1968

An evaluation of the effectiveness of moderator variables developed by three techniques

James Porter Tucker Jr.

AN EVALUATION OF THE EFFECTIVENESS
OF MODERATOR VARIABLES DEVELOPED
BY THREE TECHNIQUES

BY

JAMES PORTER TUCK, JR.

Approved:

W. A. Owens
Chairman

L. James Jerome

Henry O. Gudlling, Jr.

AN EVALUATION OF THE EFFECTIVENESS OF MODERATOR VARIABLES DEVELOPED BY THREE TECHNIQUES

BY

JAMES PORTER TUCK, JR.

A THESIS
SUBMITTED TO THE GRADUATE FACULTY
OF THE UNIVERSITY OF RICHMOND
IN CANDIDACY
FOR THE DEGREE OF
MASTER OF ARTS IN PSYCHOLOGY

AUGUST, 1968

Acknowledgments

The writer would like to express his deepest appreciation to Dr. William Leftwich for his assistance, patience, and enthusiasm throughout the duration of the study. The writer is also indebted to Dr. Henry O. Gwaltney and to Dr. L. James Tromater for their help and encouragement.

Appreciation is also extended to Mr. Peter B. Bahler, Director of the Computer Center, University of Richmond, and to his two able assistants, Mrs. Shirley Small and Mr. Jerald J. Utley, for their invaluable assistance in analysing the data of this research.

Finally, but most importantly, the writer would like to express his appreciation to his wife, Joyce, for the many hours she devoted to typing the manuscript. Particularly appreciated is the patience and understanding she demonstrated throughout the course of the study.

TABLE OF CONTENTS

TABLE OF TABLES

Table of Tables (cont.)

TABLE OF FIGURES

Chapter I

INTRODUCTION

Hull (1928) underscored the general finding that predictive validities for tests of the day usually fell within the range extending from .30 to .40. In fact validities of .50 or greater were so rare that he labeled the region beyond .50 the "region of inaccessibility." Today, forty years later, measurement specialists still view validities of .50 and above as distinct rarities. Psychologists have attributed this state of affairs to the so called criterion problem and to the nature of psychological tests. With a few exceptions there have been no concerted efforts made to solve the criterion problem. Instead investigators have resorted to the revalidation and revision of their measuring devices assuming that all is well with the criterion. Ghiselli (1963) maintained that this stagnant state of affairs has arisen primarily due to strict adherence to the classical additive prediction model.

Low predictive validities, as well as low reliability, have also been attributed to the nature of psychological tests. Tests are universally defined as samples of behavior. It follows logically from this definition that perfect or near perfect validities are unobtainable; no researcher aspires to perfect or complete sampling.

It would appear therefore that workers in the field of measurement have resigned themselves to what they apparently consider an unalterable situation. Giving lip service to the shortcomings of their techniques, they go blithefully on developing tests and test batteries which by design result in only moderate validity. It is the purpose of the present research to present an extensive treatment of a relatively new methodological approach to the aforementioned problems.

Additive regression

Basic to the classic prediction model whether bivariate or multivariate is the assumption that errors of measurement and errors of prediction are of the same magnitude for individuals within a specified group. It is recognized that on any single testing such errors vary in magnitude from one individual to another. However, it is maintained (Ghiselli, 1963) that as the number of parallel tests and criteria increase without limit, average standard errors of measurement and prediction approach the same value for all individuals. Therefore in a multiple regression equation, the regression weights are the same for all individuals. Mathematically the weights are based on the average variation of predictor and criterion scores for all individuals. In the additive regression equation, the predictors are combined in an additive fashion, the contribution of each predictor to the dependent variable being determined by its respective regression weight.

Ghiselli (1963) pointed out that the linear combination of a set of predictors having the same regression weights for all subjects in a group leads to the false assumption that the psychological structure of all the subjects is the same.

This is, of course, tantamount to saying that the error of measurement is the same for all individuals.

That is it is assumed that the variability of response to the predictors and to the criterion is the same for all individuals. Thus according to the classic model, subjects having identical predictor scores should have identical criterion scores. Since the same regression weights are used for all subjects, identical criterion scores would in fact be predicted for those having identical predictor scores. It follows that any difference between standard predictor scores and standard criterion scores is due to error of prediction. Stated in terms of standard scores, as the difference between predictor and criterion scores increases, the error of prediction is said to increase. Perfect prediction is achieved when the difference between standard predictor and criterion scores is zero. The additive model does not allow for the possibility that scores on a predictor may vary as a function of other variables. It may be that individuals having identical predictor scores obtain different criterion scores because they differ with respect to some other behavioral determinant which, to a degree, determines predictor performance. Succinctly the classic model does not allow for interaction between predictor variables and other variables, i.e., no assessment of or control for individual errors is possible.

In recent years a number of researchers have presented arguments and evidence which cast considerable doubt on the adequacy of the classic regression model. Lee (1961) presented an excellent discussion of the fallacy involved in assuming additivity. She pointed out that the function relating the criterion to a predictor is additive if the same criterion estimate is obtained by applying the function to the sum of the weighted predictor scores as is obtained by summing a series of criterion estimates obtained by applying the function separately to each weighted predictor score. If the function is not additive, then different predictor score combinations resulting in the same sum may result in different criterion scores. Lee argued that if the function is nonadditive, interactions exist in the data. In passing, the term "interaction" is used to refer to "situations in which the relation between two or more given variables is found to vary as a function of changes in the values of one or more other variables." Thus the fallacy in assuming additivity involves simply the fact that no allowance is made for the occurrence of non-additive relations, i.e., interactions. As Lee pointed out, it may be the case that interaction effects are so strong that two or more predictors which correlate zero with the criterion may show perfect correlation if considered jointly by appropriate methods.

As Ghiselli (1963) emphasized, there is a considerable body of research which shows that individual errors of measurement and prediction can be assessed, controlled, and predicted. These findings will be cited as the case for a nonadditive prediction model is developed. Basically the argument to be presented is that a nonadditive model results in a reduction of what has hithertofore been called error.

Moderated regression

Confronted with the aforementioned problems inherent in the classic model, psychologists have been working to develop a prediction model which takes into account individual errors of measurement and prediction. Toops (1948) introduced a technique by which he divided a group of individuals into "ulstriths" on the basis of what he called "statistical trait-patterns." His technique supposedly resulted in subgroups (ulstriths) homogeneous as to variability of response and criterion performance. Toops' technique was an early attempt to design a model which would permit the control and assessment of individual error. Gaylord and Carroll (1948) pointed out that a multiple regression equation optimum for an entire group may be inappropriate for subgroups included therein. As Lee (1961) was later to point out, they asserted that scores on a predictor variable may vary as a function of other variables which they called "population control variables." The term was defined as a variable used to identify subgroups having unique regression lines. The authors stated that population control was achieved by including the appropriate cross-product terms in a modified regression equation. Unfortunately no detailed description of the procedure was presented. Saunders (1954), using a somewhat similar technique, substituted for "population control variable" the term"moderator variables" and elaborated upon the technique for employing them in prediction equations.

In the usual additive model, the relation between *Y'* and *Xi* may be represented in terms of standard scores by the formula

$$y' = \bar{y} + \sum_i b_i x_i \qquad [1]$$

where *y'* is predicted *Y*; *y* is the mean of predicted *Y*; b_i is the regression weight for the predictor; and x_i is a predictor score.

As previously mentioned, the above model does not allow for interactions between predictors or interaction between predictors and other variables. Saunders (1956) presented the following model which makes allowance for such interactions in correlational data. It will be demonstrated shortly how the introduction of interaction terms permits the control and prediction of individual errors of measurement and prediction. Written in standard score form the moderated regression model takes the form

$$y' = \bar{y} + \sum_i a_i x_i + \sum_j b_j z_j + \sum_{ij} c_{ij} x_i z_j \qquad [2]$$

for the case involving one predictor and one moderator variable (z_j). It should be noted that a_i, b_j, and c_{ij} are regression weights. Excluding the last term, the equation is the usual additive multiple regression equation having two predictors. Note that z_j is treated as a predictor in the term $\sum_j b_j z_j$. It is the fourth term, $\sum_{ij} c_{ij} x_i z_j$, which characterizes the equation as a moderated regression equation. It is the interaction of $\underline{x}_i$ and $\underline{z}_j$ which permits the assessment of moderator effects. The fourth term is obtained by summing the cross-products between $\underline{x}_i$ and $\underline{z}_j$. There is apparently no limit to the number of predictors that may be used with moderated regression. For example, the case involving two predictors and one moderator would take the following form:

$$y' = y + \sum_i a_i x_i + \sum_j b_j x_j + \sum_k c_k z_k + \sum_{ik} d_{ik} x_i z_k \quad [3]$$
$$\sum_{jk} e_{jk} x_j z_k + \sum_{ij} f_{ij} x_i x_j + \sum_{ijk} g_{ijk} x_i x_j z_k$$

Note that in the above equation the seventh term, $\sum_{ij} f_{ij} x_i x_j$, represents the moderator effect that may occur as a result of the interaction between the two predictors. The last term allows for the assessment of any overall interaction that may occur between the predictors and the moderator. In a forth coming section the distinction between moderators and predictors will be discussed in detail. For purposes of discussion at this point, it suffices to say that the role of the moderator is to identify subgroups of individuals having unique regression lines. In general a subgroup is characterized by the unique manner in which its predictor scores are determined by relevant moderators.

Lykken and Rose (1963) and Rock (1965) have emphasized that the most serious limitation of the classic prediction model inheres in the assumption of homoscedasticity. This assumption is, in part, the basis for employing the same regression weights for all members of a group. Any attempt to increase the accuracy of prediction is nothing more than an attempt to decrease average errors of measurement and prediction for the entire group. As a result it has been argued (Lykken & Rose, 1963) that "the conventional prediction equation may not in general yield an optimum prediction for the valid regions of the [predictor] space having been...distorted by the attempt to make it predict equally well for the invalid regions."

The introduction of interaction terms (moderators included) allows for the control of individual errors to the extent that scores on a moderator can be shown to vary as a function of error. In other words, moderated regression permits a more accurate assessment of the hithertofore invalid regions of the predictor space. Instead of dealing solely with the average errors for the entire group, the moderated regression model takes into account individual errors while at the same time being descriptive of overall group performance.

Recognizing that empirical evidence speaks louder than supposedly logical argumentation, Ghiselli and his co-workers have conducted a series of studies which cast considerable doubt on the efficacy of the classic model and its assumptions. In four articles, Ghiselli (1956; 1960a; 1960b; 1963) demonstrated that errors of the aforementioned types varied as a function of other variables (moderators). He maintained that his results were evidence for the rejection or revision of the additive regression model. Rather than making the assumption that errors are equal for all individuals, Ghiselli showed that through the use of moderators individual errors of measurement and prediction could be predicted. Basically his technique involved the computation of difference scores between standard predictor and criterion scores. Through item analysis he developed a moderator scale which correlated with the difference or error scores. For the case involving two parallel tests, the technique has been employed to demonstrate the relation between measurement error and a specially developed moderator. In one study (Ghiselli, 1960b) it was reported that the reliability coefficients for a cross-validation group were found to increase from .82 to .97 for different subgroups identified using a moderator. In the same study increases in predictive validity from .226 for the total group to .860 for a selected subgroup were found using a moderator. Using another predictor and criterion, a similar increase from .154 (total group) to .779 for a selected subgroup was found. Thus on the basis of moderator scores Ghiselli was able to predict those individuals who were predictable and those who were unpredictable.

Reporting on three additional investigations, Ghiselli (1960a) further demonstrated the efficacy of using moderators. By extending his technique he developed a differential predictability variable which should not be confused with a predictability variable. Whereas a predictability variable distinguishes subgroups on the basis of differences between standard predictor and criterion scores, the differential predictability variable discriminates subgroups for which one of two tests is a better predictor on the basis of differences in difference scores. A more detailed discussion of Ghiselli's techniques is included in the forthcoming section on moderator development. Ghiselli found that when using one predictor ($\underline{P}_1$) alone for an entire group it correlated .17 with the criterion while a second predictor ($\underline{P}_2$) used alone correlated .51 with the criterion. By selecting out the 60% of the entire group for whom P_1 was the best predictor and the 40% for whom P_2 was the best predictor, a moderated R (R_m) of .75 was obtained. In effect, for those individuals whose scores on $\underline{P}_1$ were used, $\underline{P}_2$ was weighted zero in the prediction equation and vice versa. In the last two studies the percentages of individuals included in each subgroup were 58% & 42% and 68% & 32%. In the second study zero-order correlations of .55 and .61 were obtained and in the third .20 and .02. The moderated R's were .73 and .33 respectively. By demonstrating that subgroups could be formed on the basis of error for which there were significant increases in predictive validity, Ghiselli's data brought considerable doubt to bear on the assumption of homoscedasticity.

In a somewhat more extensive study, Ghiselli and Sanders (1967) offered further evidence which spoke against the appropriateness of the additive prediction model. The investigators studied the possibility of deriving a moderator scale which would divide a group of subjects into two subgroups of opposite heteroscedasticity. That is, for one subgroup high scores on both the dependent and independent variables would be expected to be highly related while for the same subgroup low scores on both variables would be expected to be only slightly related. In the second subgroup the opposite condition would be expected to be the case.

Ghiselli and Sanders represented the two groups graphically by use of the scatter diagram appearing in Figure 1. The diagonal line extending from the upper lefthand corner to the lower righthand corner did not appear in the original scatter diagram. It has been included in order to clarify the discussion of their findings.

Ghiselli and Sanders pointed out that in order for the assumption of homoscedasticity to hold, the average difference in standard scores for those individuals in the upper triangle should equal the average differences of those individuals in the lower triangle. In the three studies reported, the condition represented in Figure 1 was found to obtain. Since their study offered such a clear and relatively simple argument against the classic model, Ghiselli and Sanders' procedure and results will be considered in some detail. Their procedure was as follows:

1. A group of individuals was divided randomly into an experimental group and a cross-validation group

2. Within the experimental group the subjects were divided into those whose predictor and criterion scores placed them in the upper and lower triangle.

