. g. fatigue, drugs, etc.). (p. 3)

Although Hilgard went on to say that the definition is not perfect, it does illustrate a commonly accepted aspect of learning: learning involves change of the behavior of the organism which learns (Bigge, 1964, p. 1; Skinner, 1968, p. 10; Combs, 1959, p. 88).

Educators have been concerned with this change and often have sought to measure the change occurring in some situation. Several methods of analyzing change have been presented in the literature. A comparison of these methods of analysis of the measurement of change and the accompanying difficulties which arise was the focus of this study. The purpose of the study was to determine which of the several selected measures of change is most appropriate under various conditions.

The Problem of Measuring Change

In all sciences measurement is an approximation. In conducting a first order survey, a surveyor makes three measurements of a distance and takes the average of the three measurements. Physicists and engineers customarily report the relative size of the error of their measurements. Physical scientists have been fortunate in that the size of the relative error involved in their measurements often has been small, frequently less than 0.01 and sometimes less than 10^{-8}. Educators are unfortunate in this respect in that if a student's true I. Q. were 100 and an I. Q. score of 88 was observed, the relative error would be 0.14. The size of this error is not uncommon and larger relative errors do occur. The Stanford-Binet intelligence test has standard error of measurement equal to five I. Q. points (Anastasi, 1961, p. 200). Thus, relative errors of 0.14 or larger will occur 1.64 per cent of the time, assuming a normal distribution of errors.

In measuring change the problem is compounded since there is an error in both the pre score and the post score. Moreover, when the magnitude of the change is small or zero, the magnitude of the error may be larger than that of the change. This possibility makes the change difficult to detect or to separate from the error. The effect of this error of measurement on change scores was noted as early as 1924 by Thorndike:

> When the individuals in a varying group are measured twice in respect to any ability by an imperfect measure (that is one whose self-correlation is below 1.00), the average difference between the two obtained scores will equal the average difference between the true scores that would have been obtained by perfect measures, but for any individual the difference between the two obtained scores will be affected by the error. Individuals who are below the mean of the group will tend by the error to be less far below it in the second, and individuals who are above the mean of the group in the first measurement will tend by the error to be less far above it in the second. The lower the self-correlation, the greater the error and its effect.

Thorndike (1924) went on to show that there was a spurious negative correlation between initial true score and true gain. He then stated that "the equation connecting the relation of obtained initial ability with obtained gain, the unreliability of the measures and the true facts" had not been discovered. Lord (1956) developed the equation to which Thorndike alluded. Lord made the following assumptions concerning the error of measurement: the errors

> i) have zero mean for the groups tested.
> ii) have the same variances for both tests.
> iii) are uncorrelated with each other and with true score on either test. (Lord, 1956)

McNemar (1958) has extended Lord's work to the case of unequal error variances. It should be noted here that McNemar's follow-up of Lord's work is only an extension. When the variances are equal, McNemar's formulas are identical to Lord's (Lord, 1958).

Methods of Analyzing Change

The estimated true gain scores derived from Lord's and McNemar's equations have been used in either t tests or analysis of variance procedures (Soar, 1968; Tillman, 1969). There are, in addition to Lord's method, three other commonly used methods for analyzing change.

One method has been the use of a straight analysis of variance or t test on the raw difference scores as the situation warrants. It is important to note that the raw scores are used in the analysis with no correction for the unreliability of the measures.

A second method is to complete an analysis of covariance on the raw difference scores using the pretest scores as covariates. These procedures are standard statistical techniques and may be found in many texts (Hays, 1963; Snedecor, 1956; Winer, 1962).

The third method of measuring gain has been advocated by Manning and DuBois (1962): the method of residual gain. In this method the final scores are regressed on the initial scores and the difference between the final score and the score predicted by the regression equation is taken as a measure of gain. This measure is then used in t tests or analysis of variance procedures.

Thus in the case of equal variances four common methods of analyzing change have been identified:

1. Use of raw gain scores in appropriate procedures.

2. Use of Lord's true gain scores in appropriate procedures.
3. Use of Manning and DuBois' regressed gain scores in appropriate procedures.
4. Use of analysis of covariance on raw gain scores with pre scores as covariates.

The Problem

A researcher faced with these different methods and a problem in measuring change is confronted with a second and fundamental methodological problem: which of the methods for measuring change is most appropriate? It is this question that this study sought to answer. The problem of which method to use is further complicated since different writers have claimed that different techniques were appropriate. "Manning and DuBois and Rankin and Tracy feel that the method of residual gain is more appropriate for correlational procedures since it is metric free" (as quoted in Tillman, 1969, p. 2). Ohnmacht (1968) also suggested that this procedure was the best. Lord (in Harris, 1963, chapter 2) mentioned regressed gain but seemed to advocate his own method as being superior. This position is further supported by Cronbach and Furby (1970).

To determine which of the methods of analysis was most appropriate, an empirical study was conducted to compare the results of each method under known situations.

Some Limitations

This study was limited to the two group situation. To examine more than two groups would have involved such a large number of possibilities as to make the study impractical in terms of time and money. Thus, this study excluded the more general multigroup comparisons possible with analysis of variance and covariance procedures and was limited to examining t tests and the analysis of covariance using the pre score as covariate.

Additionally the case of unequal variances between the two groups was not considered.

A third limitation was that the variance of the true gains was selected *a priori* to be 3.6. True gain scores with this variance will be such that over 99 per cent will be within five units of the true mean gain. It is the ratio of the variances of the gain and error which is important. Since the reliabilities were varied, as indicated later, this study was conducted utilizing several such ratios.

One of the factors of interest was the reliability of the test used. A second factor of interest was the sample size or the relative power of the procedures under study. There is an infinite number of reliabilities and a decision was made as to the levels of reliability to be investigated: 0.50 to 0.90 in increments of 0.10. Tests with reliabilities lower than 0.50 are rarely used in practice and at best the resulting data would be highly questionable. The lower limit of 0.50 was chosen for this reason.

Sample sizes of 25 and 100 per group were chosen as being somewhat representative of sample sizes used in educational research.

Procedures

Two groups were compared under 20 conditions using t tests as follows. There were five levels of reliability (0.50, 0.60, 0.70, 0.80, 0.90) and two levels of sample size (25 and 100) used in this study. Thus there were ten different combinations of sample sizes and reliabilities. For each of these combinations two cases were investigated, one where there was no pre to post test gain in either group and the second where there was a known gain from the pre to the post test for one group. For each of these 20 instances, 100 samples were generated and analyzed using each of the four methods of analysis indicated previously.

Consequently, two questions were to be answered:

1. Does any one of the selected methods yield a disproportionate number of significant t values when there is no difference between the mean gain of the two groups?
2. Is any one of the selected methods more powerful, i. e. more successful in detecting a difference when a difference does exist?

Samples from a normal distribution were generated using techniques described by Rosenthal (1966). The method for generating random numbers was the multiplication by a constant method. With the procedures used, this method

will produce 8.5 million numbers before the series repeats. This number was more than sufficient for this study. All generation of the samples and calculation of t values using the various methods of analysis were done on the IBM 360/65 computer at the University of Florida. The significance level used for the t tests was 0.05.

