



2  
(1999)



MICHIGAN STATE LIBRARIES

3 1293 01801 7032

This is to certify that the  
dissertation entitled  
PSYCHOLOGICAL MEASUREMENT AND STATISTICAL INFERENCE:  
IMPLICATIONS OF SCALE MISSPECIFICATION  
FOR MODERATED MULTIPLE REGRESSION

presented by  
William Michael Rogers

has been accepted towards fulfillment  
of the requirements for

Ph.D. degree in Psychology

  
Major professor

Date 12/3/98

**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.  
**MAY BE RECALLED** with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
MAR 24 2001 100400		
<del>APR 05 2001</del>		

**PSYCHOLOGICAL MEASUREMENT AND STATISTICAL INFERENCE:  
IMPLICATIONS OF SCALE MISSPECIFICATION FOR MODERATED MULTIPLE  
REGRESSION**

**By**

**William Michael Rogers**

**Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of**

**DOCTOR OF PHILOSOPHY**

**Department of Psychology**

**1998**



## ABSTRACT

### PSYCHOLOGICAL MEASUREMENT AND STATISTICAL INFERENCE: IMPLICATIONS OF SCALE MISSPECIFICATION FOR MODERATED MULTIPLE REGRESSION

By

William Michael Rogers

The purpose of this thesis was to reexamine the critical relationship between scales of measurement, moderated multiple regression, and theoretical inference. The first major section critically reviews the modern measurement paradigm in psychology, and argues that psychologists have placed too much faith in both their measures and their methodologies. The second section narrows the issue, focusing on how these uncertainties in measurement scales can affect inferences and tests using moderated multiple regression methods. It is shown that weaker scales prevent the researcher from making conclusive statements about presence and strength of moderating effects. A study is conducted, by which the effects of measurement scale on interpretation of moderated multiple regression in a variety of situations is clarified. It is shown that the interpretability of obtained effect sizes for interaction effects is based, in part, on the precision of both predictor and criterion measures. In addition, the overall predictability of criterion measures appears to be a factor in the complex relationship between measurement precision and interaction effects.

Copyright by  
WILLIAM MICHAEL ROGERS  
1998

## ACKNOWLEDGMENTS

There are several individuals whose contributions to this dissertation are noteworthy. First, I would like to thank the faculty at Michigan State University, especially members of my dissertation committee. Neal Schmitt, my committee chair, kept the dissertation process streamlined, correcting my tendencies to vainly allocate time toward the mathematically intractable. Rick DeShon provided helpful advice in developing and programming the mathematical simulations used in the thesis. Alexander von Eye helped clarify the theoretical issues in defining and understanding scales of measurement. Ann Marie Ryan was instrumental in framing the practical implications of measurement theory for applied psychology.

I would also like to thank the several scholars whose work has greatly influenced my own philosophy of measurement, and motivated me toward further study: R. Duncan Luce, Louis Narens, Patrick Suppes, David Krantz, Amos Tversky, Joel Michell, and Jean-Claude Falmagne.

Finally, I would like to thank my parents, for believing in me at times when it was difficult for me to believe in myself. Without their constant encouragement, completing the dissertation would have been impossible.

## TABLE OF CONTENTS

LIST OF TABLES .....	vii
INTRODUCTION .....	1
Measurement Theory in Psychology: Campbell’s Problem and Stevens’ Solution .....	4
Psychological Data: Ordinal, Interval, or Unimportant? .....	9
Interval Scales, Ratio Scales, and Moderated Multiple Regression .....	15
Moderated Multiple Regression and Ordinal Scales: Criterion Issues .....	21
Moderated Multiple Regression and Ordinal Scales: Predictor Issues .....	29
Moderated Multiple Regression and Level of Measurement : Summary and Implications .....	36
Rationale and Overview for the Study .....	39
Simultaneous Conjoint Measurement .....	42
The MORALS Algorithm .....	49
Research Design .....	50
Study Independent Variables .....	50
Study Dependent Variables .....	52
Hypotheses .....	54
METHOD .....	60
Structure of Predictor Variables and Error Variance .....	63
Dataset Generation .....	67
RESULTS .....	69
DISCUSSION .....	85
Effects of Baseline $R^2$ .....	85
Effects of Incremental $R^2$ .....	86
Effects of Predictor Intercorrelation .....	88
Effects of Measurement Properties of Variables .....	89
Crossing vs. Non-crossing Interactions .....	90
Design Interaction Effects .....	90
Measurement, Interaction Effects, and Psychology .....	91
Practical Implications of the Study .....	95
APPENDIX A: MORALS Algorithm .....	100