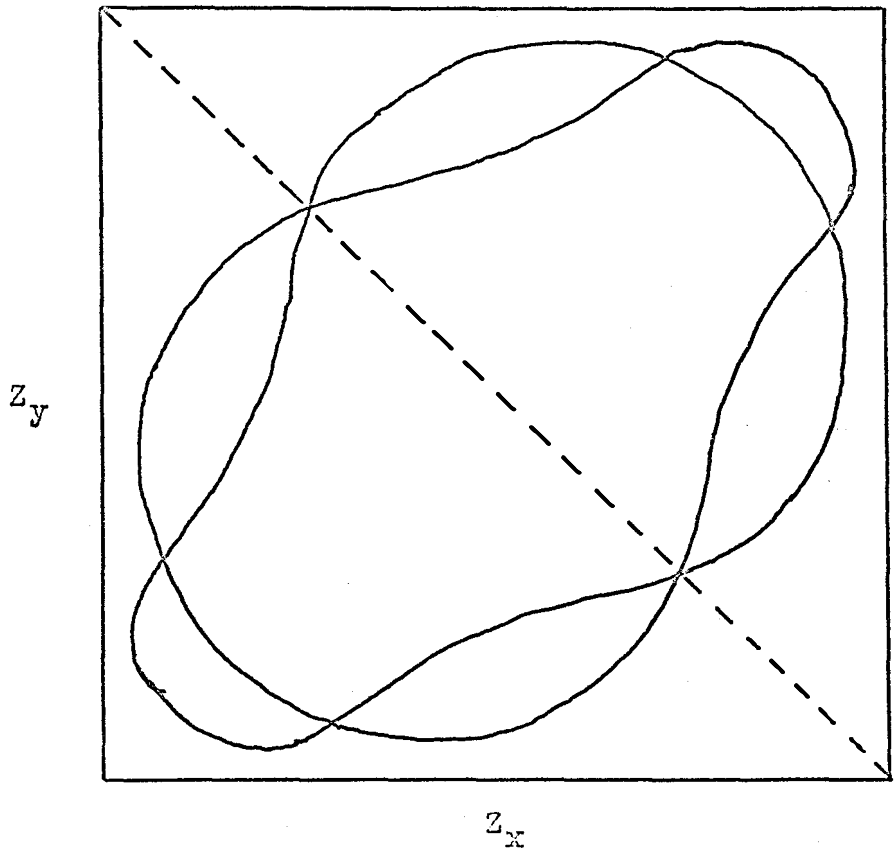

FIG. 1. Two heteroscedastic relationships pear shaped in opposite directions.

3. The absolute difference between standard predictor and criterion scores ($|z_x - z_y|$) was computed for each individual.

4. The group in each triangle was divided into two groups, one having high $|z_x - z_y|$ scores and the other having low $|z_x - z_y|$ scores.

5. Two experimental groups were formed on the basis of the four existing groups. The first group consisted of those low on $|z_x - z_y|$ in the upper triangle and those high on $|z_x - z_y|$ in the lower triangle. This group was called the upright pear subgroup. The formation of the upsidedown pear subgroup followed in the same manner.

6. Through item analysis a moderator scale was developed which differentiated the upright pear from the upsidedown pear. The moderator was so developed that those obtaining high scores on the scale formed an upright pear distribution while those scoring low formed and upsidedown pear distribution.

7. The moderator scale was then applied to a cross-validation group.

The above procedure was applied to three groups of individuals. For present purposes of discussion, the dependent and independent variables used are irrelevant. For each of the cross-validation groups the subjects were divided into two groups on the basis of those earning high or low moderator scores. Each group was then further divided into those predictor scores placed them in either the upper or lower triangle. It was hypothesized that if the moderator was effective the group scoring high on the scale would form an upright pear distribution while an upsidedown pear distribution would result for those scoring low on the scale. The following table taken from Ghiselli and Sanders' paper clearly summarizes the obtained results.

Upon close inspection of Table 1, it is readily observed that the expected results were obtained. For the upright pear group the average $|\overline{\underline{z}_x - \underline{z}_y}|$ scores for those in the lower triangle were greater than those for the individuals falling in the upper triangle. The opposite condition prevailed for the upsidedown pear group. Ghiselli and Sanders reported no significance values for their data; however, as they pointed out the results indicated that the efficacy of the additive regression model and its associated assumption of homoscedasticity are questionable.

Moderator variables

In the previous section it was mentioned that moderators and predictors differ in that a moderator divides a group of individuals into subgroups having unique regression lines. By definition and the procedures used to develop them moderators correlate with error whether it be error of measurement or error of prediction. It follows that a moderator correlates with the relationship between a predictor and a criterion (Guion, 1967). Therefore on the basis of a moderator scale a group can be divided into subgroups according to difference in error, i.e., predictability. Predictor scales, on the other hand, are not designed to correlate

TABLE 1

Average differences between standard scores ($|z_x - z_y|$) for individuals with high and low moderator scores

Group	High moderator scores (Upright pear)				Low moderator scores (Upsidedown pear)			
	Upper right-hand triangle		Lower left-hand triangle		Upper right-hand triangle		Lower left-hand triangle	
	$\lvert z_x - z_y \rvert$	N	$\lvert z_x - z_y \rvert$	N	$\lvert z_x - z_y \rvert$	N	$\lvert z_x - z_y \rvert$	N
A	.32	58	.33	19	.45	33	.33	44
B	.42	25	.70	68	.82	62	.76	29
C	1.06	74	1.15	65	1.13	52	.83	45
Average	.60		.73		.80		.67	

with error. The relationship between a predictor and a criterion is characterized by the variance that the two have in common. For the case involving two parallel tests, a moderator, by design correlates with the unexplained variation (error of measurement) between the two tests. For the case involving a predictor and a criterion, the moderator correlates with the unexplained variation between the two or, in different terms, it correlates with residual error. Using Ghiselli's terminology, it may be said that while predictors are used to predict, moderators are used to predict predictability. Grouping according to error thus results in subgroups which vary according to the accuracy with which a given predictor predicts criterion performance.

It will be recalled in the discussion of Saunders' model that a moderating effect may occur as a result of the interaction between two predictors. A moderating effect will occur if one of the predictors correlates with the unexplained variation between the other predictor and the criterion. In other words it is not absolutely necessary that scales be designed explicitly for use as moderators. Steineman (1964) has pointed out that moderators may be selected on the basis of sound theory and / or logical reasoning. This approach to the development of moderators has been called the rational approach by Banas (1964). Saunders (1954), for example, reasoned that compulsiveness should moderate the prediction of academic achievement from interest test scores. He obtained the expected results. Those subjects scoring high on the predictor but low on the criterion were found to have high compulsiveness scores. Rather than reflecting academic achievement, the high interest scores were more a reflection of compulsiveness. For the low compulsive subgroup, the correlation between interest and academic achievement was high.

Since it is the function of moderators to improve predictive validity by reducing error, some distinction between them and suppressors is in order. In a number of studies (Ewen & Kirkpatrick, 1967; Ghiselli, 1963; Ghiselli & Sanders, 1967; and Saunders, 1956) it has been emphasized that moderators like suppressors are specific to the data with which they are used. For example, two of the studies reported by Ghiselli and Sanders (1967) employed moderators developed from the same pool of items. However it was found that neither group was moderated by the same items.

Moderators and suppressors differ according to the relation each shares with the dependent variable. Suppressor variables while not correlating with the criterion do correlate with predictors which are related to the criterion. Moderators do not necessarily correlate with the criterion; however, it is not a necessary condition that the predictors with which the moderator is used must correlate with the criterion. When selecting or developing moderators it is desirable to have a scale which does not correlate highly with the criterion. If, as Tolbert (1966) pointed out, the criterion-moderator correlation is high, a moderator may serve more effectively as a predictor. Secondly, it is agreed that moderators do not have to correlate with the predictor (s). This is not a necessary condition for a moderator effect to occur at all (Saunders, 1954).

Any discussion of moderators includes many references to subgrouping. It is often falsely assumed that moderators divide total groups of subjects into separate and distinct subgroups having unique characteristics. Saunders (1956) emphasized that using moderators does not result in the formation of such groups. Rather a score on a moderator represents an individual's position on a continuous variable. Thus a moderator may be further defined as a variable which divides a group of individuals into a continuous series of subgroups.

It is important to recognize that while moderators are designed to identify subgroups, it is the subgrouping procedure which preceeds and is instrumental in the development of the moderator. Where item analysis is used, items are selected which discriminate subgroups selected on the basis of scores on some other variable. These items are used to develop the moderator. The moderator identifies subgroups to the extent that it holds up upon cross-validation.

The topic of subgrouping in the present context permits an interesting comparison of the classic and moderated regression models. The simplest representation of the classic model is offered by the following diagram

$$O \text{———} R_1 \text{———} R_c \qquad [4]$$

which can be expanded to represent the additive multiple prediction model having $\underline{R}_k$ predictors and the criterion measure $\underline{R}_c$

$$O \text{———} R_1 + R_2 + R_3 + \ldots + R_k \text{———} R_c \qquad [5]$$

Figure 2 presents Medvedeff's (1964) representation of the moderated regression model in terms of subgrouping where R_x is a response on a predictor; $\underline{O}_1 \ldots \underline{O}_n$ are subgroups of individuals homogeneous with respect to scores on a moderator; and $R_1 \ldots R_n$ are classes of criterion responses indicative of each subgroup. Medvedeff's schematization portrays quite clearly the finding that individuals having identical predictor scores may diverge as to criterion performance depending upon their scores on variables moderating the predictor-criterion relationship.

The foregoing discussion of subgrouping and moderators has been succinctly summarized in two statements made by Ghiselli (1963). He pointed out that moderators result in subgroups to the extent that they allow for the sorting of "heterogeneous aggregations of individuals into homogeneous groups..." He further reemphasized the fact that "individuals are not sorted into separate classes and a subgroup is merely those individuals who fall at the same point on the continuum [the moderator]."

No discussion of moderators would be complete without

R_x Predictor response

O_n Subgroups

R_n Criterion

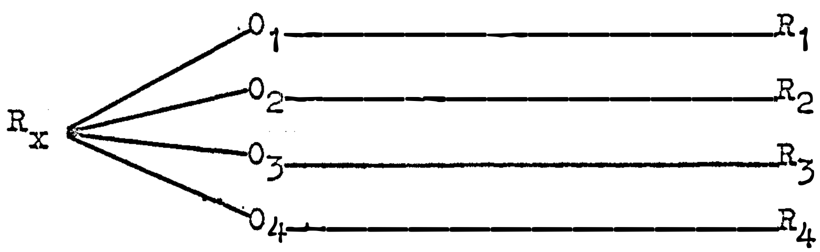

FIG. 2. Moderated regression model represented in terms of subgrouping.

a detailed examination of the manner in which they affect reliability and predictive validity. It will be recalled that the aforementioned Ghiselli studies demonstrated increases in reliability and validity as a result of using moderators. It has been pointed out that the use of moderators results in such increases because they permit the selection-out of high error subgroups. To date the best demonstration of the efficacy of subgrouping on the basis of error appeared in a study by Berdie (1961). On the basis of a variance index computed from the variability of ten subtest scores about a total test score, he divided a group of subjects into high and low variance subgroups. He further computed the difference between actual and predicted criterion scores for each individual. He hypothesized that the average prediction error for the low variance subgroup should be significantly less than for the high variance subgroup. In four of eight comparisons the expected results were obtained ($p<.05$).

While Berdie did not develop a moderator he could have easily done so. Through item analysis a scale could have been developed to correlate with the variance index. Using such a scale, reliability and validity could have been maximized for low variance subgroups.

It is often mistakenly concluded that high error (unpredictable) subgroups identified by moderators are excluded from further consideration in prediction studies. Ghiselli (1963) has underscored the fact that the role of a moderator with respect to predictive validity is not to exclude certain individuals but rather to determine or "predict the weight a test carries in determining criterion performance." Therefore the moderator may be keyed so that the higher the scores, the greater the weight a test carries for an individual. For those individuals having low moderator scores, the prediction could be made that the predictor-criterion relation would be low. Action could then be initiated to find or to develop tests which would predict for the low moderator groups. In this context the moderator is important for it allows for the prediction of unpredictability.

Moderator effects

It appears to be the consensus that a moderating effect usually will not be detected if sample size is not sufficiently large. Thorndike (1963) asserted that there is no way of determining how large a sample should be in order to obtain reliable effects. He did point out that sample size should be considerably greater than that required to establish a linear relationship. Using Thorndike's illustration, if the relationship between the criterion and a predictor takes form A for certain values of a moderator (Z) and form B for other values of Z, then the sample size must be sufficiently large to verify the form A relationship for one set of values and to verify the form B relationship for another set of values. Furthermore the sample size should be large enough to permit a determination of cut-off scores on the moderator below or above which reliable relationships between the dependent and independent variables occur.

Assuming that a study has been adequately designed and the sample size sufficiently large, we next turn to a consideration of the techniques and test statistics used to determine whether or not moderating effects exist in a set of data. Only those techniques which may be applied in general to moderated regression designs will be considered herein. Tests which are specific to certain designs will be considered in the section concerning moderator development.

Differences in subgroup standard deviations (SD) of predictor and /or criterion scores may reveal moderating effects. (Banas & Nash, 1966). Differences in SD, if reliable, actually indicate varying predictability. Those groups showing the greatest variation in scores would be the least predictable (recall Berdie's results).

A second indication that moderator effects exist may be found when a set of data is cast in the form of a scatter diagram. Kipnis (1962) found that if a moderator is operative it may show up as a lack of linearity between the predictor and criterion variables. This of course is just another way of saying that moderators, if effective, correlate with the correlation between the predictor and criterion thus producing a curvilinear relationship (Guion, 1967). For a somewhat more complex technique, see Rimland's (1960) discussion of multidimensional scatterplotting.

If the weight of a cross-product term in a moderated regression equation significantly departs from zero, evidence for interaction and thus moderating effects exist (Lee, 1961). Rock (1965) has maintained that the moderated regression model and the analysis of variance model are analogous. He pointed out that if "the function relating predictors to a criterion is nonadditive, then interactions exist in the data." The interaction referred to herein involves the nonadditive relationship between predictor and moderator variables. Concerning the similarity of the moderated regression and analysis of variance models, Marks (1964) has maintained that "every significant interaction term in an ANOVA is a...flag waved by a moderator in a plea for attention."