The two research questions generated two null hypotheses:

1. The proportion of t's significant at the 0.05 level is the same for each method of analysis when there is no gain in either group.
2. The proportion of t's significant at the 0.05 level is the same for each method of analysis when there is gain by one group but not by the other.

These hypotheses were tested at each of two combinations of reliability and sample size with chi square tests using the 0.05 level of significance.

Significance of the Study

The results of this study should either indicate empirically that one or more methods were superior to the others or that there were no great differences among the methods. If the former were true, then educational researchers may select one of the better methods. If the latter were true, then educational researchers may select any of the

methods. In either case the study provides some answer as to how change scores should be analyzed.

Organization of the Study

Chapter I has been the introduction, statement of the problem, limitations, hypotheses, and procedural overview. Chapter II reviews related literature, essentially the development of the equation and methodology of the various techniques studied. Chapter III describes the procedures and Chapter IV presents the data, conclusions, and summary.

CHAPTER II
RELATED LITERATURE

Much has been said in the literature about measuring change. However, most of this discussion is centered around the four methods investigated in this study: raw gain, Lord's true gain, regressed gain, and analysis of covariance procedures. As pointed out in Chapter I raw gain and regressed gain procedures are discussed in many texts and therefore not discussed here. Lord's true gain and regressed gain procedures are discussed in this chapter.

Derivation of Lord's True Gain Scores

The following derivation parallels Lord's (1956) development of true gain scores with one exception as noted. Lord gave the following equations as a model for the observed pre and post scores:

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

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

where

X = observed pre score;

Y = observed post score;

T = true pre score;

G = true gain;

E_1 = error of measurement in pre observed score;

E_2 = error of measurement in post observed score.

Lord then made the following assumption concerning E_1 and E_2, the errors of measurement.

The Errors

i) have zero mean in the group tested.
ii) have the same variance (σ_e^2) for both tests.
iii) are uncorrelated with each other and with true score on either test (Lord, 1956).

The derivation can be considerably shortened at this point by examining a standard regression equation which predicts one variable, X_1, from two other variables, X_2, X_3. The equation is (Tate, 1965, p. 171)

$$(3) \quad X_1 = B_{12.3} \frac{\sigma_1}{\sigma_2} (X_2 - \bar{X}_2) + B_{13.2} \frac{\sigma_1}{\sigma_3} (X_3 - \bar{X}_3) + \bar{X}_1$$

where

$$B_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2} \quad ;$$

and

$$B_{13.2} = \frac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2} \quad .$$

If we let

X_1 = G = gain;

X_2 = X = observed pre score;

X_3 = Y = observed post score;

the following elements in the regression equation can be identified as

X_1 = estimated gain;

$\bar{X}_2$ = mean of the observed pre scores;

$\bar{X}_3$ = mean of the observed post scores;

S_1 = standard deviation of the gain scores;

S_2 = standard deviation of observed pre scores;

S_3 = standard deviation of observed post scores;

r_{23} = correlation of observed pre and post scores.

Lord has pointed out that r_{12} and r_{13}, the correlations between observed score and true score, are the reliabilities.

From (1) and his stated assumptions, Lord writes

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2 \quad ; \tag{4}$$

$$\sigma_y^2 = \sigma_t^2 + \sigma_g^2 + \sigma_e^2 \quad ; \tag{5}$$

$$\sigma_{xy} = \sigma_t^2 + \sigma_{te} \quad . \tag{6}$$

Lord solves these equations to find

$$\sigma_g^2 = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy} - 2\sigma_e^2 \quad , \tag{7}$$

the variance of the true gain scores.

At this point the only element in the regression equation which is undefined is $\bar{X}_1$. This element is found by considering the mean of the observed pre scores which from (1) can be seen to be equal to the mean of the gain scores plus the mean of the errors of measurement, that is

$$\bar{X} = \bar{T} + \bar{E} \quad . \tag{8}$$

But by Lord's assumption (i) $\bar{E} = 0$, therefore

(9) $$\bar{X} = \bar{T} \quad .$$

Similarly from (2) Lord shows

(10) $$\bar{Y} = \bar{T} + \bar{G} \quad .$$

Then (9) is subtracted from (10) and rewritten to yield

(11) $$\bar{G} = \bar{Y} - \bar{X} \quad .$$

Thus all elements are defined and (3) may be rewritten as in terms of T, G, X, and Y as

(12) $$\hat{G} = B_{12.3} \frac{\sigma_g}{\sigma_x} (X - \bar{X}) + B_{13.2} \frac{\sigma_g}{\sigma_y} (Y - \bar{Y}) + \bar{Y} - \bar{X}$$

which Lord has asserted to be an estimate of true gain.

It may be noted that no notational scheme or other method has been presented to distinguish between statistics and parameters in the preceding derivation. This lack is in keeping with Lord's derivation. It is assumed here that Lord was referring to parameters until the point at which he obtained the final equation and that he then intended to use sample values to estimate the appropriate parameters in the regression equation.

Comparison of Lord's True Gain Scores with Other Scores

In his original article Lord (1956) made no comparison of his method with any other method. In a subsequent article (1959) Lord again made no mention of other methods. In a chapter written in Problems in Measuring Change (Harris, 1963, Chapter 2) he made reference to regressed gain scores,

but no discussion or comparison was presented. In Statistical Theories of Mental Test Scores (Lord and Novick, 1968) no comparison of Lord's true gain scores with other procedures is presented.

Comparison of Regressed Gain Scores with Other Scores

Manning and DuBois (1962) have compared per cent, raw, and residual gain scores. A per cent gain score is raw gain score divided by the pre score (Manning and DuBois, 1962). The comparisons were made on the bases of metric requirements, reliability, and appropriateness of use in correlation procedures. On each of these bases residual gain scores were recommended over per cent and raw gain scores. Manning and DuBois pointed out that per cent and raw gain scores require at least equal interval scales on both pre and post scores and that the scales be the same on both pre and post scores, i. e. the same equal interval scale must be used on both tests. According to Manning and DuBois these qualities are not possessed by educational and psychological test scores. In contrast, residual gain scores do not require the same equal interval scales and therefore are appropriate for use with test scores (Manning and DuBois, 1962). Manning and DuBois summarily list formulas showing that residual gain scores are more reliable and more appropriate measures for correlational procedures than are raw or per cent gain scores. These formulas were only listed, not derived, and no reference was made to their derivation.

Another Study and Summary

Madansky (1959) reported or derived several methods for fitting straight lines to two variables when both were measured with error. One of these procedures is applicable in the case when the variance of the error of measurement is unknown. However, there has apparently been no attempt to apply the method to the analysis of change.

A search of the literature has revealed no comparative empirical examination of the four methods examined in this study. Further, the advocates and authors of two of the reported procedures, each of whom has been shown to know of the existence of the other procedure, continue to advocate their own method even though they offer no reason or data for this advocacy. This study should provide some knowledge as to any difference in the four methods.

CHAPTER III
METHODS AND PROCEDURES

Procedures: An Overview

As stated in Chapter I, Monte Carlo techniques were employed to generate pre and post test scores for two groups. One group is referred to as the gain group, the other as the no gain group.