APPENDIX B1: $f^2$ Values of Crossing Interactions with Two Continuous Predictors .....	102
APPENDIX B2: $f^2$ Values of Crossing Interactions with One Continuous and One Binary Predictor .....	103
APPENDIX B3: $f^2$ Values for Crossing Interactions with Two Binary Predictors .....	104
APPENDIX B4: Pre-Post Correlation Coefficients for Crossing Interactions with Two Continuous Predictors .....	105
APPENDIX B5: Pre-Post Correlation Coefficients for Crossing Interactions with One Continuous and One Binary Predictor .....	106
APPENDIX B6: Pre-Post Correlation Coefficients for Crossing Interactions with Two Binary Predictors .....	107
LIST OF REFERENCES .....	108

## LIST OF TABLES

Table 1.	Datasets for Slope Bias Example: GPA by Race and SAT Score .....	23
Table 2.	Datasets for Social Behavior Example: Social Behaviors by Work Experience and Time .....	26
Table 3.	Dataset for Predictor Rescaling Example: Y by X and Z .....	32
Table 4.	Dataset in Table 1a presented as SAT by Levels of Race and GPA .....	33
Table 5.	Performance by Levels of Motivation and Ability .....	45
Table 6.	Mean $\Delta f^2$ by $R^2$ of Additive Model and Measurement Level of Variables .....	70
Table 7.	Mean Correlation by $R^2$ of Additive Model and Measurement Level of Variables .....	72
Table 8.	Mean Correlation by Incremental $R^2$ of Interaction Effect and Measurement Level of Variables .....	74
Table 9.	Mean $\Delta f^2$ by Incremental $R^2$ of Interaction Effect and Measurement Level of Variables .....	75
Table 10.	Mean $\Delta f^2$ by Pre-Transformation Predictor Intercorrelation and Measurement Level of Variables .....	77
Table 11.	Mean Pre-Post Transformation Correlation by Pre-Transformation Predictor Intercorrelation and Measurement Level of Variables .....	78
Table 12.	Incremental Attenuation ( $\Delta\Delta f^2$ ) by Variable and Variables Already Transformed .....	80
Table 13.	Mean $\Delta f^2$ by Non-Crossing / Crossing Interaction and Measurement Level of Variables .....	82
Table 14.	Mean $\Delta f^2$ Values for Study Design Factors .....	96

## INTRODUCTION

Many theories and procedures in applied psychology predict interactive or moderating relationships between independent variables in determining their effects on a dependent variable. Moderating effects are usually defined (e.g. Zedeck, 1971) as situations in which the bivariate relationship between two variables (X, Y) is influenced by a third variable (Z). Personnel psychologists typically use the moderator concept in assessing test bias. Evidence for moderating effects of categorical variables such as gender or ethnicity is considered differential prediction, and the test is deemed biased against a subgroup defined by the moderator (Cleary, 1968). Use of moderators is also prevalent in other applied domains, such as organizational behavior (e.g. Pierce, Gardner, Dunham, & Cummings, 1993), and training / skill acquisition (e.g. Kanfer & Ackerman, 1989). As the theoretical models generated to explain or predict human behavior in organizational settings become more complex, the development of theories specifying moderator variables will grow in importance.

The primary purpose of early studies using moderators was not to test theories or detect test bias, but to assess differential validity across subgroups defined by a third variable (e.g. Ghiselli, 1956; Saunders, 1956). This usage was predicated on the notion that moderator variables defined homogenous subgroups, within which criterion-related validity was generally thought to be more accurately assessed, and, in some cases, of greater magnitude. As such, this initial use of moderator variables was primarily atheoretical and focused on validity maximization, rather than on substantive relationships between grouping variables, predictors, and criteria (Lubinski & Humphreys, 1990).

As the early use of moderator variables was in differential validity assessment, the primary method of examining moderator effects was based on the comparison of subgroup correlations. The use of this method was restricted to categorically-defined subgroups, but this was less severe of a restriction, given the nature of most subgrouping variables (e.g. gender, race). However, it posed problems for the treatment of continuous moderator variables. Since it is generally not desirable to collapse continuous scores into categories, the preferred modern method used to assess interactive or moderating relationships is moderated multiple regression (MMR) (Saunders, 1956; Zedeck, 1971). This method has been shown to provide more information than subgroup correlational analysis, in the form of subgroup slopes (Stone-Romero & Anderson, 1994), and can be applied to situations with either dichotomous or continuous moderator variables.