Saunders (1956) presented the following modified $\underline{t}$ test for assessing the difference between moderated multiple $\underline{R}$ (R_m) and linear multiple $\underline{R}$ ($\underline{R}_1$) having n - 3 degrees of freedom (degrees of freedom equal to $\underline{n}$ minus the number of independent variables in the larger $\underline{R}$).

$$t = \sqrt{\frac{(n-3)(R_m^2 - R_1^2)}{(1 - R_1^2)}} \qquad \boxed{6}$$

Finally Ewen and Kirkpatrick (1967) presented a discussion of a technique for determining if increases in predictive validity are due to moderators or to suppressors. Their technique simply involved adding a moderator to a prediction equation as a predictor. If significant increases in $\underline{R}_1$ resulted, it was assumed that the variable was acting as a suppressor. If increases resulted only as a result of using the variable in an interaction term, then it was adjudged to be acting as a moderator. This test is meaningful only if moderators are used which correlate low with the dependent variable.

Development techniques

One question of crucial importance which has largely been neglected concerns whether moderators are developed or discovered. Many investigators imply that moderators can be invented for use with almost any data. To date only one investigator, Guion (1967), has addressed himself to the question. He maintained that moderators are to be found only where they exist and in whatever manner they operate. He further stated that moderators cannot "be invented to fit an investigators methodological preference."

It would appear that the controversy has arisen primarily due to the fact that moderators and moderator effects have been considered one in the same thing. Moderator variables are developed or invented in order to allow for the detection of moderator effects if they exist. Keeping this in mind, our attention now turns to a detailed consideration of the six major techniques for developing moderators.

Absolute-difference Technique

The development of this technique was presented in a series of articles by Ghiselli (1956; 1960a; 1960b; 1963). For the case involving one predictor, refer to Ghiselli (1956). The technique for the case involving two predictors follows in outline form.

1. For subjects in an experimental group convert their

predictor and criterion scores to standard scores, $\underline{z}_p$ and $\underline{z}_c$ respectively.

2. For each subject compute $\underline{z}_{p1} - \underline{z}_c$ and $\underline{z}_{p2} - \underline{z}_c$ and call these differences $\underline{d}_1$ and $\underline{d}_2$ respectively. Algebraic signs are disregarded.

3. For each subject find $\underline{d}_2 - \underline{d}_1$ and call this difference D. Retain algebraic signs. A positive $\underline{D}$ indicates that test 1 is a better predictor than test 2. A negative D indicates that test 2 is a better predictor than test 1.

4. Select or develop through item analysis a test which correlates highly with D and call this test a differential predictability test.(moderator scale). Determine the cut-off score above which one test is to be used and below which the other is to be used.

5. For those subjects in a cross-validation group scoring high on the moderator scale, use their scores on test 1. For those scoring low, use their scores on test 2.

6. Compute $\underline{r}$ using those test scores selected by the foregoing procedure. Thus in computing $\underline{r}$ standard scores on test 1 are used for some subjects while scores on test 2 are used for the others.

Ghiselli did not use moderators in multiple regression equations. He used them to identify those groups for which one of two tests was the best predictor. In his computation of $\underline{r}$ the different scores ($\underline{z}_{p1}$ and $\underline{z}_{p2}$) were equally weighted. His technique however is well suited for use with a moderated regression equation. Such an equation would take the form of equation [3] for the case involving two predictors and one moderator. Whether one adheres to the Ghiselli procedure or uses a moderated prediction equation, the obtained results should be the same.

In discussing the effectiveness of moderators developed by this procedure, Ghiselli pointed out that as the moderator cut-off score is increased the validity coefficient should first increase and then decrease as depicted in Figure 3. If the cut-off is set very low or very high then scores on one of the two tests are used for the entire group. The optimum cut-off score is that score on the moderator above which test 1 is the best predictor and below which test 2 is the better of the two predictors.

Algebraic-difference Technique

This technique, developed by Banas (1964), is very similar to the Ghiselli technique. As the name implies, this procedure takes into account the algebraic difference between $\underline{z}_p$ and $\underline{z}_c$. Banas' procedure not only results in the identification of predictable and unpredictable subgroups, it further allows for the identification of overpredicted and underpredicted subgroups. Those subjects having high predictor scores but low criterion scores ($\underline{z}_p - \underline{z}_c = +d$) are said to be overpredicted. Conversely those having low predictor scores and high criterion scores are said to be underpredicted ($\underline{z}_p - \underline{z}_c = -\underline{d}$). Thus as Hobert and Dunnette (1967) pointed out, the Banas approach is superior to the Ghiselli approach because it results in more homogeneous subgroups. Ghiselli's procedure allows for the

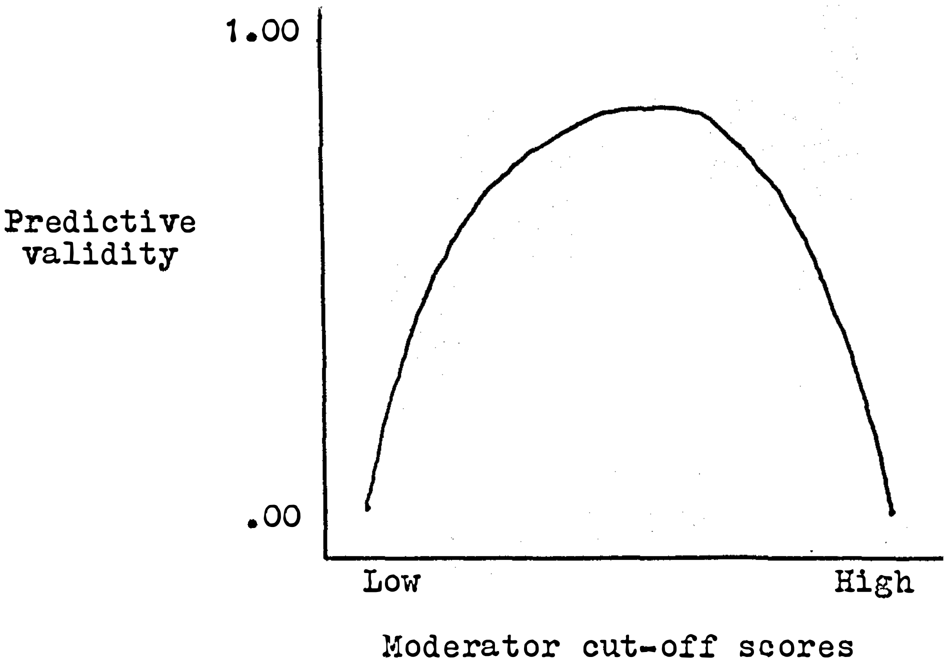

FIG. 3. The curvilinear relationship between predictive validity and cut-off scores on a differential predictability moderator.

identification of predictable and unpredictable subgroups. Banas' approach allows for the subgrouping of the unpredictable subgroup into over- and underpredicted subgroups. While Ghiselli's procedure results in the development of one moderator, Banas' results in the development of two. One moderator is developed to discriminate between the predictable and underpredicted subgroups while the second is used to discriminate between the overpredicted and predictable subgroups.

<u>Quadrant Analysis</u>

Quadrant analysis is supposedly superior to other techniques because it results in more homogeneous subgroups, (Hobert & Dunnette, 1967). The technique proceeds by dividing a group of individuals into four subgroups on the basis of standard predictor and criterion scores. The subgrouping is performed by dividing a scatter diagram into four sections by erecting lines perpendicular to the *X* and *Y* axes at the point represented by the median score on each variable. Figure 4 demonstrates what the scatter diagram should look like after subgrouping.

Through item analysis two moderators are developed. One is used to discriminate between the low hit and underpredicted subgroups while the other is used to discriminate between the high hit and overpredicted subgroups. The moderators are usually developed so that a high score on the moderator used with the low predictor groups represents underprediction. For the high predictor groups a high moderator score should be keyed to represent overprediction.

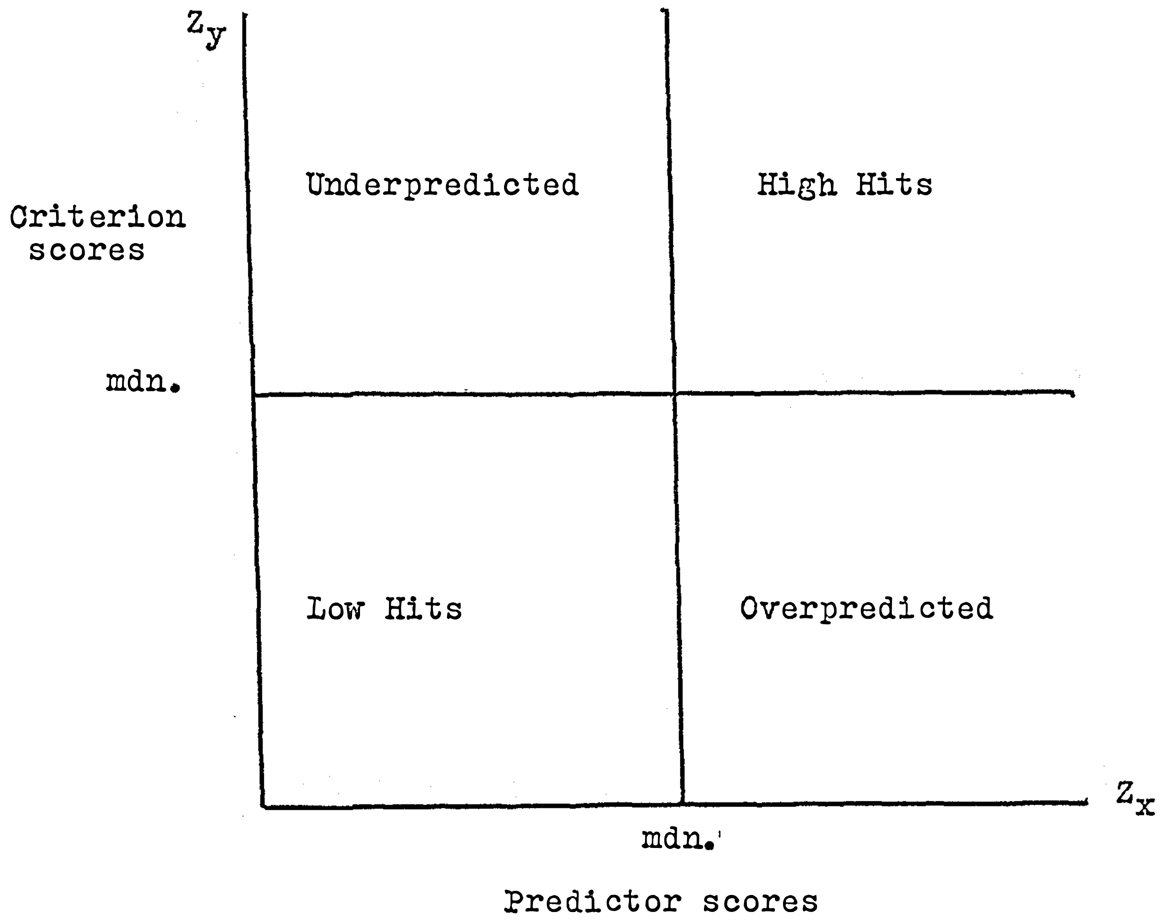

FIG. 4. Subgroups resulting from the Quadrant Analysis developmental technique.

Hobert and Dunnette pointed out that since the moderators are developed for groups having common predictor scores but different criterion scores, scores on the moderators should correlate with the criterion scores. Thus if the moderator is effective for low predictor groups it should show a positive correlation with the criterion. If the second moderator is effective it should show a negative correlation with the criterion. The moderators increase in effectiveness as the aforementioned correlations increase.

<u>Deviate Technique</u>

Introduced by Niedt and Malloy (1954) and England (1960), this technique involves correlating item responses with the difference between actual and predicted criterion scores. Through item analysis a scale is developed which correlates with residual error, i.e., Y - $\underline{Y}$'. Thus the scores on such a scale are related to the error of prediction. To the extent that the scale holds up upon cross-validation it may be used to predict the error of prediction for different subgroups. Following the usual procedure, scores on the scale would be added to a prediction equation as a moderator. Like the absolute-difference technique, this approach results in two subgroups. The unpredictable subgroup is composed of those individuals having high Y - $\underline{Y}$' scores while those having low $\underline{Y}$ - $\underline{Y}$' scores make up the predictable subgroup.

<u>Intraindividual Variability</u>

Developed by Berdie (1961), the technique proceeds as follows:

1. Compute for each individual a variance index based

on differences between subscale scores on a test and the mean or total score on the same test. For a test having ten subscales, the variance index would be of the form

$$\sum_{i=1}^{10} = \frac{\left((X_i - \overline{X}\right)^2}{n} \qquad \boxed{7}$$

2. On the basis of the variance index divide the subjects into high and low variance subgroups. Since the low variance subgroup is more consistent in responding, the error of measurement and thus the error of prediction should be smaller for this group than for the high variance subgroup.

3. Originally Berdie used the above procedure only as a means to identify predictable and unpredictable groups of individuals. The design readily lends itself to moderator development. It would be quite simple to design a moderator scale to correlate with the variance index. Unlike the aforementioned technique which resulted in a moderator correlating with the error of prediction, the present procedure results in a scale which correlates with the error of measurement. As in the absolute-difference technique and the deviate technique, the present approach only permits the identification of two subgroups.

Response Inconsistency

This technique has been used in the past to develop validity or verification scales for such tests as the SVIB (Filbeck & Callis, 1961), the MMPI (Campbell & Trockman, 1963), and the Kuder Personal Preference Record (Kuder, 1960). Briefly the technique involves item analysis of test responses to develop a scale comprised of those items rarely answered in a certain fashion by most individuals. The inclusion of scores from such a scale in a moderated prediction equation would serve to moderate the predictor-criterion relationship.

A Modified Approach

Kogan and Wallach (1964) introduced a new approach which does not actually qualify as a separate technique. In addition to considering subgroups identified by one of two moderators, they studied gains in predictive validity for subgroups identified by moderator pairs. They divided their total group into two subgroups on the basis of high and low scores on two moderators. Then subgroups high on one moderator and low on the other, high on both, low on both, and so on were studied. The procedure has the marked disadvantage that a very large number of subjects is required to effectively assess any moderating effects that may exist.