The model for the observed pre scores is

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

where

X is the observed pre score;

T is the true pre score;

E_1 is a normally distributed random error with mean 0.0 and variance $\sigma^2_{e_1}$.

The model for the post scores is

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

where

Y is the observed post score;

T is the true pre score;

G is the true gain from pre to post score;

E_2 is a normally distributed random error with mean 0.0 and variance $\sigma^2_{e_2}$.

The generated scores were subsequently analyzed for the difference in the amount of gain or change between the two groups. The scores were analyzed by the four selected methods;

1. a t test on the raw difference scores
2. a t test on Lord's true gain scores
3. a t test on regressed gain scores
4. a t test from an analysis of covariance on the raw difference scores using pre score as covariate.

The results of these analyses were then compared. For the gain group an appropriate mean gain, μ_g, from pre to post scores was obviously selected to be 0.0 in the case of no gain for either group and selected to be of such size as to make the power 0.50 when there was a gain in the gain group. The post scores were generated by adding a random normal gain, G, and a random normal error, E_2, to the generated true pre scores. The variables G and E_2 had means μ_g and 0 respectively and variances as discussed later. For the no gain group there was no gain from pre to post scores.

The pre scores for both the gain and the no gain group were taken from a normal population with mean 50.0 and variance 100.0. The mean and variance of the population of post scores for both the gain and the no gain groups were functions of the mean true gain and of the reliability.

After the samples were generated, the hypothesis of no difference in average gain between the two groups was tested using each of the four methods. The t values for each of these tests were recorded.

This procedure was repeated over 100 samples for each of the selected reliabilities 0.50, 0.60, 0.70, 0.80 and 0.90. The method for introducing the effect of the selected reliability into each generated score is presented in a following section.

Thus 100 t values were calculated and recorded for each method of analysis and at each level of reliability. This entire procedure was repeated for each of the following conditions:

1. group size = 25, $\mu_g = 0.0$ for both groups;
2. group size = 25, $\mu_g \neq 0.0$ for the gain group
   $= 0.0$ for the no gain group;
3. group size = 100, $\mu_g = 0.0$ for both groups;
4. group size = 100, $\mu_g \neq 0.0$ for the gain group
   $= 0.0$ for the no gain group.

Where the gain was not equal to 0.0 it was such that the power of the t tests on the raw difference scores was 0.50, i. e. the expected proportion of rejected null hypotheses was 0.50. The following sections describe in more detail some of the previously mentioned procedures.*

---

* The reader is also referred to the FORTRAN listing in Appendix A for the exact computer routines by which these procedures were carried out.

Sampling from a Normal Population with Specified Mean and Variance

If F is the cumulative density function of a random variable R_1, then the random variable R_2 defined by

$$R_2 = F(R_1) \tag{3}$$

is uniformly distributed over the interval [0,1] (Meyer, 1965, p. 256, Theorem 13.6). Here F is the cumulative density function of the random variable R_1. It follows then that R_1, where

$$R_1 = F^{-1}(R_2) \ , \tag{4}$$

is normally distributed if F^{-1} is the inverse cumulative density function of a normal distribution and if R_2 is a uniform random number on the interval [0,1] (Meyer, 1965, pp. 256-257).

Thus random samples from a normal distribution may be obtained using uniform random numbers and by (4) where F is the cumulative density function of a normal distribution. For a normal distribution, $F^{-1}(R_2)$ must be calculated using numerical approximation methods. This calculation as well as the generation of the uniform random numbers were done using a routine described by Rosenthal (1966, pp. 270, 287). Rosenthal's techniques were adapted to the IBM 360/65 computer installed at the University of Florida (see Appendix A FUNCTION RAND). The normal population sampled had mean 0 and variance 1.0. If a different mean or variance

was required, it was obtained by addition or multiplication by an appropriate constant.

Selecting Reliability

Reliabilities of 0.50, 0.60, 0.70, 0.80 and 0.90 were selected as representative of reliabilities found in test scores. The reliability, rel, of a test may be defined as

$$(5) \qquad \text{rel} = 1 - \frac{\sigma^2_{e_1}}{\sigma^2_x} \qquad \text{(Nunnally, 1967, p. 221)},$$

where σ^2_e is the error of measurement variance and σ^2_x is the observed score variance. Since σ^2_x had been selected *a priori* to be 100.0, we have from (5)

$$(6) \qquad \sigma^2_{e_1} = 100\ (1 - \text{rel}) \quad .$$

Moreover, since

$$(7) \qquad X = T + E_1$$

and since the error, E_1, is assumed to be independent of the true score, T, we have

$$(8) \qquad \sigma^2_x = \sigma^2_t + \sigma^2_{e_1} \quad ,$$

or, combining (6) and (8) and solving for σ^2_t,

$$(9) \qquad \sigma^2_t = 100 - \sigma^2_{e_1} \quad .$$

For the post scores the desired variances are also easily found from the model for a post score,

(10) $$Y = T + G + E_2 \quad ,$$

and for which

(11) $$\sigma_y^2 = \sigma_t^2 + \sigma_g^2 + \sigma_{e_2}^2 \quad .$$

As stated in Chapter I, σ_g^2 was selected to be 3.6. If (5) is rewritten with σ_y^2 and $\sigma_{e_2}^2$ instead of σ_x^2 and $\sigma_{e_1}^2$, respectively, then (5) and (11) may be used to find

(12) $$\sigma_{e_2}^2 = \frac{\sigma_t^2 + \sigma_g^2}{rel} - \sigma_t^2 - \sigma_g^2 \quad .$$

The effect of the selected reliability may be obtained by selecting the error of measurement variances and the variance of the true scores in accordance with (6), (9) and (12).

Thus it is seen that if true scores are selected from a distribution with variance σ_t^2 and if the errors of measurement are selected independently from a distribution with variance σ_e^2, then by (8) X has variance σ_x^2, if (7) holds.

Selecting Gain

When there was no gain in either group the value of μ_g would then be 0.0. When μ_g was nonzero for the gain, its value was selected so as to make the power of the t test on the raw difference scores equal to 0.50. The power of 0.50 was selected in order to permit maximum difference between the four methods of analysis.

The value of G was determined by examining the difference scores (D).

(13) $$D = Y - X,$$

and from

(14) $$D = (T + G + E_2) - (T + E_1) \quad ,$$

or

(15) $$D = G + E_2 - E_1 \quad .$$

The elements in the right side of (11) are mutually independent normally distributed random variables whose variances have been found and thus

(16) $$\sigma_d^2 = \sigma_{e_1}^2 + \sigma_g^2 + \sigma_{e_2}^2 \quad .$$

Furthermore since the only difference in (15) for the gain and no gain groups is the mean of G, the variance of the difference for the gain group, $\sigma_{d_g}^2$, and the variance of the difference for the no gain group, $\sigma_{d_{ng}}^2$, are equal, i. e.

(17) $$\sigma_{d_g}^2 = \sigma_{d_{ng}}^2 = \sigma_d^2 \quad .$$

This common value σ_d^2 may then be used to determine the appropriate value of μ_g to produce the desired power of 0.50 for the t test on the raw difference scores.