Using MMR, the test for interactive or moderating effects is a test on the regression weight of a multiplicative term composed of both predictor components. Expressed in terms of a linear model, this is as follows:

$$Y = b_0 + b_1 X + b_2 Z + b_3 XZ + e \quad (1)$$

where Y is a continuous dependent variable, X is a predictor variable, and Z is a predictor variable thought to have a moderating effect on the X-Y relationship. A test of the significance of  $b_3$ , in this case, is a test of the relevant moderating effect. This test is mathematically equivalent to a hierarchical F-test of the incremental  $R^2$  for the above model over a reduced model without the XZ product term.



Despite the theoretical importance of interactions in applied psychology, confirming evidence has often been difficult to gather. Cronbach (1987) notes the difficulties in finding interaction effects to be statistically significant. Zedeck (1971) has termed moderator effects “as elusive as suppressor variables”. Moderator effects have also been characterized as more difficult to detect in non-experimental field settings than in experimental settings (McClelland & Judd, 1993; Morris, Sherman, & Mansfield, 1986). Although the failure to discover significant moderating effects using MMR led some researchers to advocate alternative methodologies, such methods were eventually shown to be invalid (c.f. Wise, Peters, & O’Connor, 1984).

In response to these difficulties, many researchers have investigated statistical artifacts which may contribute to Type II errors using MMR. Type I errors (detecting an interaction when one is not present in the population) are assumed to be controlled for by the significance level of either the t-test for the product term in the MMR equation, or the F-test for incremental  $R^2$  after inclusion of the product term. Several factors, such as small sample size (Alexander & DeShon, 1994), measurement error (Busemeyer & Jones, 1983), small population effect sizes (Stone-Romero & Anderson, 1994), and range restriction (Aguinas & Stone-Romero, 1997) have all been shown to increase Type II errors and reduce the power of the MMR method. These factors reduce the probability of concluding interaction effects are present when they are, in fact, present in the population.

While the above findings are of considerable practical utility, they have perhaps overshadowed more fundamental issues related to the erroneous interpretation of moderated multiple regression analysis. These issues are rooted in the measurement properties of the variables used by the investigator. These measurement properties, and

the associated scales of measurement, are typically defined based on the prevalent theory of scale types (Stevens, 1946).

### Measurement Theory in Psychology: Campbell's Problem and Stevens' Solution

In the majority of psychological circles, the name S.S. Stevens is synonymous with scales of measurement. His nominal, ordinal, interval, and ratio categorizations (Stevens, 1946) have been almost uniformly accepted in the psychological literature, and rarely do methodological or statistical textbooks go beyond these concepts when discussing measurement scales. The unanimity of this acceptance cannot, however, be explained solely by the utility of Stevens' model. Rather, Stevens' impetus for developing such a measurement taxonomy, and the taxonomy's subsequent widespread acceptance, owe themselves, in part, to a reactive stance by early 20<sup>th</sup> century psychologists against influential measurement theories, most notably the ideas of Campbell (1920, 1928).

Campbell, a physicist, attempted to formalize *extensive* measurement within physics. Extensive measurement is the numerical representation of physically additive properties of objects. Mathematicians such as Helmholtz (1887) and Hölder (1901) had developed complex theorems and proofs for such measurement systems, and physical concepts were readily applied to the theory. Length, mass, and distance are common examples. Measurement of these properties is based on the empirical *concatenation*, or physical addition, of identified subunits (e.g. meters, grams) which correspond to the object being assessed. Campbell, and others (e.g. Bridgman, 1922), proposed that extensive measurement is the only basis for measurement, and any scale or measurement

system must, at some level, be based on extensively measured entities. Accordingly, Campbell called extensive measurement *fundamental* measurement. The measurement of empirical properties which could not be extensively measured, but was instead composed of fundamental measures, was called *derived* measurement. Concepts such as density and acceleration are derived measures, as they are determined by simple mathematical relationships between fundamental measures (e.g. density derived from mass and volume), or powers thereof (e.g. acceleration derived from velocity). Campbell's theory essentially classified measurement as fundamental or derived, and any scale or representation which was neither was voided as measurement.