The section to follow presents a review of the studies which have used moderator designs. The review is organized according to the dependent variables used. The studies have in general reported differential validities for different subgroups identified by relevant moderators. Only limited use has been made of moderator scales in moderated regression equations.

Literature search

Grade Point Average (GPA)

Hoyt and Norman (1954) hypothesized that the correlation between freshman grades in college and aptitude test scores would be moderated by adjustment as assessed by the MMPI. It was found that the correlation between the dependent and independent variables was significantly higher for a normal subgroup as compared to a maladjusted subgroup ($p < .05$). The investigator made the observation that maladjustment affected college achievement by producing over- and under-achievement. In passing Thorndike (1963) has stated that underachievement and overachievement are synonymous to under- and overprediction respectively. Hoyt and Norman pointed out that maladjustment may have affected achievement to the extent that "one student may defensively overcompensate for felt deficiencies through intensive concentration on his studies..." while another "may dwell on his felt problems at such length that he pays no attention to his studies...." Thus on the basis of the variables used, academic achievement was more predictable for adjusted students than for maladjusted students.

In three studies (Frederiksen & Gilbert, 1960; Frederiksen & Melville, 1954; and Saunders, 1956) it was shown that compulsiveness moderated the relationship between interest test scores and engineering school grades. In all three studies the Accountant scale of the SVIB was used as a measure of compulsiveness. It was found that low compulsive subjects were more predictable on the basis of their interest scores. Frederiksen and Gilbert found that their results only held up for those keys on the SVIB which were most logically related to engineering--Mathematician, Physicist, Engineer, and Chemist scales. This of course reiterates the important point made by Guion (1967) concerning the discovery of moderator effects. Frederiksen and Gilbert further pointed out that such results are understandable because it would be expected that "compulsive students would tend to expend an amount of effort which is unrelated to interest in engineering while noncompulsive students would expend effort in relation to degree of interest."

It has been found that anxiety moderates the relation between aptitude test scores and academic achievement. Grooms and Endler (1960) found the correlation between aptitude tests and GPA for a group of male college students to be .30. When the total group was subgrouped on the basis of an anxiety measure, it was found that the aforementioned correlation for a high anxiety subgroup was .63. Coefficients of .13 and .19 were found for the medium and low anxiety groups respectively. The writers called their measure of anxiety a modifier variable instead of a moderator variable. The distinction was made because anxiety as used in the study was considered to be a trichotomized variable. Not unlike other studies using moderators which are defined as continuous variables, Grooms and Endler's measure may be assumed to have had underlying continuity.

Malnig (1959) subdivided a sample of college freshmen into high anxiety (*HA*), middle anxiety (*MA*), and low anxiety (*LA*) subgroups on the basis of scores on the *TMAS*. Unlike Grooms and Endler, he found that *HA* individuals were significantly less predictable than those individuals in the _LA_ subgroup. Such contradictory results could probably be linked to procedural differences, but more than likely they are a result of moderator specificity.

Many researchers have stated emphatically that a moderator variable is an independent continuous variable. However several studies (Abelson, 1952; Ewen & Kirkpatrick, 1967) have investigated the moderating effects of demographic variables which cannot be assumed to have underlying continuity. Abelson found that the prediction of college grades from high school grade average was more accurate for girls than for boys. Ewen and Kirkpatrick investigated the possibility of improving the prediction of success in nursing school by using race and cultural deprivation as moderators. Cultural deprivation was not found to serve as an effective moderator. The use of race, however, led to a significant improvement in validity except when success in pediatric nursing was being studied. Specifically white students were found to be more predictable than Negro students.

Hewer (1967) divided a group of 4,283 college freshmen into nine subgroups on the basis of socio-economic status. She investigated the efficiency of predicting college grades from verbal and quantitative aptitude test scores for the different subgroups. No significant moderator effects were obtained. In no case was one group significantly over- or underpredicted when compared to the other subgroups.

In the next set of articles to be discussed, the subgrouping procedure has been based on some measure of ability or aptitude. Kipnis (1962) designed a study to evaluate the efficacy of the Hand Skills Test (a device which measures "persistence beyond minimum standards on a tiring task") in predicting school grades and job performance evaluations. He divided each of four groups of Navy personnel (three enlisted man groups and one group of officer candidates) into high and low aptitude subgroups on the basis of their scores on measures of verbal aptitude, math aptitude, and mechanical aptitude. Kipnis found that aptitude did moderate the relation between the Hand Skills Test (HST) and the criterion measure. Specifically he found that the HST predicted school grades and job performance significantly better for low aptitude subgroups. He also found that for the high aptitude subgroups the validities in each case were not significantly different from zero.

Goodstein and Heilbrun (1962) reported finding evidence for differential validities for the prediction of college achievement from the EPPS at three levels of intellectual ability. It was found that for the most part personality factors were important in determining achievement for the average ability student. The success of high and low ability subgroups was found to be determined more by intellectual factors. Hakel (1966) followed Goodstein and Heilbrun's procedure using different subjects and found results which showed very little agreement with their results. Hakel maintained that results such as those obtained in the aforementioned study have very little generality. This of course re-emphasizes the general finding that moderator effects are highly specific. Hakel, like Dunnette (1963), stressed the need for cross-validation studies plus a careful study of the generality of results for correlational designs using moderators. In fact, one of the major problems in research using moderator designs has been the lack of cross-validation studies.

Bowers (1967) compared the predictive validity obtained using an additive regression model with that obtained using moderated regression model. He used high school percentile rank and an ACT composite score to predict first term GPA for a group of college freshmen employing an additive regression equation. For the moderator design, he subgrouped his subjects on the basis of ACT score levels. He found that the moderated equation permitted significantly better prediction than did the additive equation.

The studies discussed thus far have employed moderators which were selected on the basis that they were logically related to the predictor-criterion relationship. This approach has been called by Banas (1964) the rational approach to moderator development. It will be recalled that moderators may be selected, or more correctly, developed. Such moderators are developed by item analysis techniques. Banas has called this the empirical approach to moderator development.

In a study employing the deviate technique, Niedt and Malloy (1954) found that the use of two moderator keys resulted in a significant improvement in predictive effectiveness. The moderator keys were developed by correlating item responses with the unexplained variation between first semester average course marks and scores on the ACE-FORM L (linguistics) test and an English test. When applied to a cross-validation group, the addition of scores from the moderator keys lead to significantly better prediction of GPA as compared to using only scores on the two predictors.

Rock (1965) item analyzed items from a life history questionnaire to develop a predictability test (moderator scale) which would discriminate between predictable and unpredictable subgroups. Responses to the SVIB, the Purdue Math Placement Test, and the Purdue English Placement Test were scored for a group of freshmen engineering students. It was found that responses to the biographical predictability test permitted better than chance discriminations of predictable and unpredictable subgroups when considering a dichotomous criterion of survival in an engineering program. The findings suggested a curvilinear relationship between scores on the moderator and the predictor-criterion relationship. It will be recalled that the presence of such a curvilinear relationship is taken as evidence that a moderator developed by the absolute-difference technique is effective.

In a study using the absolute-difference technique, Richardson (1965) failed to find the expected results. In the first two of three studies, he attempted to predict GPA from scores on the ACE and the CPI. The predictability test or moderator was developed from items contained in the CPI. The third study involved the prediction of GPA from scores on the ACT and the MMPI with the moderator being developed from items contained in the MMPI. All three studies resulted in negative results. Richardson reported that on the basis of his data it was impossible to develop an effective moderator which could be cross-validated. He accounted for his results in terms of the nature of his criterion measure pointing out that it is difficult to predict a multidimensional variable like GPA using only two predictors.

A frequently occurring source of error, apparently operative in Richardson's study and the other studies mentioned thus far, has been discussed by Chansky (1964). He maintained that grades do not meet the assumption of normality and thus cannot be assumed to be interval level measurement. He suggested that in the future GPA be treated as ordinal level measurement with correlations being of the rank type. Such a modification is easily made. The moderated regression model lends itself quite well to use with dichotomous criterion measures. Moderated point biserial designs have frequently occurred in the literature.

Job Proficiency and Production

Lawler (1966) conducted an investigation in which he studied managerial ability as a moderator of the prediction of job performance from contingency attitudes. Contingency attitudes were measured by a questionnaire on which a group of managers expressed the degree to which they felt that their pay was contingent upon their job performance. Job performance and managerial ability measures were obtained from supervisor ratings and self-ratings. The subjects were subgrouped into those who indicated that pay was highly contingent on performance and those who indicated that pay was only slightly contingent on performance. These groups were then divided into subgroups judged to be either high or low in managerial ability. It was hypothesized that there would be no significant difference between the high and low contingency groups for the low ability managers but that there would be a significant difference therein for the high ability managers. The expected results were obtained. In other words, ability moderated the relation between contingency attitudes (assumed to be a measure of motivation) and performance. The results seemed to indicate, as Lawler pointed out, that performance = f(Ability x Motivation).

Banas and Nash (1966) found evidence for a moderating effect in their study of differential validity for groups of handicapped and non-handicapped individuals. It was shown that the prediction of job performance using the Clerical (Q), Manual Dexterity (M), Spatial (S), and Intelligence (G) scales of the GATB was significantly better for the non-handicapped individuals. Consistently lower validities were obtained for the handicapped group.

In a study involving taxicab drivers, Ghiselli (1956) sought to predict job proficiency (production during the first 12 weeks on the job) from scores on a tapping and dotting test and two inventories which assessed appropriateness of occupational level and interests in jobs involving personal relationships. He computed the difference in standard scores on the criterion and the tapping-dotting test for all the subjects in an experimental group. Then the correlations between these difference scores and the two inventories were determined. It was found that the difference scores correlated moderately with the occupational level inventory and low with the personal relationships scale. Therefore it was hypothesized that those scoring low on the occupational level scale would obtain low difference scores, i.e., relatively higher correlations between the criterion and the tapping-dotting test. In a cross-validation group, it was found that for the one third of the subjects scoring lowest on the moderator, the validity coefficient was .664; for the two thirds scoring lowest on the moderator it was .323; and for the entire group it was .220.

Dawis, Weiss, Lofquist, and Betz (1967) investigated the prediction of satisfactoriness from ability test scores using satisfaction as a moderator for a group of factory workers. Satisfactoriness was a measure of average productivity and supervisor evaluations. Employee satisfaction was assessed by a 20 scale test designed to measure satisfaction on 20 different dimensions. A battery of tests was used to assess verbal comprehension, numerical ability, visual pursuit, visual speed and accuracy, numerical reasoning, verbal reasoning, and manual speed and accuracy. The data were analyzed for each sex group. Within each sex group the subjects were subdivided into three subgroups; high satisfaction (HS), medium satisfaction (MS), and low satisfaction (LS). The validity coefficients for the HS subgroups were .63 and .69. For the LS and MS subgroups, the coefficients ranged from .34 to .52. Therefore the results supported the hypothesis that satisfaction moderates the prediction of satisfactoriness from ability test scores. It must be pointed out, however, that Dawis' et. al. results are of limited value in that no cross-validation analysis was performed.

Using age, organizational tenure, salary position, education, group size, and level of the group in the organizational hierarchy as moderators, Friedlander (1967) investigated change in work groups due to laboratory training. It was found that groups in which there was heterogeniety of educational background, in which the leader was older or had attained a higher education made significant gains due to training when compared to eight groups not receiving training. It was also found that groups high in salary position and heterogeneous with respect to tenure also benefited significantly from training.[1]

Hobert and Dunnette (1967) used item analysis to develop two moderators which discriminated between over- and underpredicted managers against a criterion of managerial effectiveness. The moderators were developed using

---

[1]Permission for citation granted by Dr. Frank Friedlander via personal communication.

the aforementioned quadrant analysis technique. The investigators reported that the underpredicted individuals were characterized as having emotional stability in interpersonal relationships, self-confidence, dominance, and aggression. The overpredicted individuals were characterized as lacking these qualities.

Personality

Self-esteem has been found to moderate the prediction of vocational choice from a measure of self perceived abilities (Korman, 1967). It was shown that high self-esteem persons saw themselves as able to meet the ability requirements of their chosen occupations while low self-esteem individuals tended to seek out those occupations not requiring their high abilities. Also the low self-esteem individual was reported as more likely to accept situations in which he felt inadequate.

In the last study to be discussed herein, Steineman (1964) found that informativeness among 13,448 Navy enlisted men moderated the prediction of career decisions from a biographical information blank. The study was based on the assumption that the career intention questionnaire would be more valid for better informed recruits. The total sample was subdivided into high, middle, and low subgroups on the basis of scores on the Naval Knowledge Test (NKT). It was found that validity coefficients were higher for subgroups scoring high on the NKT than for the total group.

Problem

Hobert and Dunnette (1967) maintained that compared to the Absolute-difference and Algebraic-difference techniques Quadrant Analysis should permit the development of more effective moderators. They pointed out that those techniques resulting in more homogeneous subgroups should result in the development of more effective moderators, i.e., moderators which more effectively enhance prediction or which result in higher multiple $\underline{R}$'s. Accordingly, the use of moderators developed by the Algebraic-difference technique should result in a greater reduction of error (residual) when compared to using moderators developed by the Absolute-difference technique. Hobert and Dunnette presented no empirical evidence to support their claims.

The present study is an empirical investigation of the three techniques and their ability to improve the strength of relationships in correlational designs. Hopefully the study will shed some light on the mechanics involved in moderator variables and the moderator effects to which they are sensitive. Since the investigation is strictly empirical, no hypotheses concerning between technique differences are tested.

Chapter II

METHOD

Subjects. In the present research, 333 white males served as subjects. The Ss were members of a class of 352 freshmen at a small four year institution during the first semester of the school year extending from September 1966 to January 1967. Nineteen Ss were excluded from the sample because complete data on the variables to be used were lacking for them.

Criterion. The criterion measure used was first semester grade point average (GPA). The GPA index is determined by computing the ratio of quality credits to academic hours attempted.

No assessment of the reliability of the dependent variable was made for two reasons. Any attempt to determine criterion reliability for the specified sample would have necessitated the computation of the intercorrelations between six-week GPA indices and the overall semester GPA index. Since GPA is determined cumulatively from each six week period to the next, reliability coefficients would be expected to be spuriously high. A second alternative would have been to correlate first semester GPA with second semester GPA. Reliability coefficients computed in this manner would probably have been spuriously low due to range restriction in the sample.