The t test on the raw difference scores is found from the following formula:

$$(18) \qquad t = \frac{\bar{D}_g - \bar{D}_{ng}}{\sqrt{\left[\frac{(n-1)\,S^2_{d_g} + (n-1)\,S^2_{d_{ng}}}{n_1 + n_2 - 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

If the group size is 25 and reliability 0.50, the value of $_g$ is found as follows:

(note: rel = 0.50 implies $\sigma^2_d = 106.72$)

$$(19) \qquad t = \frac{\bar{D}_g - 0}{\sqrt{\left[\frac{24(2\sigma^2_{d_g}) + 24(2\sigma^2_{d_{ng}})}{25 + 25 - 2}\right]\left[\frac{1}{25} + \frac{1}{25}\right]}}$$

$$t = \frac{\bar{D}_g}{2.9154} \quad .$$

This value of t is greater than the critical value of t (2.01) only if

$$(20) \qquad 2.01 \leq \frac{\bar{D}_g}{2.9154} \quad ,$$

that is, only if

$$(21) \qquad 5.86 \leq \bar{D}_g \quad .$$

Thus if a value of 5.86 is chosen for the mean gain, the power is .50. Appropriate values for other group sizes and reliabilities were similarly determined.

Analysis of the t Values for the Four Methods

The number of t's significant at the 0.05 level was recorded for each of the four methods of analysis. These

data were recorded for each of the 20 combinations of sample size, reliability and gain. A chi square statistic was calculated for each of these 20 sets to test the null hypothesis of no difference in the proportion of significant t values for the four methods of analysis. These data may be seen in Tables 1 and 2 of Chapter IV.

CHAPTER IV
RESULTS, CONCLUSIONS, AND SUMMARY

Results

The number of significant t's for each method of analysis under the no gain condition is presented in Table 1, and for the gain condition in Table 2. Additionally, the computed chi square statistics for each reliability level are given. In each case the null hypothesis tested was that the proportion of significant t's was the same for each of the four methods of analysis. The chi square values were computed from the 2 x 4 contingency tables implied by the corresponding line of the table. For example, for group size of 25 and a reliability of 0.50, the 2 x 4 contingency table implied by the first line of Table 1 is:

	Raw gain	Lord's gain	Regressed gain	Analysis of covariance
Significant	5	58	2	2
Non significant	95	42	98	98

As may be seen by inspection of Tables 1 and 2, all the chi square values were significant at the 0.05 level and in each case the hypothesis of equal proportion of significant t values for the four methods of analysis was rejected.

TABLE 1

NUMBER OF SIGNIFICANT t's WHEN THE TRUE MEAN GAIN WAS 0.0 FOR BOTH GROUPS

GROUP SIZE = 25

RELIABILITY	RAW GAIN	LORD'S TRUE GAIN	REGRESSED GAIN	ANALYSIS OF COVARIANCE	CHI SQUARE
0.50	5	58	2	2	163.13
0.60	5	53	4	4	128.98
0.70	7	51	6	6	103.69
0.80	5	24	5	5	30.76
0.90	6	27	4	4	40.95

GROUP SIZE = 100

RELIABILITY	RAW GAIN	LORD'S TRUE GAIN	REGRESSED GAIN	ANALYSIS OF COVARIANCE	CHI SQUARE
0.50	6	83	5	5	242.95
0.60	5	69	6	6	178.28
0.70	6	55	7	7	115.05
0.80	5	50	5	5	111.60
0.90	8	38	6	6	59.61

CHI SQUARE (3,.95) = 7.82

TABLE 2

NUMBER OF SIGNIFICANT t's WHEN THE POWER OF THE t TEST ON THE RAW DIFFERENCE SCORES WAS 0.50

GROUP SIZE = 25

RELIABILITY	RAW GAIN	LORD'S TRUE GAIN	REGRESSED GAIN	ANALYSIS OF COVARIANCE	CHI SQUARE
0.50	44	89	54	54	48.80
0.60	52	89	64	64	33.00
0.70	52	86	65	66	28.23
0.80	47	78	50	51	25.43
0.90	50	75	53	52	16.90

GROUP SIZE = 100

RELIABILITY	RAW GAIN	LORD's TRUE GAIN	REGRESSED GAIN	ANALYSIS OF COVARIANCE	CHI SQUARE
0.50	58	95	66	66	38.80
0.60	57	94	63	63	39.48
0.70	50	82	58	58	24.45
0.80	52	88	54	54	38.37
0.90	55	83	59	59	21.35

CHI SQUARE (3,.95) = 7.82

Further inspection of Tables 1 and 2 reveals a higher number of significant t's for Lord's true gain procedure than for any other methods. Moreover, examination of Table 1 shows that this particular technique gives a considerably greater frequency of significant t values than one would expect by chance. The expected frequency is 5 for the *a priori* established condition of no actual difference in the two populations sampled. These results indicate that use of Lord's true gain procedure tends to create a higher significance level than the user would intend. If the sample proportion of significant t's found in the analysis is used as an estimate of the significance level, that estimate is 0.58 for the case when the group size was 25. For the same group size the lowest estimate of the significance level is 0.39.

Conclusions

Since the hypothesis of equal proportion of significant t's for the four methods of analysis was rejected in each of the 10 cases where the mean gain was 0.0 and since the use of Lord's true gain scores provided estimated levels of significance which were considerably higher than those intended, the use of Lord's true gain scores is strongly suspect and therefore is not recommended.

No apparent differences were found among the remaining three methods of analysis. However, there is a similarity between the regressed gain scores procedure and the

analysis of covariance procedure that should be examined. The data in Table 1 indicate that the same number of significant t's was found by both of these methods in the case where there was no gain for either group.

The 100 t values for each of the four methods of analysis when the group size was 25 and there was no gain in either group are presented in Appendix B. Inspection of the t values for the regressed gain procedure and the analysis of covariance procedure reveals a striking similarity between the t values; for each sample the t values are identical to at least the first decimal place. As a descriptive statistic it is noted that the correlation between the t values found by these two methods is 1.00 (rounded to 3 digits). Thus, the two methods are providing very similar results.

In contrast the correlation between the t's for the raw difference and regressed gain procedures is 0.898. The two methods, regressed gain and analysis of covariance, are not entirely similar to the raw difference procedure. It may also be seen from Appendix B that the signs of the t's from both the regressed gain and analysis of covariance procedures sometimes are opposite from the sign of the t for the raw difference procedure.

Snedecor (1956, pp. 397,398) has indicated that the regressed gain procedure and the analysis of covariance procedure on the post scores using pre score as covariate are identical procedures. No source was found indicating a

similarity between the regressed gain procedure and the analysis of covariance procedure on the difference scores using pre score as covariate. However, the two methods may be shown to be equivalent by writing the linear model for the regressed gain, or, equivalently, for the analysis of covariance on the post scores using pre score as covariate, and the model for the analysis of covariance on the difference scores using pre score as covariate. The model for covariance analysis on the post scores is

(1) $$Y = B_0 + B_1X + B_2Z + E$$ (Mendenhall, 1968, p. 170),

where X and Y are defined as previously and

$Z = 1$, if Y is from the gain group,

$= 0$, if Y is not from the gain group.