By constraining measurement to extensive attributes, Campbell's theory had thrown down a gauntlet to psychology. The vast majority of psychological variables were not amenable to empirical concatenation operations, and next to none consisted of additive physical units. This was even true of the psychophysics discipline, which was, in the early-to-mid 20<sup>th</sup> century, considered to be the most rigorously quantitative of any field in psychology. Many psychophysicists at the time (e.g. McGregor, 1935; Johnson, 1936; Smith, 1938) attempted to integrate psychological measurement within Campbell's theory, but to little avail. In 1940, a committee of the British Association for the Advancement of Science, on which Campbell was an influential member, provided another damaging blow, formally declaring fundamental measurement in psychology an impossibility (Ferguson et al, 1940):

“Why do not psychologists accept the natural and obvious conclusion that subjective measurements of loudness in numerical terms (like those of length or

weight or brightness) ... are mutually inconsistent and cannot be the basis of measurement?"

As Campbell's theory gained support from the committee's pronouncement, the prospects of acceptable measurement in psychology grew dimmer. Given the vital role ascribed to measurement by the scientific community as a whole, the mood of many was that psychology was on the defensive in a philosophical battle for its existence as a science.

Stevens, a psychophysicist, was especially influenced by the edict of the British committee, as his own loudness sensation scale, the sone scale (Stevens & Davis, 1938), was among those that the association chose to examine in detail. Stevens proposed an alternative to Campbell's theory by relaxing the requirement of extensively measurable entities. Stevens suggested that any numerical coding which somehow represents an empirical reality should be considered measurement, regardless of the presence of additivity in either the numbers used or the empirical objects in question. Stevens' complete theory argues that empirical non-additivity does not preclude measurement itself, but only restricts the ways in which the measurements can be used.

Stevens identified four primary scale types: *nominal*, *ordinal*, *interval*, and *ratio*. His nominal scale is the numerical coding of attributes based on equality or inequality. Thus, using a '1' to represent a male and a '2' to represent a female tells us only that "males are not equal to females". Stevens' ordinal scale uses numbers to denote order properties of an attribute. A simple example is order of finish in a marathon. 1<sup>st</sup> place finishes ahead of 2<sup>nd</sup>, which finishes ahead of 3<sup>rd</sup>, and so on. Ordinal scales also contain

equality and inequality information, in the form of “ties” at any given rank. In the marathon example, if after 1<sup>st</sup> and 2<sup>nd</sup> place, three people all crossed the finish line at exactly the same moment, they could all be given rank “3”, as they finished after “2”, and before “4”. Stevens’ interval scale possesses the properties of the aforementioned scales (i.e. equality and order), and has the additional property of equality of differences. A classic example of an interval scale is temperature measurement using a thermometer. The liquid in the thermometer is known to increase in volume linearly with an increase in temperature. The marked gradations on the outside of the thermometer are set at equal intervals of the volume within the thermometer. Thus, the change in volume of the liquid from the 10° mark to the 20° mark is equal to the volume change between the 20° and 30° marks. Given the linear relationship between volume and external temperature, one can conclude that the physical temperature differences are also equivalent. Ratio scales possess the properties of equality, order, and difference, and, in addition, reflect a physically additive structure by the presence of a true zero point. Ratio scales are used when the object in question has a meaningful point of absence or non-existence. Mass and length are common examples.

Each scale type defines a set of permissible transformations, under which the information contained in the scale remains invariant. Nominal scales permit only one-to-one transformations, where any value in the transformed scale  $s_t$  has only one corresponding value in original scale  $s_o$ , and vice versa. Ordinal scales permit monotonic increasing transformations, as these preserve order from the original scale. Interval scales permit positive linear, or *affine*, transformations, of the form:  $s_t = a \cdot s_o + b$ . Ratio scales permit positive similarity transformations, which preserve the ratio of two scale values, of

the form:  $s_t = a \cdot s_o$ , where  $a$  is a positive real number. Non-permissible transformations of any scale result in loss of information, and the resulting scale can only be treated at the level of measurement which permits the transformation. For example, a non-linear monotonic transformation of an interval or ratio scale results in an ordinal scale, and a linear transformation of a ratio scale results in an interval scale. In the case of linear transformations and monotonic transformations with functional formulas (e.g.  $X^n$  or  $\log(X)$ ), a non-permissible transformation can be reversed to recover original scale information. This is not true of all monotonic transformations, however.

The notion of permissible statistics was a natural extension to permissible transformations. Permissible statistics were defined by Stevens to be functions whose meaning and statistical inference remained invariant across permissible transformations of a given scale type. Non-parametric statistics, such as frequency-based and rank-order concordance indices, were the only statistics applicable to nominal and ordinal scales, respectively. Advanced parametric methods, such as t-tests, F-tests (analysis of variance), and Pearson correlational indices were restricted to interval and ratio scales.