Predictor. Verbal Scale scores of the College Entrance Examination Board Scholastic Aptitude Test (SAT) were used as the independent variable. The verbal scale of the SAT includes antonyms, sentence completion, analogies, and reading comprehension items.

Zimmerman (1965) reported that validity coefficients for the verbal scale ranging from .16 to .61 with a median of .35 have been obtained for predicting academic achievement of male liberal arts students. Bowers (1965) reported that test-retest reliabilities for 14 SAT forms administered between 1959 and 1962 consistently approached .90. Bowers further pointed out that the verbal scale has been found to predict freshman GPA in liberal arts colleges better than the math scale.

Only one independent variable was used, the reason being that the introduction of additional predictors would have made the experimental design unnecessarily cumbersome.

<u>Moderator pool.</u> The items used in the development of the moderator scales were contained in a biographical data blank composed of 45 items and a 72 item adjective check list. The 45 items deal with such topics as self-satisfaction, health information, secondary education, leadership experiences, motivation, parental education, and relationship with parents.

The form was administered to the sample early in the first semester of the school year. S was instructed to circle the letter corresponding to one of four alternatives following each item considered to be most descriptive of him. Information on the adjective check list was not used. The 45 items used have been reproduced in Appendix A.

<u>Procedure.</u> The total sample ($\underline{n}$=333) was randomly divided into an developmental group ($\underline{n}$=167) and a cross-

validation group (n=166). The placement was accomplished by selecting every other name from an alphabetized list of the *Ss'* last names. The moderator scales were developed on the developmental group and then applied to the cross-validation group.

For the developmental group the correlation (r) between GPA and the verbal scale scores was computed. In preparation for the subgrouping procedures to be used, each *S*'s score on the verbal scale and his GPA index was converted to z scores.

All item analyses were conducted using the X^2 test for two independent samples (α=.30, df=1). All tests of differences were conducted at the .05 confidence level. Correlation coefficients, whether r, multiple linear (R_l), or moderated multiple (R_m), were computed with the aid of the IBM 1620 Single and Multiple Linear Regression Analysis Program. The program is described in detail in Appendix C.

As a means of simplifying the presentation of the procedure, the developmental steps for each of the three techniques is discussed separately.

<u>Absolute-difference Technique</u>

For each *S* the absolute difference between the GPA index (z_c) and the verbal score (z_p) was computed. The resulting d scores were arranged in ascending order and the median d score computed. Two subgroups were formed on the basis of the d scores. The unpredictable subgroup (n=83) was composed of those having d scores above the median. The predictable subgroup (n=84) was composed of those *Ss* having d scores below the median.

The biographical data blank was item analyzed for the two subgroups. The resulting moderator key was scored so that a high score represented unpredictability. A discriminating item was scored +1 for *Ss* in the unpredictable subgroup and -1 for *Ss* in the predictable subgroup. Following Ghiselli's (1956) suggestion, the correlation between the moderator scores and the *d* scores was computed. This coefficient offered some indication of the effectiveness of the moderator scale.

R_1 between the dependent variable and the verbal and moderator scores was computed. In this instance, the moderator scores were treated as a second independent variable. Employing the appropriate *F* test, the difference between *r* and R_1 was assessed to determine if the addition of the moderator as a predictor variable resulted in a significant increase of R_1 over *r*.

As a first step in the computation of R_m, the cross-product terms obtained by multiplying the moderator scale scores by the verbal scale scores were computed for each *S*. R_m was then computed by introducing the cross-product values as a third independent variable. Using the appropriate *F* test, the difference between R_1 and R_m was assessed to determine if the moderator was operating as a moderator or as a suppressor.

Algebraic-difference Technique

The predictable and unpredictable subgroups obtained by the absolute-difference approach were used for the present technique. The unpredictable subgroup was divided into two additional groups. The overpredicted subgroup ($\underline{n}$=39) was composed of those $\underline{Ss}$ having +d scores. The underpredicted subgroup ($\underline{n}$=44) was composed of those $\underline{S}$s having -$\underline{d}$ scores.

The items in the biographical data blank were item analyzed against the overpredicted and predictable subgroups. The resulting moderator scale (M_{op}) was scored by assigning a +1 to a discriminating item for overpredicted Ss and by assigning a -1 to the same item for Ss in the predictable subgroup. A second item analysis was conducted for the underpredicted and predictable subgroups. Again the moderator (M_{up}) was keyed so that a high score represented unpredictability (underprediction) and a low score represented predictability.

The effectiveness of the moderators was assessed by determining their correlation with the $\underline{d}$ scores. If effective, $\underline{M}_{op}$ should correlate positively with the +d scores. $\underline{M}_{up}$, if effective, should correlate negatively with the -$\underline{d}$ scores.

$\underline{R}_1$ between the dependent variable and the moderator and verbal scores was computed twice, once for each moderator. Employing $\underline{F}$ tests, the difference between $\underline{r}$ and $\underline{R}_1$ for each moderator was assessed. This test permitted a determination of the ability of the moderators to operate as predictors.

$\underline{R}_m$ was computed by including the moderator-verbal cross-product term as a third independent variable. $\underline{R}_m$ was computed twice, once for each moderator. Employing

F tests, the difference between R_1 and R_m for each moderator was examined to determine if the moderators were operating as suppressor variables.

Finally, the overall moderated multiple R (R_M) employing both moderators (M_{op} and M_{up}) and all cross-product terms was computed. The cross-product term between the two moderators was not introduced for reasons of maintaining simplicity of design and interpretation. Using the proper F test, the difference between R_M and the R_m for each moderator was assessed. The difference between between the two R_m's was also tested.

Quadrant Analysis

The median z scores for the GPA indices and for the verbal scores were computed. This technique resulted in the development of two moderators based on four subgroups. The underpredicted subgroup was composed of those Ss having criterion scores above the median and predictor scores below the median. The low hit subgroup was composed of those Ss having both scores below the median. The high predictor subgroups (high hit and overprediction) were determined in the same fashion relative to median z scores.

Two item analyses of the items included in the biographical data blank were performed. The first was performed for the underpredicted and low hit subgroups. The resulting moderator scale (M_{up}) items were keyed +1 for underpredicted Ss and -1 for low hit Ss. The second item analysis was performed for the overpredicted and high hit subgroups. Discriminating items for the resulting moderator scale (M_{op}) were scored +1 for overpredicted Ss and -1 for high hit Ss. Following the procedure recommended by Hobert and Dunnette (1967), the correlation between the moderator scores and the dependent variable was computed to assess the effectiveness of the moderator scales.

The procedure for the computation of R_1, R_m, and R_M was identical to the procedure used with the algebraic-difference approach. The procedure for assessing differences between r, R_1, R_m, and R_M was also identical to the procedure used for the algebraic-difference approach. Again, for R_M the cross-product term between the two moderators was not used.

Cross-validation

The moderator scales (keys) developed on the developmental group were applied to the cross-validation group. R_m was computed using scores obtained with the absolute-difference moderator keys. Actual GPA served as the dependent variable. R_m was computed in the same manner employing the algebraic-difference and quadrant analysis keys. Using the keys developed by the latter two techniques, R_m was computed twice, once for each moderator. R_M was computed using the keys developed by the algebraic-difference and quadrant analysis approaches.

Using the appropriate F tests, between technique comparisons were made between the R_m's and between the R_M's as an attempt to determine which of the three subgrouping procedures resulted in the development of the most effective moderators. $\underline{R}_m$ was compared for all three techniques. was compared for the latter two techniques. Appropriate within technique comparisons were also made.

Chapter III

RESULTS

Developmental Sample

Mean GPA for the group was 2.2825, s = .7022. The mean verbal scale score was 495.72, s = 80.24. The correlation (r) between the GPA indices and the verbal scores was .100 ($p > .05$).

Table 2 presents the results of the item analyses. For the Absolute-difference technique, the item analysis resulted in 10 items which comprised the absolute-difference moderator scale (M_{ab}). The second two analyses for the Algebraic-difference technique (ALGD.) yielded a total of 19 items. The underpredicted moderator scale (M_{up}) was comprised of 11 items. The overpredicted moderator scale (M_{op}) was comprised of eight items.

It is noteworthy that five of the items contained in M_{ab} also appeared in M_{up} (ALGD.). The keying (+1 or -1) was identical for the shared items. The M_{ab} and M_{op} (ALGD.) scales had four items in common. Keying for the items was identical. M_{up} and M_{op} had one item in common, the keying being the same for both scales.

The item analyses for the Quadrant analysis technique (QA) resulted in a total of 24 items. The underpredicted moderator scale (M_{up}) was made up of 10 items. The remaining 14 items comprised the M_{op} scale. Seven of the items appearing in M_{up} went into making up M_{op}. However for five of the items, the keying was reversed. Of the 24 items making up the QA scales, a total of 10 were

TABLE 2

Results of Item Analyses

Item number	Chi Square	α level
	Absolute-difference	
1	1.17	$p<.30$
2	1.09	$p<.30$
3	1.87	$p<.20$
8	2.44	$p<.20$
10	1.78	$p<.20$
11	2.68	$p<.20$
27	1.11	$p<.30$
28	2.59	$p<.20$
34	1.35	$p<.30$
40	1.12	$p<.30$
	Algebraic-difference (M_{up})	
2	1.93	$p<.20$
3	3.73	$p<.10$
11	1.38	$p<.30$
13	1.66	$p<.20$
20	1.33	$p<.30$
21	2.10	$p<.20$
25	1.22	$p<.30$
27	2.93	$p<.10$
28	2.74	$p<.10$

TABLE 2 (Con't)

Item number	Chi Square	α level
32	1.50	$p<.30$
45	5.19	$p<.05$
	Algebraic-difference (M_{op})	
8	6.75	$p<.01$
10	1.78	$p<.20$
11	1.89	$p<.20$
22	3.38	$p<.10$
33	2.33	$p<.20$
40	5.09	$p<.05$
43	3.51	$p<.10$
44	1.32	$p<.30$
	Quadrant analysis (M_{up})	
2	2.21	$p<.10$
10	1.41	$p<.30$
11	2.51	$p<.20$
13	2.08	$p<.20$
21	3.41	$p<.10$
24	2.65	$p<.20$
29	7.45	$p<.01$
32	2.64	$p<.20$
36	16.48	$p<.01$
45	1.39	$p<.30$

TABLE 2 (Con't)

Item number	Chi Square	α level
Quadrant analysis (M_{op})		
9	1.16	$p < .30$
11	1.08	$p < .30$
13	1.16	$p < .30$
19	2.96	$p < .10$
21	1.18	$p < .30$
23	1.81	$p < .20$
27	1.32	$p < .30$
29	1.36	$p < .30$
32	4.12	$p < .05$
35	1.49	$p < .30$
36	3.66	$p < .10$
43	2.31	$p < .20$
44	3.11	$p < .10$
45	7.98	$p < .01$

shared with M_{ab} and the ALGD. scales. The five scales were made up of a total of 27 different items. The five scales have been reproduced in Appendix B. Table 3 presents the mean, standard deviation (s), and range of scores for the scales.

The check on the effectiveness of M_{ab}, as proposed by Ghiselli (1956), resulted in an r of .34. The checks for M_{up} and M_{op} (ALGD.) were -.42 and .39 respectively. For M_{up} and M_{op} (QA), the checks were .43 and -.47 respectively. All checks were significant beyond the .01 level.

Table 4 presents the linear and moderated multiple R's based on the moderator scale scores. As indicated the only R's not significant beyond the .05 level were R_l (M_{ab}) and R_l (M_{op}, ALGD.).

Within Technique Comparisons

Tables 5, 6, and 7 present the results of the within technique comparisons between r, R_l, R_m, and R_M.

Absolute-difference. The difference between r and R_l was not significant. However R_m was significantly greater than R_l, $F(1,163)=7.45$; $p<.01$.

Algebraic-difference. Comparisons involving the M_{up} scale revealed R_l to be significantly greater than r, $F(1,164)=16.44$; $p<.01$. No significant difference was found between R_l and R_m. The difference between R_{mup} and R_M was significant, $F(3,160)=3.58$; $p<.05$; R_M was the larger of the two coefficients. Comparisons between the coefficients based on M_{op} resulted in the finding of no

TABLE 3

Mean, Standard Deviation, and Range of Moderator Scales: Developmental Group

Moderator scales by techniques	Mean	SD	Range
Absolute-diff.			
M_{ab}	.69	2.91	-6--+8
Algebraic-diff.			
M_{up}	1.53	3.32	-7--+9
M_{op}	-.36	2.84	-8--+8
Quadrant anal.			
M_{up}	1.59	3.31	-8--+10
M_{op}	-1.10	4.34	-12--+12

TABLE 4

Linear Multiple (R_l) and Moderated Multiple (R_m) Coefficients of Correlation Obtained Using Three Moderator Development Techniques

Techniques	Coefficients of correlation		
	R_l	R_m	R_M
Absolute-diff.			
M_{ab}	.15	.25*	
Algebraic-diff.			.40**
M_{up}	.32**	.32**	
M_{op}	.12	.24*	
Quadrant anal.			.49**
M_{up}	.34**	.36**	
M_{op}	.40**	.41**	

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

TABLE 5

Differences Between Correlation Coefficients:

Absolute-difference Technique

Number of independent variables	Coefficients of correlation	F
r	.10	
		2.00
R_1	.15	
		7.45**
R_m	.25	

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

TABLE 6

Differences Between Correlation Coefficients:

Algebraic-difference Technique

Number of independent variables	Coefficients of correlation	F
	M_{up}	
r	.10	
		16.44**
R_l	.32	
		.01
R_m	.32	
		3.58*
R_M	.40	
	M_{op}	
r	.10	
		.85
R_l	.12	
		7.86**
R_m	.24	
		6.12**
R_M	.40	
	M_{up} and M_{op}	
R_{mup}	.32	
		7.27**
R_{mop}	.24	

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

TABLE 7

Differences Between Correlation Coefficients: Quadrant Analysis Technique

Number of independent variables	Coefficients of correlation	F
	M_{up}	
r	.10	
		19.78**
R_1	.34	
		1.95
R_m	.36	
		7.78**
R_M	.49	
	M_{op}	
r	.10	
		30.17**
R_1	.40	
		.59
R_m	.41	
		4.99**
R_M	.49	
	M_{up} and M_{op}	
R_{mup}	.36	
		7.80**
R_{mop}	.41	

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

significant difference between $\underline{r}$ and $\underline{R}_1$. $\underline{R}_{mop}$ was significantly greater than $\underline{R}_1$, $F(1,163)=7.86$; $p<.01$. $\underline{R}_M$ was significantly greater than $\underline{R}_{mop}$, $F(3,160)=6.12$; $p<.01$. $\underline{R}_{mup}$ was significantly greater than $\underline{R}_{mop}$, $F(1,163)=7.27$; $p<.01$.