E = a normally distributed random error with mean 0.

The model for covariance analysis on the difference scores is

(2) $$D = B_0 + B_1X + B_2Z + E$$

where

(3) $$D = Y - X$$

and all other elements are defined as in (1). Now if the right side of (3) is substituted into (2) and the resultant equation rearranged to yield

(4) $$Y = B_0 + (B_1 + 1.0)X + B_2Z + E \quad ,$$

it is seen that (1) and (4) are identical except for the addition of 1.0 to B_1 of equation (1) and thus the two methods will yield the same t values for testing the hypothesis that B_2 is equal to 0.0.

Since no clear difference was found among the raw gain procedure, the regressed gain procedure and the analysis of covariance procedure, none of these is recommended as more appropriate for the analysis of change than the other. All of these three procedures are recommended above Lord's true gain procedure.

Discussion

It is reasonable to ask if there is some questionable logic in Lord's derivation of true gain scores. Two things become apparent upon examination of the derivation. First the formula which Lord uses to begin his derivation, (3) of Chapter II, requires that the independent variable be known exactly (Madansky, 1959; Scheffé, 1959, p. 4), i. e. without error of measurement. The problem of estimating true gain arises in that the pre and post test scores are not known exactly, but instead the observed scores, or the true scores plus measurement errors, are known as may be seen from Lord's models of the observed scores, (1) and (2) in Chapter II. If true pre and true post score were known these could be put into the regression equation. However, if true pre score and true post score were known there would be no need for the regression equation to

estimate gain. The gain could be obtained simply by subtracting true pre score from true post score. In short, Lord seems to have assumed his conclusion in his derivation.

Second, a look at the basic method for estimating true gain is enlightening. In order to estimate true gain from observed score it is necessary to somehow remove the error of measurement since, assuming the test to be valid, this is the factor that obscures the true score. Mendenhall (1968, p. 1) says that "...statistics is a theory of information...." What information is known concerning the errors of measurement? By Lord's assumption (iii) of Chapter II the errors are uncorrelated with true score and with each other. Thus neither the observed pre score nor the observed post score should provide any information concerning the size of the error. Since this error is random it would seem that it could not be removed from the observed scores. Consider two equal observed scores, one obtained from a higher true score by the addition of a negative error, the other obtained from a lower true score by the addition of an error of the same size but opposite sign of the previous error. How does one decide which score is to have a positive correction added and which is to have a negative correction added? It would appear that one cannot make this decision without having some information besides the observed scores.

A Direction for Future Research

Since this study shows no difference among the proportion of significant t's for the raw gain, regressed gain, and the analysis of covariance procedures, it would be interesting to investigate the use of these procedures under assumptions other than the models listed in the first section of Chapter III. Differences may occur, for example, when the gain is a linear function of the pre score.

Summary

This study compared four selected measures of change. The four measures were: raw difference, Lord's true gain, regressed gain, and analysis of covariance procedures. An empirical comparison was made among these four methods. Samples were generated using Monte Carlo techniques and the data in each sample were analyzed by each of the four methods.

It was found that Lord's true gain procedure produced a number of spurious significant t values, greater than would be expected by chance, when there was no real difference in amount of gain between the two populations sampled. No apparent differences were noted among the remaining three methods and these three methods did not appear to have inflated significance levels. With such data use of Lord's true gain procedure is not recommended and none of the remaining three methods was recommended over the others.

APPENDIX A

FORTRAN PROGRAM WHICH PERFORMED THE CALCULATIONS

```
      DIMENSION X(100,2,2),LORD(100,2),DIFF(100,2)
     1,AMAT(3,3),DIFFX(3),DUM(3)
      DOUBLE PRECISION SEED
      REAL KR21,LORD,MPRG,MPRNG,MPOG,MPONG,MLG,MLNG,MRGG,MRGNG
      READ (5,1) N,SEED,NSAMP,KSAMP,NOPT
    1 FORMAT(I3,F11.0,3I4)
      IF(NOPT.EQ.0) GO TO 3
      READ(5,2) IREL,ISAMP
C     IREL - RESTART RELIABILITY AT IREL FOR ABORTED RUN
C     ISAMP - RESTART SAMPLE NUMBER AT ISAMP FOR ABORTED RUN
    2 FORMAT(2I4)
      GO TO 4
    3 IREL=1
      ISAMP=1
    4 CONTINUE
C
C     N - SIZE OF SAMPLE
C     GAIN - AVERAGE GAIN IN GAIN GROUP
C     SEED - SEED FOR RANDOM NUMBER GENERATOR
C     KSAMP - NUMBER OF SAMPLES TO BE TAKEN AT EACH LEVEL
C     NSAMP - NUMBER OF THE LAST SAMPLE FROM THE PREVIOUS RUN
C           - INCREMENTED AND PRINTED OUT AS THE SAMPLE NUMBER
```

```
C     OF RELIABILITY
C
      TT=0.0
      DO 1000 IR=IREL,5
      READ(5,14) GAIN
   14 FORMAT(F6.4)
      DO 902 J1=ISAMP,KSAMP
      NSAMP=NSAMP+1
C
C     S - SUM
C     SS - SUM OF SQUARES
C     D - DIFFERENCE SCORE
C     G - GAIN GROUP
C     NG - NO GAIN GROUP
C     PR - PRE SCORE
C     PO - POST SCORE
C     PP - PRE X POST
C     PRDIFF - SUM PR X DIFF
C
      SDG   =0.0
      SDNG  =0.0
      SSDG  =0.0
      SSDNG =0.0
      SPRG  =0.0
      SPRNG =0.0
      SSPRG =0.0
```

```
      SSPRNG=0.0
      SPOG  =0.0
      SPONG =0.0
      SSPOG =0.0
      SSPONG=0.0
      SSPPG =0.0
      SSPPNG=0.0
      PRDIFF=0.0
      REL=0.40+ 0.10*IR
      SE1= SQRT(100.0-SE1*SE1)
      SX=SQRT(100.0-SE1*SE1)
      SE2= SQRT((SE1*SE1+3.36)/REL-3.36-SE1*SE1)
      GAIN=TT*SQRT(2.0*(SE1*SE1+SE2*SE2+3.36)/N)
C     X(I,J,K)    I=STUDENT
C                 J=1, PRE SCORE
C                  =2, POST SCORE
C                 K=1 GAIN
C                  =2 NO GAIN
C
C     DIFF(I,J)    I= STUDENT
C                  J=1, GAIN
C                   =2, NO GAIN
C
      DO 10 I=1,N
      D1=SX*RAND(SEED)
      D2=SX*RAND(SEED)
```

```
      X(I,1,1)=50+D1+SE1*RAND(SEED)
      X(I,2,1)=50+D1+GAIN+1.83*RAND(SEED)+SE2*RAND(SEED)
      X(I,1,2)=50+D2+SE1*RAND(SEED)
      X(I,2,2)=50+D2+1.83*RAND(SEED)+SE2*RAND(SEED)
      DIFF(I,1)=X(I,2,1)-X(I,1,1)
      DIFF(I,2)=X(I,2,2)-X(I,1,2)
      SDG=SDG +DIFF(I,1)
      SDNG=SDNG +DIFF(I,2)
      SSDG=SSDG+DIFF(I,1)*DIFF(I,1)
      SSDNG=SSDNG+DIFF(I,2)*DIFF(I,2)
      SPRG=SPRG +X(I,1,1)
      SPRNG=SPRNG +X(I,1,2)
      SSPRG=SSPRG +X(I,1,1)*X(I,1,1)
      SSPRNG=SSPRNG+X(I,1,2)*X(I,1,2)
      SPOG=SPOG+X(I,2,1)
      SPONG=SPONG+X(I,2,2)
      SSPOG=SSPOG+X(I,2,1)*X(I,2,1)
      SSPONG=SSPONG+X(I,2,2)*X(I,2,2)
      SSPPG=SSPPG+ X(I,1,1)*X(I,2,1)
      SSPPNG=SSPPNG +X(I,1,2)*X(I,2,2)
   10 PRDIFF=PRDIFF+DIFF(I,1)*X(I,1,1)+DIFF(I,2)*X(I,1,2)
      VAPRG= (SSPRG-SPRG*SPRG/N)/(N-1)
      VAPRNG= (SSPRNG-SPRNG*SPRNG/N)/(N-1)
      VAPOG= (SSPOG-SPOG*SPOG/N)/(N-1)
      VAPONG= (SSPONG-SPONG*SPONG/N)/(N-1)
      CPPG= (SSPPG-SPRG*SPOG/N)/SQRT((SSPRG-SPRG*SPRG/N)*(SSPOG-
```

```
1SPOG*SPOG/N))
 CPPNG= (SSPPNG-SPRNG*SPONG/N)/SQRT((SSPRNG-SPRNG*SPRNG/N)*(
1SSPONG-SPONG*SPONG/N))
 DBARG= SDG/N
 DBARNG = SDNG/N
 VADG= (SSDG-SDG*SDG/N)/(N-1)
 VADNG =(SSDNG-SDNG*SDNG/N)/(N-1)
 T1= (DBARG-DBARNG)/SQRT(((N-1)*(VADG+VADNG)/(2*N-2))*(2.0/N
1))
 B1G= (((1.0-REL)*CPPG*SQRT(VAPOG))/SQRT(VAPRG)-REL+CPPG*CPP
1G)/
1(1.0-CPPG*CPPG)
 B2G= (REL-CPPG*CPPG-((1.0-REL)*SQRT(VAPRG)*CPPG)/SQRT(VAPOG
1))/(1.0-CPPG*CPPG)
 B1NG=(((1.0-REL)*CPPNG*SQRT(VAPONG))/SQRT(VAPRNG)-REL+CPPNG
1*CPPNG))/ (1.0-CPPNG*CPPNG)
 B2NG= (REL-CPPNG*CPPNG-((1.0-REL)*SQRT(VAPRNG)*CPPNG)/SQRT(
1VAPONG))/(1.0-CPPNG*CPPNG)
 SLG   =0.0
 SSLG  =0.0
 SLNG  =0.0
 SSLNG =0.0
 MPRG=SPRG/N
 MPOG=SPOG/N
 MPRNG=SPRNG/N
 MPONG=SPONG/N
```

```
      DO 110 I=1,N
      LORD(I,1)=DBARG+B1G*(X(I,1,1)-MPRG)+B2G*(X(I,2,1)-MPOG)
      LORD(I,2)= DBARNG+B1NG*(X(I,1,2)-MPRNG)+B2NG*(X(I,2,2)-MPON
     1G)
      SLG=SLG+LORD(I,1)
      SLNG=SLNG+LORD(I,2)
      SSLG=SSLG+LORD(I,1)*LORD(I,1)
  110 SSLNG=SSLNG+LORD(I,2)*LORD(I,2)
      MLG=SLG/N
      MLNG=SLNG/N
      VALG=(SSLG-SLG*SLG/N)/(N-1.0)
      VALNG=(SSLNG-SLNG*SLNG/N)/(N-1.0)
      T2=(MLG-MLNG)/SQRT(((SSLG-SLG*SLG/N)+SSLNG-SLNG*SLNG/N)/(N*
     1(N-1)))
      A=(SSPPG+SSPPNG-(SPRG+SPRNG)*(SPOG+SPONG)/(2*N))/
     1(SSPRG+SSPRNG-(SPRG+SPRNG)*(SPRG+SPRNG)/(2*N))
      B=(SPOG+SPONG)/(2*N)-A*(SPRG+SPRNG)/(2*N)
      SRGG=0.0
      SSRGG=0.0
      SRGNG=0.0
      SSRGNG=0.0
      DO 210 I=1,N
      RGSG=X(I,2,1)-A*X(I,1,1)-B
      RGSNG=X(I,2,2)-A*X(I,1,2)-B
      SRGG=SRGG+RGSG
      SSRGG=SSRGG+RGSG*RGSG
```

```
      SRGNG=SRGNG+RGSNG
  210 SSRGNG=SSRGNG+RGSNG*RGSNG
      MRGG=SRGG/N
      MRGNG=SRGNG/N
      VARGG=(SSRGG-SRGG*SRGG/N)/(N-1)
      VARGNG=(SSRGNG-SRGNG*SRGNG/N)/(N-1)
      T3= (MRGG-MRGNG)/SQRT((SSRGG-SRGG*SRGG/N+SSRGNG-SRGNG*
     1SRGNG/N)/(N*(N-1)))
      AMAT(1,1)=2*N
      AMAT(1,2)=N
      AMAT(1,3)=SPRG+SPRNG
      AMAT(2,1)=AMAT(1,2)
      AMAT(2,2)=N
      AMAT(2,3)=SPRG
      AMAT(3,1)=AMAT(1,3)
      AMAT(3,2)=AMAT(2,3)
      AMAT(3,3)=SSPRG+SSPRNG
  900 CONTINUE
      CALL INV(AMAT)
      DIFFX(1)= SDG+SDNG
      DIFFX(2)=SDG
      DIFFX(3)=PRDIFF
      YXXXXY=0.0
      DO 410 I=1,3
      DUM(I)=0.0
      DO 405 J=1,3
```

```
405 DUM(I)=DUM(I)+DIFFX(J)*AMAT(I,J)
410 YXXXXY=YXXXXY+DUM(I)*DIFFX(I)
    SSE=SSDG+SSDNG-YXXXXY
    VAACOV=SSE/(2.0*N-3)
    T4=DUM(2)/SQRT(VAACOV*AMAT(2,2))
    AA=(PRDIFF-(SPRG+SPRNG)*(SDG+SDNG)/(2*N))/
   1(SSPRG+SSPRNG-(SPRG+SPRNG)*(SPRG+SPRNG)/(2*N))
    AMG=DBARG-AA*(MPRG-(MPRG+MPRNG)/2)
    AMNG=DBARNG-AA*(MPRNG-(MPRG+MPRNG)/2)
    KR21=1.0-(MPRG*(100-MPRG))/(100*VAPRG)
    WRITE(6,501) NSAMP,REL,KR21,T1,T2,T3,T4,MPRG,MPRNG,MPOG,
   1MPONG,MLG,MLNG,MRGG,MRGNG,AMG,AMNG,SEED,VAPRG,VAPRNG,VAPOG,
   2VAPONG,VALG,VALNG,VARGG,VARGNG,VAACOV
501 FORMAT(I6,1X,2(F3.2,1X),1X,4(F7.3),5(4X,2F6.2)/5X,F23.11,
   116X,4(4X,2F6.1),8X,F6.2)
    WRITE(7,502)NSAMP,REL,KR21,T1,T2,T3,T4,MPRG,MPRNG,MPOG,
   1MPONG,MLG,MLNG,NSAMP,MRGG,MRGNG,AMG,AMNG,VAPRG,VAPRNG,
   2VAPOG,VAPONG,VALG,VALNG,VARGG,VARGNG,VAACOV
502 FORMAT(I6,2F3.2,4F7.3,6F6.2,3X,'1'/I6,4F6.2,8F5.1,F6.2,3X,'
   12')
902 CONTINUE
    ISAMP=1
1000 CONTINUE
    STOP
    END
    FUNCTION RAND (RO)
```

```
      DOUBLE PRECISION RO
      RO=DMOD(RO*30517578125.,34359738368.)
      X=RO/34359738368.
      Y=SIGN(1.0,X-0.5)
      V=SQRT(-2.0*ALOG(0.5*(1.0-ABS(1.0-2.0*X))))
      RAND=Y*(V-(2.515517+0.802853*V+.010328*
     1V**2)/(1.0+1.432788*V+0.189269*V**2+0.001308*V**3))
      RETURN
      END
      SUBROUTINE INV(A)
C     PROGRAM FOR FINDING THE INVERSE OF A 3X3 MATRIX
      DIMENSION A(3,3),L(3),M(3)
      DATA N/3/
C     SEARCH FOR LARGEST ELEMENT
      DO80 K=1,N
      L(K)=K
      M(K)=K
      BIGA=A(K,K)
      DO20 I=K,N
      DO20 J=K,N
      IF(ABS (BIGA)-ABS (A(I,J))) 10,20,20
   10 BIGA=A(I,J)
      L(K)=I
      M(K)=J
   20 CONTINUE
C     INTERCHANGE ROWS
```

```
      J=L(K)
      IF(L(K)-K) 35,35,25
   25 DO30 I=1,N
      HOLD=-A(K,I)
      A(K,I)=A(J,I)                                                    C
   30 A(J,I)=HOLD                                                      C
C     INTERCHANGE COLUMNS
   35 I=M(K)
      IF(M(K)-K) 45,45,3                                               C
   37 DO40 J=1,N
      HOLD=-A(J,K)
      A(J,K)=A(J,I)
   40 A(J,I)=HOLD
C     DIVIDE COLUMN BY MINUS PIVOT
   45 DO55 I=1,N
   46 IF(I-K)50,55,50
   50 A(I,K)=A(I,K)/(-A(K,K))
   55 CONTINUE
C     REDUCE MATRIX
      DO65 I=1,N
      DO65 J=1,N
   56 IF(I-K) 57,65,57
   57 IF(J-K) 60,65,60
   60 A(I,J)=A(I,K)*A(K,J)+A(I,J)
   65 CONTINUE
C     DIVIDE ROW BY PIVOT
```

```
      DO75 J=1,N
   68 IF(J-K)70,75,70
   70 A(K,J)=A(K,J)/A(K,K)
   75 CONTINUE
C     CONTINUED PRODUCT OF PIVOTS
C     REPLACE PIVOT BY RECIPROCAL
      A(K,K)=1.0/A(K,K)
   80 CONTINUE
C     FINAL ROW AND COLUMN INTERCHANGE
      K=N
  100 K=(K-1)
      IF(K) 150,150,103
  103 I=L(K)
      IF(I-K) 120,120,105
  105 DO110 J=1,N
      HOLD=A(J,K)
      A(J,K)=-A(J,I)
  110 A(J,I)=HOLD
  120 J=M(K)
      IF(J-K) 100,100,125
  125 DO130 I=1,N
      HOLD=A(K,I)
      A(K,I)=-A(J,I)
  130 A(J,I)=HOLD
      GO TO 100
  150 RETURN
```