Quadrant analysis. The comparisons for M_{up} and M_{op} resulted in relatively the same findings. In both cases $\underline{R}_1$ was significantly greater than $\underline{r}$; $F(1,164)=19.78$; $p<.01$ and $F(1,164)=30.17$; $p<.01$ respectively. In neither case was $\underline{R}_m$ significantly different from R_1. $\underline{R}_M$ was significantly greater than $\underline{R}_{mup}$, $F(3,160)=7.78$; $p<.01$. $\underline{R}_M$ was also significantly greater than $\underline{R}_{mop}$, $F(3,160)=4.99$; $p<.01$. $\underline{R}_{mop}$ was significantly greater than $\underline{R}_{mup}$, $F(1,163)=7.80$; $p<.01$.

Zero-order correlations. Tables 8-14 present the zero-order correlations among the independent variables and between the independent variables and the dependent variable. Particular attention should be focused on the correlations between M and VM in Tables 8 and 10 (M_{ab} and M_{op}--ALGD. respectively). As indicated the $\underline{r}$'s are .98 and .99 respectively. In both cases the introduction of the VM variable resulted in a significant increase of the $\underline{R}_m$'s over the R_1's ($p<.01$). In no instance was the correlation between the moderator variables and the verbal-moderator interaction variables for the same scale less than .98. The correlations between the verbal scale and the five moderator variables ranged from -.19 to .37. The correlations between V and VM ranged from -.16 to .38.

TABLE 8

Zero-order Correlations:

Absolute-difference Technique

R_{mab}

	GPA	V	M	VM
GPA	1.00	.10	-.11	-.14
V		1.00	-.03	.03
M			1.00	.98**
VM				1.00

**p<.01

TABLE 9

Zero-order Correlations:

Algebraic-difference Technique

R_{mup}

	GPA	V	M	VM
GPA	1.00	.10	.28**	.28**
V		1.00	-.19*	-.11
M			1.00	.99**
VM				1.00

TABLE 10

Zero-order Correlations:

Algebraic-difference Technique

R_{mop}

	GPA	V	M	VM
GPA	1.00	.10	-.03	-.06
V		1.00	.37**	.38**
M			1.00	.99**
				1.00

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

TABLE 11

Zero-order Correlations:

Algebraic-difference Technique

R_M

	GPA	V	M_{up}	VM_{up}	M_{op}	VM_{op}	VMM
GPA	1.00	.10	.28**	.28**	-.03	-.06	-.04
V		1.00	-.19*	-.11	.37**	.38**	.08
M_{up}			1.00	.99**	.10	.09	-.08
VM_{up}				1.00	.12	.10	-.03
M_{op}					1.00	.99**	.21**
VM_{op}						1.00	.18*
VMM							1.00

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

TABLE 12

Zero-order Correlations:

Quadrant Analysis Technique

R_{mup}

	GPA	V	M	VM
GPA	1.00	.10	.33**	.32**
V		1.00	.07	.17*
M			1.00	.98**
VM				1.00

*p<.05
**p<.01

TABLE 13

Zero-order Correlations:

Quadrant Analysis Technique

R_{mop}

	GPA	V	M	VM
GPA	1.00	.10	-.40**	-.41**
V		1.00	-.11	-.16*
M			1.00	.98**
VM				1.00

*p<.05
**p<.01

TABLE 14

Zero-order Correlations:

Quadrant Analysis Technique

R_M

	GPA	V	M_{up}	VM_{up}	M_{op}	VM_{op}	VMM
GPA	1.00	.10	.33**	.32**	-.40**	-.41**	-.12
V		1.00	.07	.17*	-.11	-.16*	-.15
M_{up}			1.00	.98**	-.29**	-.28**	-.23**
VM_{up}				1.00	-.30**	-.30**	-.26**
M_{op}					1.00	.98**	.37**
VM_{op}						1.00	.38**
VMM							1.00

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

Noteworthy was the finding that in those cases where $\underline{R}_m$ did not result in a significant increase over $\underline{R}_1$, the zero-order correlations between the dependent variable and the moderator and interaction variables were relatively strong. It should further be noted that in the two instances where $\underline{R}_m$ was significantly greater than $\underline{R}_1$, the correlations of the moderator and interaction variables with the dependent variable were nonsignificant ($p > .05$).

Cross-validation Sample

Mean GPA for the sample was 2.1899, s = .6892. The mean verbal scale score was 497.26, s = 76.37. The correlation ($\underline{r}$) between GPA and the verbal scores was .37 ($p < .01$). Table 15 presents the mean, standard deviation (s), and range of scores for the five moderator scales. Table 16 presents the seven moderated multiple R's. As indicated all coefficients were significant beyond the .01 level.

$\underline{R}_{mab}$ and $\underline{R}_{mup}$ and $\underline{R}_{mop}$ (ALGD.) were not significantly different from $\underline{r}$ ($p > .05$). There was no significant difference between $\underline{R}_{mup}$ and R_{mop}. In neither case was $\underline{R}_M$ (ALGD.) significantly different from $\underline{r}$, $\underline{R}_{mup}$, or $\underline{R}_{mop}$.

The three moderated $\underline{R}$'s based on QA scales were significantly greater than $\underline{r}$ ($p < .01$). $\underline{R}_M$ was significantly greater than $\underline{R}_{mup}$, $F(3,159)=5.07$; $p < .01$. $\underline{R}_M$ was also significantly greater than R_{mop}, $F(3,159)=3.51$; $p < .05$. $\underline{R}_{mop}$ was found to be significantly greater than $\underline{R}_{mup}$, $F(1,162)=4.49$; $p < .05$.

TABLE 15

Mean, Standard Deviation, and Range

of Moderator Scales: Cross-validation Group

Moderator scales by techniques	Mean	SD	Range
Absolute-diff.			
M_{ab}	1.02	2.75	-6--+6
Algebraic-diff.			
M_{up}	2.02	3.31	-7--+11
M_{op}	-.23	2.85	-6--+8
Quadrant anal.			
M_{up}	1.84	3.22	-8--+9
M_{op}	-1.50	3.62	-9--+7

TABLE 16

Moderated Multiple Coefficients of Correlation for Cross-validation Group

Technique	Moderated R's	
	R_m	R_M
Absolute-diff.		
M_{ab}	.37**	
Algebraic-diff.		
M_{up}	.40**	
M_{op}	.39**	
		.42**
Quadrant anal.		
M_{up}	.43**	
M_{op}	.46**	
		.51**

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

Between Technique Comparisons

No significant difference was found between $\underline{R}_{mab}$ and $\underline{R}_{mop}$ (ALGD.). However $\underline{R}_{mup}$ (ALGD.) was significantly greater than $\underline{R}_{mab}$, $F(1,162)=3.97$; $p<.05$.

$\underline{R}_{mup}$ (QA) was significantly greater than $\underline{R}_{mab}$, $F(1,162)=9.23$; $p<.01$. $\underline{R}_{mop}$ (QA) was also greater than $\underline{R}_{mab}$, $F(1,162)=13.97$; $p<.01$. For the $\underline{R}_m$'s based on the scores obtained with the ALGD. and QA scales, all between technique comparisons were significant. $\underline{R}_{mup}$ (QA) was significantly greater than $\underline{R}_{mup}$ (ALGD.) and $\underline{R}_{mop}$ (ALGD.); $F(1,162)=5.13$; $p<.05$ and $F(1,162)=6.36$; $p<.05$ respectively. $\underline{R}_{mop}$ (QA) was significantly greater than $\underline{R}_{mop}$ (ALGD.) and $\underline{R}_{mup}$ (ALGD.); $F(1,162)=11.03$; $p<.01$ and $F(1,162)=9.76$; $p<.01$ respectively.

$\underline{R}_M$ (QA) was found to be greater than $\underline{R}_M$ (ALGD.), $F(1,159)=16.81$; $p<.01$. $\underline{R}_M$ (QA) was also significantly greater than $\underline{R}_{mup}$ and $\underline{R}_{mop}$ (ALGD.); $F(3,159)=6.91$; $p<.01$ and $F(3,159)=7.35$; $p<.01$ respectively.

Zero-order correlations. Contained in Tables 17-19 are the zero-order correlations among the independent variables and between the independent variables and the dependent variable. As indicated in Tables 17 and 18, the only independent variable which correlated significantly with GPA was the verbal scale. With the exception of the double interaction variable ($VM_{up}M_{op}$), all independent variables for the $\underline{R}$'s based on the QA scales had significant correlations with the dependent variable.

The correlation between the moderator variables and the verbal-moderator interaction variables was in no case

TABLE 17

Zero-order Correlations:

Absolute-difference Technique

Cross-validation Group

R_{mab}

	GPA	V	M	VM
GPA	1.00	.37**	-.04	-.03
V		1.00	.01	.04
M			1.00	.99**
VM				1.00

*p<.05
**p<.01

TABLE 18

Zero-order Correlations:

Algebraic-difference Technique

Cross-validation Group

R_{mup}, R_{mop}, and R_M

	GPA	V	M_{up}	VM_{up}	M_{op}	VM_{op}	VMM
GPA	1.00	.37**	.10	.12	-.01	.00	-.04
V		1.00	-.12	-.04	.22**	.20*	.12
M_{up}			1.00	.98**	.03	.03	-.18*
VM_{up}				1.00	.05	.04	-.13
M_{op}					1.00	.99**	.38**
VM_{op}						1.00	.36**
VMM							1.00

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

TABLE 19

Zero-order Correlations:

Quadrant Analysis Technique

Cross-validation Group

R_{mup}, R_{mop}, and R_M

	GPA	V	M_{up}	VM_{up}	M_{op}	VM_{op}	VMM
GPA	1.00	.37**	.16*	.22**	-.30**	-.33**	-.14
V		1.00	.04	.14	-.13	-.20*	-.01
M_{up}			1.00	.98**	-.15	-.13	-.36**
VM_{up}				1.00	-.14	-.13	-.37**
M_{op}					1.00	.99**	.37**
VM_{op}						1.00	.36**
VMM							1.00

*p<.05
**p<.01

less than .98. For $\underline{R}_{mab}$ (Table 17) and $\underline{R}_{mup}$ (Table 18) the correlations between V and M and between V and VM were not significant. For $\underline{R}_{mop}$ (ALGD.) both correlations were significant; $p<.01$ and $p<.05$ respectively. For $\underline{R}_{mop}$ (QA) only the relation between V and VM was significant, $p<.05$. Neither of the relationships was significant for $\underline{R}_{mup}$ (QA).

Chapter IV

DISCUSSION

Without exception the studies reviewed earlier did not present data sufficient for an adequate description of the mechanics of moderators and the effects they assess. Research in the area has been characterized by a controversy between those who are convinced that moderators are the answer to long standing problems in psychological measurement and those who maintain that moderators contribute nothing useful or additional in making measurement more precise. The data reported herein afforded a clearer and somewhat revealing description of moderator function. While casting considerable doubt on the tenability of the moderator model, the present research by no means resolved the controversy.

In addition to maintaining that an effective moderator need not correlate with accompanying predictors, Saunders (1954) pointed out that it is not necessary for a moderator and the predictors with which it is used to correlate with the dependent variable. He made no mention of the expected zero-order correlations (necessary and/or sufficient) involving the moderator-predictor interaction variable. Saunders simply characterized a moderator as a variable which correlates with error, i.e., the variance not shared in common by an independent variable and a dependent variable. He further stated that the introduction of an interaction variable containing an effective moderator should result in a significant increase in the size of $\underline{R}$. Finally, Ewen and Kirkpatrick (1967) have pointed out that a significant increase of an $\underline{R}_m$ over an $\underline{R}_1$ is evidence that a moderator is in fact operating as a moderator and not as a suppressor.

The present data indicated that the use of M_{ab} and M_{op} (ALGD.) resulted in the detection of significant moderator effects in the developmental sample, i.e., the two scales were effective as moderators. Indeed, assuming that the above mentioned investigators were correct in their reasoning, it would appear that the data warrant no other conclusion. However, close scrutiny of the zero-order correlations for $\underline{R}_{mab}$ and $\underline{R}_{mop}$ reveals the tenability of a somewhat different conclusion. For both $\underline{R}_m$'s, the interaction variables (VM) did not correlate significantly with GPA. Obviously the VM variables did not function as predictors. It is somewhat doubtful that the increase resulting from the addition of VM was necessarily due to the action of moderators. In other words the data do not necessarily lead to the conclusion that a significant moderator effect was operative in the data. It would appear that the results can be explained parsimoniously in terms of the suppression concept. Even though M_{ab} and VM_{ab} did not correlate significantly with GPA, the intercorrelation between the two was apparently of such magnitude that a significant increase due to suppression between the two occurred. The tenability of the foregoing conclusion is readily demonstrated by the example below. For

$$R^2_{1.23} = r^2_{12} + r^2_{13} - 2r_{12}r_{13}r_{23}$$

assume that $\underline{r}_{12} = .10$; $\underline{r}_{13} = .00$; $\underline{r}_{23} - .98$; and $\underline{n} = 167$. Solution of the equation reveals that the addition of $\underline{r}_{13}$ results in an $\underline{R}$ of .50 as compared to $\underline{r}_{12} = .10$. Such an increase is significant at the .01 level as assessed by an appropriate $\underline{F}$ ratio assuming a sample of comparable size as used in the study. While not directly analogous to the data under consideration, the above example demonstrates that it is possible for suppressor effects to occur when the validities of predictors are very low and the intercorrelations between the variables very high. Unlike the example, the suppressor effects in $\underline{R}_{mab}$ and $\underline{R}_{mop}$, assuming they were present, would be more complexly determined due to the presence of a third independent variable. In other words, the ability of a variable to operate as a suppressor would be more complexly determined by its intercorrelations with a greater number of variables.