END

APPENDIX B

LISTING OF t's FOR EACH METHOD OF ANALYSIS
WITH THE GROUP SIZE 25 AND GAIN 0.0

SAMPLE NUMBER	RAW GAIN	LORD'S TRUE GAIN	REGRESSED GAIN	ANALYSIS OF COVARIANCE
1	0.334	0.935	0.597	0.592
2	-0.052	-0.175	-0.257	-0.254
3	1.428	3.180	1.340	1.330
4	0.888	3.837	1.441	1.453
5	0.318	0.925	-0.153	-0.152
6	0.005	0.013	0.903	0.906
7	2.297	25.767	0.899	0.986
8	0.288	1.746	0.252	0.249
9	-0.904	-2.464	-1.383	-1.370
10	-1.008	-1.691	-1.013	-1.003
11	-0.186	-0.664	-0.243	-0.241
12	-0.051	-0.040	-0.255	-0.253
13	-0.232	-1.325	-0.161	-0.159
14	-0.585	-2.182	-0.959	-0.958
15	2.239	10.504	1.702	1.720
16	-0.575	-2.038	0.143	0.152
17	-1.297	-2.736	-0.717	-0.725
18	-0.538	-2.794	-0.269	-0.267
19	0.674	4.600	0.638	0.632
20	-0.954	-5.083	-1.758	-1.767
21	0.235	0.937	0.379	0.375
22	-0.036	-0.075	-0.867	-0.864
23	0.451	2.305	0.564	0.558
24	-1.324	-2.824	-1.661	-1.648
25	-0.856	-2.220	-0.518	-0.516
26	1.240	4.498	0.735	0.739
27	0.221	0.910	-0.112	-0.112
28	0.006	0.019	-0.163	-0.161
29	0.165	0.609	0.250	0.247
30	0.076	0.251	0.164	0.163
31	-0.610	-2.692	-0.920	-0.912
32	1.242	2.416	1.441	1.427
33	-0.696	-2.768	-0.608	-0.603
34	0.410	2.427	0.225	0.224
35	-0.076	-0.379	-0.254	-0.252
36	-0.670	-2.169	-0.894	-0.884
37	2.034	6.352	1.791	1.790
38	1.216	7.274	0.966	0.961
39	-0.029	-0.143	0.813	0.829
40	-1.205	-3.964	-1.347	-1.334
41	0.361	0.995	0.729	0.722

LISTING OF t's (CONTINUED)