Due to the negligible correlations of M and VM with GPA for $\underline{R}_{mop}$ (ALGD.) it is highly improbable that any suppressor effects between the two occurred. However both variables correlated significantly with V. It is unlikely that M or VM taken separately contributed significantly as suppressors due to their moderate correlations with V. However it would seem plausible that the combined suppressor effects produced by both may have resulted in the increase of $\underline{R}_m$ over $\underline{R}_1$. Such an explanation seems tenable in view of the fact that the addition of M alone resulted in no significant increase of $\underline{R}_1$ over $\underline{r}$.

The increases of the $\underline{R}_M$'s over the $\underline{R}_m$'s based on the QA and ALGD. scales for the developmental sample appear to be readily explained in terms of the action of suppressor and predictor variables. For the ALGD. scales, the effectiveness of M_{op} and VM_{op} as suppressors has already been considered. M_{up} and VM_{up} obviously contributed to the size of $\underline{R}_M$ as predictors as indicated by their significant validities. Due to the high degree of overlap between the two variables, the exclusion of either would result in no appreciable decrease in the size of $\underline{R}_M$. It seems reasonable to conclude that the double interaction variable did not contribute to the size of $\underline{R}_M$. Even though it correlated significantly with M_{op} and VM_{op}, the negligible validities of those two variables would make it improbable that VMM acted as a suppressor.

M_{up} and M_{op} for the QA scales functioned as predictors as indicated by their significant correlations with GPA. $\underline{R}_M$ for the same scales was significantly greater than the two $\underline{R}_m$'s. The size of $\underline{R}_M$ was most likely, for the greater part, due to the combination of the moderators as predictors. The conclusion seems justified that VMM acted as a suppressor in view of its insignificant validity and significant overlap with M_{up}, VM_{up}, M_{op}, and VM_{op}.

The foregoing findings do not necessarily mean that research on the moderator model should be abandoned. The preceeding explanations of moderator effects in terms of the suppression concept do however point up a crucial fallacy in the logic employed to justify the use of certain tests which are supposedly sensitive to moderator effects and moderator-suppressor differences. Saunders' (1954) and

Ewen and Kirkpatrick's (1967) arguments would appear to have been valid based on the premisses employed. However their arguments were apparently invalid due to the exclusion of conditional premisses concerning possible zero-order correlations involving the interaction variable. Based on the present study, the importance of the interaction variable in accounting for the results has been clearly demonstrated.

The test suggested by Lee (1961), though not carried out in the present research, would also seem to be of questionable usefulness. Lee pointed out that evidence for a moderator effect may exist if the regression weight of the interaction variable departs significantly from zero. However, this test, like the one suggested by Ewen and Kirkpatrick does not permit a distinction between moderators and suppressors. An interaction variable could have a significant weight and still correlate insignificantly with the dependent variable. Such a state of affairs is within the realm of possibility due to the fact that the intercorrelations among a set of variables contribute to the size of the regression weight for each variable. Thus a suppressor may have a significant weight as a result of its correlations with other independent variables.

As previously pointed out, Hobert and Dunnette (1967) maintained that those developmental techniques employing the finest subgrouping should yield the most effective moderators. Additionally it was maintained that the use of the more effective moderators should result in the largest $\underline{R}$'s.

The oross-validation data for the present study appear to have partially confirmed their expectations.

Considering the first technique, the ability of M_{ab} to increase $\underline{R}$ significantly above $\underline{r}$ did not hold up on cross-validation. The M and VM variables had negligible validities. VM though correlating .99 with M produced no suppression due to the extremely low validity of M. Such findings as the above are not unusual in light of one characteristic that moderators and suppressors share-speoificity. Another explanation for effects not cross-validating, particularly in the present data, would seem to be the high initial correlation that existed between V and GPA (.37).

M_{op} (ALGD.) did not cross-validate in its ability to produce a significant increase in correlation nor was it significantly larger that $\underline{R}_{mab}$. $\underline{R}_{mup}$ (ALGD.) though not significantly different from $\underline{r}$, was significantly larger than $\underline{R}_{mab}$. This seemingly incongruous finding is easily accounted for when one considers the difference in error terms for the $\underline{F}$ ratios employed to compare $\underline{r}$ with $\underline{R}_{mab}$ and $\underline{R}_{mup}$ as opposed to the $\underline{R}_{mab}$ and $\underline{R}_{mup}$ comparison. It would appear that the greater size of $\underline{R}_{mup}$ can be accounted for in terms of the action of a suppressor variable. Even though the correlations of M_{up} and VM_{up} with GPA were insignificant, the intercorrelation between VM and M was apparently of such magnitude that a suppression effect oocurred between the two variables. Recall the example demonstrating the effectiveness of suppressors among variables having low validities but extremely high intercorrelations. Obviously a necessary condition for the occurrence of such effects is a relatively large sample size. For smaller samples, the increase produced by such suppressor effects would probably not be of such magnitude that they would reach significance.

The $\underline{R}$'s based on the QA scales were significantly greater than the $\underline{R}$'s based on scales developed by the first two techniques. Examination of the intercorrelations involving the M_{up} variables reveals that M_{up} and VM_{up} acted as predictor variables. M_{op} and VM_{op} also acted as predictors. Thus the ability of the variables to function as predictors cross-validated. $\underline{R}_{mop}$ was apparently greater than $\underline{R}_{mup}$ due to the increased effectiveness of the M_{op} variables as predictors. It is unlikely that the VM's in either case contributed anything additional as indicated by the high overlap between the M and VM variables.

$\underline{R}_{M}$ for the QA scales was significantly larger than any $\underline{R}$ obtained in the cross-validation sample. Due to the insignificant intercorrelations between the M_{op}-VM_{op} variables and the M_{up}-VM_{up} variables, it may be assumed that the size of $\underline{R}_{M}$ was partially due to the combined predictor effects produced by the combination of the M_{up} and M_{op} variables (VM's included). Special attention should also be paid to the double interaction variable VMM. The variable did not contribute anything as a predictor. However it probably produced a multiple suppression effect which may have contributed to the size of $\underline{R}_{M}$.

While seemingly lending support to Hobert and Dunnette's assertions, results of the present research do not necessarily warrant the conclusion that one technique is better than another for developing moderators for the simple reason that the moderator model is of questionable tenability. The data would seem to warrant the conclusion that the size of $\underline{R}$ is directly related to the number of subgroups employed by a developmental technique.

There would appear to be several conditions in a set of data that might permit a distinction between moderator variables and other independent variables. One such condition would involve an intercorrelation matrix for an $\underline{R}_m$ in which the V, M, and VM variables intercorrelated negligibly (V=predictor; M=moderator; VM=interaction variable). The moderator model would appear to be tenable if in such a matrix it could be shown that the addition of the VM variable resulted in an increase in R due to its action as a predictor. Further the moderator model would gain additional tenability if it could be shown that a VM variable can function as an effective predictor when V and M have negligible validities. This is apparently just what Saunders (1954) had in mind in his discussion of the VM variable and interactive effects in a set of data. In the absence of sufficient data it cannot be determined whether such results obtained in Saunders' research or in the research of anyone else. The present study would seem to indicate that conditions such as those just mentioned are mathematical improbabilities. In the interaction variable, M serves the function of a weight for the V variable. Thus when the M and VM variables are correlated, very high overlap is not unusual. In view of the nature of the M and VM variables, it is highly improbable that a VM variable could operate as an effective predictor under the aforementioned hypothetical conditions.

Undoubtedly a great deal of research needs to be carried out in order that a comprehensive enumeration of moderator characteristics may be obtained. One worthwhile undertaking would be a detailed examination of the intercorrelation matrices for different subgroups identified using moderators. It would be interesting to see if variation in the _R_'s for the groups might be due to the action of suppressors. Depending on the unique combination of independent variables for certain so called unpredictable subgroups, it may be found that one or more variables functions as a suppressor. Research along these lines would appear crucial to a better understanding of moderators and how they differ from suppressors, assuming that they do. Indeed, it is incumbent upon those in the "moderator camp" to demonstrate the uniqueness of the phenomenon with which they are working.

Chapter V

SUMMARY

In a recent article, Hobert and Dunnette (1967) maintained that compared to the Absolute-difference and Algebraic-difference techniques Quadrant Analysis should yield the more effective moderators. Their assertion was based on the reasoning that more homogeneous subgrouping should yield moderators with increased sensitivity to error. The present study was carried out as an empirical investigation of their assertions. Further, the design of the study permitted an investigation of moderator function.

The total sample (_n_=333) of male college students was randomly divided into a developmental sample and a cross-validation sample. Employing the aforementioned techniques a total of five moderator scales were developed and applied to the cross-validation sample. Based on the obtained findings, the following conclusions were drawn:

1. The tenability of the moderator model is questionable in the face of apparently fallacious reasoning concerning moderator characteristics and function.

2. Moderator function apparently can be more parsimoniously accounted for in terms of the suppression concept.

3. Previously suggested tests for the presence of moderator effects are inadequate in that they do not necessarily distinguish moderator effects from effects produced by suppressor variables.

4. The use of scales developed by those techniques employing more homogeneous subgrouping results in the attainment of larger _R_'s.

BIBLIOGRAPHY

Abelson, R. P. Sex differences in predictability of college grades. Educational and Psychological Measurement, 1952, 12, 638-644.

Banas, P. A. An investigation of transituational moderators. Unpublished doctoral dissertation, University of Minnesota, 1964.

Banas, P. A., & Nash, A. N. Differential predictability: Selection of handicapped and non-handicapped. Personnel and Guidance Journal, 1966, 45, 227-230.

Berdie, R. F. Intra-individual variability and predictability. Educational and Psychological Measurement, 1961, 21, 663-676.

Bowers, J. E. Prediction of freshman achievement with a moderator model. Educational and Psychological Measurement, 1967, 27, 427-428.

Bowers, J. E. Review of Form KSA_{45}. 1965. Cited by O. K. Buros (Ed.), The Sixth Mental Measurements Yearbook. New Jersey: The Gryphon Press, 1965.

Campbell, D. P., & Trockman, R. A verification scale for the Minnesota Vocational Interest Inventory. Journal of Applied Psychology, 1963, 47, 276-279.

Chansky, N. M. A note on the grade point average in research. Educational and Psychological Measurement, 1964, 24, 95-99.

Dawis, Rene V., Weiss, D. J., Lofquist, L. H., & Betz, Ellen. Satisfaction as a moderator in the prediction of satisfactoriness. Proceedings, 75th Annual Convention, APA, 1967.

Dunnette, M. D. A modified model for test validation and selection research. Journal of Applied Psychology, 1963, 47, 317-323.

England, G. W. The interest factor in undergraduate engineering achievement. Personnel and Guidance Journal, 1960, 38, 401-405.

Ewen, R. B., & Kirkpatrick, J. J. The selection of minority groups: Predictive and moderated validities for samples of nursing students. Paper presented at American Psychological Association, Washington, D.C., September 1967.

Filbeck, R., & Callis, R. A verification scale for the Strong Vocational Interest Blank. Journal of Applied Psychology, 1961, 45, 318-324.

Frederiksen, N., & Gilbert, A. C. F. Replication of a study of differential predictability. Educational and Psychological Measurement, 1960, 20, 759-767.

Frederiksen, N., & Melville, S. D. Differential predictability in the use of test scores. Educational and Psychological Measurement, 1954, 14, 647-656.

Friedlander, F. The impact of laboratory training on groups of differing demographic composition. Working paper No. 3. Case Institute of Technology, January 1967.

Gaylord, R. H., & Carroll, J. B. A general approach to the problem of the population control variable. American Psychologist, 1948, 3, 310. (Abstract)

Ghiselli, E. E. Differentiation of individuals in terms of their predictability. Journal of Applied Psychology, 1956, 40, 374-377.

Ghiselli, E. E. Differentiation of tests in terms of the accuracy with which they predict for a given individual. Educational and Psychological Measurement, 1960, 20, 675-684. (a)

Ghiselli, E. E. Moderating effects and differential reliability and validity. Journal of Applied Psychology, 1963, 47, 81-86.

Ghiselli, E. E. The prediction of predictability. Educational and Psychological Measurement, 1960, 20, 3-8. (b)

Ghiselli, E. E., & Sanders, E. P. Moderating heteroscedasticity. Educational and Psychological Measurement, 1967, 27, 581-590.

Goodstein, L. D., & Heilbrun, A. B. Prediction of college achievement from the Edwards Personal Preference Schedule at three levels of intellectual ability. Journal of Applied Psychology, 1962, 46, 317-320.

Grooms, L. D., & Endler, N. S. The effect of anxiety on academic achievement. Journal of Educational Psychology, 1960, 51, 299-304.

Guion, R. M. Personnel selection. Psychological Review, 1967, 18, 202-204.

Hakel, M. D. Prediction of college achievement from the Edwards Personal Preference Schedule using intellectual ability as a moderator. Journal of Applied Psychology, 1966, 50, 336-340.

Hewer, Vivian H. Are Tests fair to college students from homes with low socio-economic status? Personnel and Guidance Journal, 1965, 43, 764-769.

Hobert, R., & Dunnette, M. D. Development of moderator variables to enhance the prediction of managerial effectiveness. Journal of Applied Psychology, 1967, 51, 50-64.

Hoyt, D. P., & Norman, W. T. Adjustment and academic predictability. Journal of Counseling Psychology, 1954, 1, 96-99.

Hull, C. L. Aptitude testing. Yonkers, New York: World Book Company, 1928.

Kipnis, D. A noncognitive correlate of performance among lower aptitude men. Journal of Applied Psychology, 1962, 46, 76-80. (Abstract)

Kogan, N., & Wallach, M. A. Risk taking: A study in cognition and personality. New York: Holt, Rinehart and Winston, 1964.

Korman, A. K. Self-esteem as a moderator of the relationship between self-perceived abilities and vocational choice. Journal of Applied Psychology, 1967, 51, 65-67.

Kuder, G. F. Administrator's manual: Kuder Preference Record, Form C. Chicago: Science Research Associates Inc., 1960.

Lawler, E. E. Ability as a moderator of the relationship between job attitudes and job performance. Personnel Psychology, 1966, 19, 153-164.

Lee, Marilyn C. Interaction, configurations, and non-additive models. Educational and Psychological Measurement, 1961, 21, 797-805.

Lykken, D. T., & Rose, R. Psychological prediction from actuarial tables. Journal of Clinical Psychology, 1963, 2, 139-151.

Malnig, L. R. Differential predictability of academic achievement from aptitude tests administered under psychological stress to subjects with varying amounts of anxiety. Dissertation Abstracts, 1960, 20, 3197-3198.

Marks, M. R. How to build better theories, tests, and therapies: The off quadrant approach. American Psychologist, 1964, 19, 793-798.

Medvedeff, E. The utility of sub-grouping analysis in the prediction of sales success. Unpublished doctoral dissertation, Purdue University, 1964.

Neidt, C. O., & Malloy, J. P. A technique for keying items of an inventory to be added to an existing test battery. Journal of Applied Psychology, 1954, 38, 308-312.

Richardson, H. Utility of a new method for predicting college grades. Journal of General Psychology, 1965, 72, 159-164.

Rimland, B. Multidimensional scatter-plotting: a graphic approach to profile analysis. Journal of Applied Psychology, 1960, 44, 404-406.

Rock, D. A. Improving the prediction of academic achievement by population moderators. Unpublished doctoral dissertation, Purdue University, 1965.

Saunders, D. R. Moderator variables in prediction. Educational and Psychological Measurement, 1956, 16, 209-222.

Saunders, D. R. The "moderator variable" as a useful tool in prediction, 1954. Cited by Anne Anastasi (Ed.), Testing problems in perspective. Educational Testing Service, Princeton, New Jersey, 1965.

Steinemen, J. H. Use of a logically related predictor in determining intra-group differential predictability. Journal of Applied Psychology, 1964, 48, 336-338.

Thorndike, R. L. The concepts of over- and underachievement. New York: Bureau of Publications, Teachers College, Columbia University, 1963.

Tolbert, E. L. Predictability in testing and counseling. Personnel and Guidance Journal, 1966, 45, 219-226.

Toops, H. A. The use of addends in experimental control, social census, and man managerial research. Psychological Bulletin, 1948, 45, 41-74.

Zimmerman, W. S. Review of Form KSA_{45}.1965. Cited by O. K. Buros (Ed.). The Sixth Mental Measurements Yearbook. New Jersey: The Gryphon Press, 1965.

APPENDIX A

Moderator pool

1. How do you feel about your share of happiness in life?

    a. Have had nothing but bad breaks.
    b. Have had more than your share of bad breaks.
    c. Have had more good breaks than bad ones.
    d. Luck has been your way practically all the time.

2. How often do you feel dissatisfied with yourself?

    a. Frequently.
    b. Occasionally.
    c. Rarely.
    d. Hardly ever.

3. How often do you feel discouraged?

    a. Frequently
    b. Occasionally.
    c. Rarely.
    d. Hardly ever.

4. Up to the age of 21 years, approximately how often did you suffer minor illnesses?

    a. More often than the average person.
    b. About as often as the average person.
    c. Less often than the average person.
    d. Never.

5. In recent years, has your health been:

    a. Excellent.
    b. Good.
    c. Fair.
    d. Poor.

6. Does a hard day's work tire you out?

    a. Much more than the average person my age.
    b. Somewhat more that the average person my age.
    c. Somewhat less than the average person my age.
    d. Much less than the average person my age.

7. How long does it usually take you to fall asleep?

    a. Can go to sleep right away, at any time of the day or night.
    b. Can go to sleep in 15 minutes to half an hour.
    c. Usually need half an hour or more to fall asleep.
    d. No consistent pattern; depends on how tired, etc.

8. On the average, how much sleep do you require to feel really good?

    a. Less than 5 hours.
    b. 5 to 7 hours.
    c. 7 to 8 hours.
    d. More than 8 hours.

9. How many days were you sick in bed last year?

a. None.
b. 1 to 2 days.
c. 3 to 5 days.
d. Over 5 days.

10. How much education did your father have?

a. Grade school or less.
b. High school.
c. College.
d. A graduate degree (M.A., M.S., Ph.D., etc.).

11. How much schooling did your mother have?

a. Grade school or less.
b. High school.
c. College.
d. A graduate degree (M.A., M.S., Ph.D., etc.).

12. How much independence do you feel your parents allowed you while in high school?

a. Quite restrictive.
b. About as much as the rest of your friends.
c. Quite lenient.
d. As much as you wanted.

13. When you were growing up, about how many books were around the house?

a. A large library.
b. Several bookcases full.
c. One bookcase full.
d. A few books.

14. How often were you allowed to use the family car?

a. Had your own, did not use their car.
b. Not at all.
c. As often as you asked.
d. Only on special occasions.

15. Who did most of the repair work around your home?

a. Yourself.
b. Another member of the family.
c. Someone hired to do the job.
d. No special person.

16. For commendable behavior as a child, how were you usually rewarded?
a. Praised.
b. Given a present.
c. Given no special attention.
d. Something else.

17. How were you usually punished as a child?

    a. Punished physically.
    b. Reprimanded verbally, or deprived of something.
    c. Told how you should have acted.
    d. Warned not to do it again, but seldom punished.

18. Who influenced your conduct most when you were a child?

    a. Your father.
    b. Your mother.
    c. A brother or sister.
    d. Someone else.

19. Who made the major decisions in your family?

    a. Your mother.
    b. Your father.
    c. Some other person.
    d. Discussion and common agreement.

20. While in high school, how many hours a week did you spend doing chores and tasks around the home?

    a. One hour or less.
    b. 2 to 4 hours.
    c. 5 to 7 hours.
    d. More than 7 hours.

21. When you were a child were you punished by your parents for not doing well in school?

    a. Yes, frequently.
    b. Yes, occasionally.
    c. Very seldom.
    d. Never.

22. In high school, did you:

    a. Lead a clique or gang.
    b. Belong to a clique or gang.
    c. Keep to yourself.
    d. None of the above.

23. With regard to taking risks, which best describes you:

    a. Hardly ever take a risk.
    b. Sometimes take a risk.
    c. Generally take a risk.
    d. I'm a gambler at heart.

24. How many times during the past five years have you held a position as president, captain, or chairman of any clubs, teams, committees, or study groups?

24. (cont.)

a. Never.
b. Once.
c. Two or three times.
d. Four or more times.

25. How many elective offices have you held in the last five years?

a. None.
b. 1 or 2.
c. 3 to 5.
d. 6 or more.

26. How do you feel concerning the adequacy of your high school preparation for college?

a. Was very adequate.
b. Was weak in certain areas.
c. Was very inadequate.
d. Unable to answer.

27. As you grew up, how did you feel about school?

a. Liked it very much.
b. Liked it most of the time.
c. Just accepted it as necessary.
d. Was often unhappy with it.

28. During your teens, how did you compare with others of your own sex in rate of progress through school?

a. Advanced much more rapidly than most.
b. Advanced just a little faster than most.
c. About the same as most.
d. Progressed just a little slower than most.

29. How would you classify your potential as a student in college?

a. Considerably above average.
b. Somewhat above average.
c. Average.
d. Below average.

30. How did your teachers generally regard you in school?

a. As able to get things done with ease.
b. As a hard worker.
c. As not interested in school subjects.
d. As something of a "problem".

31. At what time of day did you do most of your best studying?

   a. Morning.
   b. Afternoon.
   c. Night.
   d. No particular time.

32. What was your standing in your high school class?

   a. Below the average.
   b. Above average.
   c. In the upper 25%.
   d. In the upper 10%.

33. How difficult was high school work for you?

   a. Fairly easy.
   b. Neither easy nor hard.
   c. Fairly hard.
   d. Quite hard.

34. Most teachers in college:

   a. Require far too much work of their students.
   b. Require slightly too much work of their students.
   c. Require about the right amount of work.
   d. Require too little work of their students.

35. What do you think is the most important thing a person should get out of college?

   a. Training for a profession.
   b. General cultural knowledge.
   c. Personal maturity.
   d. Social polish.

36. Which one of the following types of teachers would you prefer to have (as a college student)?

   a. Very hard to get good grades from.
   b. Harder than average to get good grades from.
   c. About average in difficulty.
   d. Easier than the average to get good grades from.

37. How well do you do most things you have decided to do?

   a. You almost always succeed in the things you attempt and do them better than most people could.
   b. You often find you have bitten off more than you can chew and have to give up.
   c. You usually get the things done that you attempt, but you seldom do them as well as you want to.
   d. You find that you do most things as well as other people do.

38. Do you generally do your best:

    a. At whatever job you are doing.
    b. Only in what you are interested.
    c. Only when it is demanded of you.
    d. On few if any jobs.

39. How greatly disturbed are you if something is left unfinished.

    a. Slightly.
    b. Moderately.
    c. Considerably.
    d. Highly.

40. What do you consider to be the major motivating force in your life?

    a. Prestige.
    b. Material gains.
    c. To gain a position of security.
    d. Something else.

41. Assuming you had sufficient musical ability and training to perform in the following capacities, which one do you believe would give you the greatest personal satisfaction?

    a. Soloist -- instrumental or vacal.
    b. Composer.
    c. Conducter.
    d. Member of orchestra or choral group--not soloist.

42. Which do you enjoy most?

    a. A good "bull session".
    b. Working or studying hard.
    c. Listening to music.
    d. Reading for pleasure.

43. Which one of the following seems most important to you?

    a. A pleasant home and family life.
    b. A challenging and exciting job.
    c. Getting ahead in the world.
    d. Being active and accepted in community affairs.

44. Which of the following is most important to you?

    a. Professional status or authority.
    b. Money.
    c. Family and Friends.
    d. Religion.

45. Where do you feel that you gained the most knowledge?
    a. School.
    b. Home.
    c. Personal experience.
    d. Examples set by others.

APPENDIX B

Moderator scales

Absolute-difference M_{ab}

Item number	Alternatives	Key
1.	a., b.	-1
	c., d.	+1
2.	a., b.	+1
	c., d.	-1
3.	a., b.	+1
	c., d.	-1
8.	a., b.	+1
	c., d.	-1
10.	a., b.	+1
	c., d.	-1
11.	a., b.	+1
	c., d.	-1
27.	a., b.	-1
	c., d.	+1
28.	a., b.	-1
	c., d.	+1
34.	a., b.	+1
	c., d.	-1
40.	a., c.	-1
	b., d.	+1

Algebraic-difference M_{up}

Item number	Alternatives	Key
2.	a., b.	+1
	c., d.	-1
3.	a., b.	+1
	c., d.	-1
11.	a., b.	+1
	c., d.	-1
13.	a., b.	-1
	c., d.	+1
20.	a., b.	+1
	c., d.	

Item number	Alternatives	Key
21.	a., b.	-1
	c., d.	+1
25.	a., b.	-1
	c., d.	+1
27.	a., b.	-1
	c., d.	+1
28.	a., b.	-1
	c., d.	+1
32.	a., b.	-1
	c., d.	+1
45.	a., b., d.	+1
	c.	-1

Algebraic-difference M_{op}

Item number	Alternatives	Key
8.	a., b.	+1
	c., d.	-1
10.	a., b.	+1
	c., d.	-1
11.	a., b.	+1
	c., d.	-1
22.	a., b.	+1
	c., d.	-1
33.	a., b.	+1
	c., d.	-1
40.	a., c.	-1
	b., d.	+1
43.	a.	-1
	b., c., d.	+1
44.	a., b.	+1
	c., d.	-1

Quadrant Analysis M_{up}

Item number	Alternatives	Key
2.	a., b.	+1
	c., d.	-1
10.	a., b.	+1
	c., d.	-1
11.	a., b.	+1
	c., d.	-1
13.	a., b.	-1
	c., d.	+1
21.	a., b.	-1
	c., d.	+1
24.	a., b.	+1
	c., d.	-1
29.	a., b.	+1
	c., d.	-1
32.	a., b.	-1
	c., d.	+1
36.	a., b.	-1
	c., d.	+1
45.	a., b., d.	+1
	c.	-1

Quadrant Analysis M_{op}

Item number	Alternatives	Key
9.	a., b.	-1
	c., d.	+1
11.	a., b.	+1
	c., d.	-1
13.	a., b.	+1
	c., d.	-1
19.	a., b.	-1
	c., d.	+1
21.	a., b.	+1
	c., d.	-1
23.	a., b.	-1
	c., d.	+1

Item number	Alternatives	Key
27.	a., b.	-1
	c., d.	+1
29.	a., b.	-1
	c., d.	+1
32.	a., b.	+1
	c., d.	-1
35.	a., b.	-1
	c., d.	+1
36.	a., b.	-1
	c., d.	+1
43.	a.	+1
	b., c., d.	-1
44.	a., b.	-1
	c., d.	+1
45.	a., b., d.	-1
	c.	+1

APPENDIX C.

Single and multiple linear regression analysis program

Library Listing: 6.0.148 - <u>1620 Single and Multiple Linear Regression Analysis Program</u>, by Anthony J. Capato, Columbia University. The program uses a least squares solution in computing the multiple <u>R</u>. The maximum number of independent variables is ten, the number of data points being unlimited. Included in the output are the partial regression coefficients, simple correlations, the multiple correlation, standard error of the Y data, standard error of the estimate, significance of regression, and the standard error of the partial regression coefficients.

VITA

James Porter Tuck Jr., born on May 14, 1944, in Richmond, Virginia, attended Lee-Davis High School from which he graduated in June, 1962. He entered the University of Richmond in September, 1962, and majored in Psychology. In April, 1965, he was initiated into Psi Chi. In April, 1966 he was elected to membership in Phi Beta Kappa. He was awarded the degree of Bachelor of Arts in June, 1966. In the latter part of June, 1966, he was married to Joyce Leigh Wilkinson. In September, 1966, he began work toward the degree of Master of Arts in Psychology at the University of Richmond. He expects to be awarded the Master of Arts degree in August, 1968. In September, 1968, he will begin work toward a doctoral degree in Psychological Measurement and Statistics at Purdue University.