SAMPLE NUMBER	RAW GAIN	LORD'S TRUE GAIN	REGRESSED GAIN	ANALYSIS OF COVARIANCE
42	1.042	1.762	0.975	0.966
43	1.556	16.696	1.728	1.710
44	0.511	3.038	0.184	0.183
45	-1.295	-7.504	-1.783	-1.780
46	0.669	4.977	1.026	1.026
47	0.382	1.210	0.640	0.634
48	0.228	0.216	-0.110	-0.110
49	0.922	5.810	0.943	0.933
50	-1.263	-3.021	-1.114	-1.106
51	-0.276	-0.763	-0.941	-0.938
52	0.496	1.975	0.757	0.749
53	1.385	6.511	1.583	1.572
54	-0.087	-0.187	-0.161	-0.160
55	-1.594	-2.591	-1.646	-1.630
56	0.489	3.276	-0.238	-0.239
57	-1.450	-3.217	-1.298	-1.290
58	1.146	4.044	0.657	0.657
59	1.754	7.198	1.535	1.530
60	-0.538	-8.090	-0.255	-0.253
61	-0.484	-2.971	-0.083	-0.083
62	1.155	3.037	1.766	1.749
63	1.321	20.233	1.620	1.604
64	0.130	1.233	-0.008	-0.008
65	0.721	2.349	0.514	0.510
66	-0.556	-1.149	-0.084	-0.084
67	0.178	0.332	0.434	0.432
68	-0.146	-0.465	-0.496	-0.493
69	1.290	4.543	1.526	1.514
70	-1.904	-4.395	-1.858	-1.844
71	0.572	5.132	0.470	0.466
72	0.923	2.998	1.140	1.128
73	-0.034	-0.106	-0.004	-0.004
74	0.123	0.546	0.026	0.026
75	-0.586	-2.404	-0.920	-0.911
76	-2.622	-8.344	-2.917	-2.895
77	-0.034	-0.078	-1.368	-1.382
78	-0.259	-0.308	-0.257	-0.254
79	-0.597	-2.141	0.577	0.605
80	0.065	0.197	-0.404	-0.412
81	0.614	1.834	0.162	0.162
82	0.729	0.579	1.078	1.073
83	-0.967	-2.813	-1.069	-1.058
84	-1.275	-5.166	-1.077	-1.071
85	1.259	4.430	0.797	0.802

LISTING OF t's (CONTINUED)

SAMPLE NUMBER	RAW GAIN	LORD'S TRUE GAIN	REGRESSED GAIN	ANALYSIS OF COVARIANCE
86	-1.736	-16.635	-0.996	-1.010
87	-0.835	-4.328	-0.406	-0.408
88	-0.956	-4.608	-0.633	-0.630
89	0.930	3.109	0.924	0.915
90	0.551	1.689	0.579	0.573
91	0.121	0.420	0.167	0.165
92	1.352	9.886	1.212	1.204
93	2.337	12.778	2.149	2.145
94	0.921	2.005	0.815	0.809
95	0.599	3.848	0.442	0.438
96	0.756	1.445	1.030	1.020
97	0.155	0.659	-1.004	-1.019
98	-0.091	-0.168	-0.162	-0.160
99	-1.449	-57.894	-0.194	-0.204
100	0.125	0.638	-0.218	-0.216

BIBLIOGRAPHY

Anastasi, Anne Psychological Testing. New York: Macmillan, 1961.

Bigge, Morris L. Learning Theories for Teachers. New York: Harper and Row, 1964.

Combs, A. W., Snygg, Donald Individual Behavior. New York: Harper and Row, 1959.

Cronbach, Lee J., Furby, Lita "How Should We Measure Change--or Should We?" Psychological Bulletin, 1970, 74: 68-80.

Hays, William L. Statistics for Psychologists. New York: Holt, Rinehart, and Winston, 1963.

Hilgard, Ernest R. Theories of Learning. New York: Appleton, 1956.

Lord, F. M. "The Measurement of Growth," Educational and Psychological Measurement, 1956, 16: 421-437.

Lord, F. M. "Statistical Inferences about True Scores," Psychometrika, 1959, 24: 1-17.

Lord, F. M. "Elementary Models for Measuring Change," in C. W. Harris Problems in Measuring Change. Madison, Wisconsin: University of Wisconsin Press, 1963, pp. 21-38.

McNemar, Q. "On Growth Measurement," Educational and Psychological Measurement, 1958, 18: 47-55.

Madansky, Albert "The Fitting of Straight Lines When Both Variables Are Subject to Error," Journal of the American Statistical Association, 1959, 54: 173-205.

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

Mendenhall, William Introduction to Linear Models and the Design and Analysis of Experiments. Belmont, California: Wadsworth, 1968.

Meyer, Paul Introductory Probability and Statistical Application. Reading, Massachusetts: Addison-Wesley, 1965.

Nunnally, Jum C. Psychometric Theory. New York: McGraw-Hill, 1967.

Ohnmacht, Fred W. "Correlates of Change in Academic Achievement," Journal of Educational Measurement, 1968, 5: 41-44.

Rosenthal, Myron R. Numerical Methods in Computer Programming. Homeward, Illinois: Irwin, 1966.

Scheffé, Henry The Analysis of Variance. New York: Wiley, 1959.

Schick, George B. (ed), May, Merril M. (ed) The Psychology of Reading. Milwaukee: The National Reading Conference, Inc., 15th Yearbook, 1969, Ranking, Earl F., Jr., Dale, Lothar H. pp. 17-26.

Skinner, B. F. The Technology of Teaching. New York: Appleton, 1968.

Snedecor, George W. Statistical Methods. Ames, Iowa: Iowa State College Press, 1956.

Soar, Robert S. "Optimum Teacher-Pupil Interaction for Summer Growth," Educational Leadership Research Supplement, 1968, 26(3): 275-280.

Tate, Merle W. Statistics in Education and Psychology. New York: Macmillan, 1965.

Thorndike, E. L. "The Influence of Chance Imperfections of Measures Upon the Relation of Initial Score to Gain or Loss," Journal of Experimental Psychology, 1924, 7: 225-232.

Tillman, Chester E. Crude Gain VS. True Gain: Correlates of Gain in Reading after Remedial Tutoring. Doctoral dissertation, University of Florida, 1969.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

BIOGRAPHICAL SKETCH

John Howard Neel was born July 27, 1944, at Waynesburg, Pennsylvania. He graduated from William R. Boone High school, Orlando, Florida, in June, 1962. In August, 1965, he received the degree Bachelor of Arts with a major in mathematics from the University of Florida. He taught algebra and general mathematics at John F. Kennedy Junior High School from September, 1965, until June, 1966. In September, 1966, he enrolled in the College of Education at the University of Florida under a United States Office of Education fellowship program directed by Dr. Wilson H. Guertin. In September, 1968, he accepted a research assistantship under the same program. In June, 1968, he received the degree Master of Arts in Education. He was an instructor in the College of Education at the University of South Florida from September, 1968, until August, 1969, and is currently on leave from that position. In September, 1969, he was appointed Interim Instructor in the College of Education at the University of Florida and he holds that position currently.

John Howard Neel is married to the former Carol Lynn Ramft. They have two daughters, Sarah Elizabeth and Lia Suzanne.

John Howard Neel is a member of the Florida Educational Research Association, The American Educational Research Association, Phi Delta Kappa, and the American Statistical Association.

This dissertation was prepared under the direction of the chairmen of the candidate's supervisory committee and has been approved by all members of that committee. It was submitted to the Dean of the College of Education and to the Graduate Council, and was approved as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1970

B Sharf by MC Baker

Dean, College of Education

Dean, Graduate School

Supervisory Committee:

Chairman

Co-Chairman

V. A. Hines

William B. Ware

R. V. Rees