Accepting Stevens' theory of scale types, even with the restrictions it placed on transformations and statistics, would have represented considerable gains in measurement theory for psychology. Despite these potential gains, however, many psychologists still saw a problem. Stevens' theory essentially told them that the variables they study are indeed measurable, but due to scale properties, only certain statistics are allowable, hence only certain hypotheses could be tested. Acceptance of this, combined with the lack of evidence for an interval nature of a majority of psychological variables, would amount to an admission that the large group of meaningful statements limited to interval and ratio

scales could rarely be made in psychology. To psychologists attempting to expand measurement practice and psychological science to the boundaries of physical science, these restrictions were unacceptable. The nature of psychological data, and the methods used to analyze them, became the major point of contention.

### Psychological Data: Ordinal, Interval, or Unimportant?

The earliest attacks on Stevens' theory were based on the notion of statistical methods being closed systems. Typified by Burke (1953), Lord (1953), and Anderson (1961), these criticisms essentially stated that numerical calculations are independent of measurement scales and empirical phenomena, and thus any calculation can be conducted on any numbers. After all, states Lord (1953), "the numbers don't know where they came from." Though the logic of this statement is unclear to some<sup>1</sup>, Lord's conceptual separation of numerical computation and statistical meaning is evident. He uses examples such as calculating the arithmetic mean of jersey numbers for freshmen on a football team. Gaito (1980) supports such reasoning, suggesting that statistical theory and measurement theory are independent and unrelated considerations. This type of argument, sardonically termed "computational libertarianism" by Michell (1990), boils down to both a difference in semantics and a lack of consideration of empirical meaning. It would have been ridiculous for Stevens to suggest that calculations *cannot be done* using scale values. No researchers, at least to date, have been legally or otherwise restricted from performing mathematical calculations. These calculations should instead be judged by their eventual meaning or use in hypothesis testing. Unless researchers

following Lord's logic can propose meaningful hypotheses about the mean of nominal scales such as jersey number, the calculation of the mean remains theoretically and empirically meaningless.

Other researchers have questioned the relationship between statistical methodology and measurement theory based on statistical assumptions. Gaito (1960) presents the argument that statistical tests are only mathematically based on distributional assumptions. According to Gaito, verification of these distributional assumptions, rather than scale type, validates the use of a particular statistical method. The problem with this argument is similar to that of the early criticisms of Lord (1953) and Burke (1953), in that no explicit linkage is made with the meaning of the statistical test or hypothesis. A variety of transformations of a given variable could be conducted in order to produce a normally distributed result. Gaito is technically correct, as this would indeed validate the use of statistical methods assuming normality, but the hypotheses tested under the transformation may become meaningless (e.g. difference in logarithms of attitudes toward an object), or, at the very least, difficult to interpret. Again, the key is not valid statistical methodology, but valid empirical inference and theoretical meaning. Stine (1989) sums up the critical relationship between substantive theory, measurement, and statistical methodology, and the flaws in arguments such as Gaito's:

“In short, for the statistician or mathematician, statistical methods are closed systems. For the scientist, statistical methods are but one component of a larger, more complex system. The full potential of a statistical technique is realized only

---

<sup>1</sup>Townsend & Ashby (1984): “Just exactly what this curious statement has to do with statistics or



when its proper role as a component of the scientific endeavor is realized. A failure to recognize this role can lead to scientific decision making on the basis of nonsense.”

Another set of responses to Stevens’ theory were motivated by the desire to use parametric methodology with data that were not shown to be interval or ratio scaled. Recall that the only difference between ordinal and interval data is an equivalent distance between scale points. This meaningful distance, according to Stevens’ (1946) theory, permitted the use of advanced parametric statistical methods of testing mean differences. The difficulties in verifying equal intervals in measurement instruments, combined with the desire for parametric methodology, led many psychologists to seek proxy indicators for an interval scale. Perhaps because of its role in many parametric methods, the most popular proxy indicator of an interval scale has been the normal distribution.

Gaito (1959) reasoned that the normal distribution is evidence for an interval scale because one can divide a normal distribution into equal units based on the standard deviation. Achenbach (1978) writes: “In effect, then, the assumption of a normal curve also implies an assumption about the type of measurement scale employed.” Jensen (1974) states: “if ... normality of the population distribution of the trait is correct, we have ... a true interval scale of measurement”. In 1980, he writes: “Ipso facto, any test of intelligence that yields a normal distribution of scores must be an interval scale.”

Despite all claims to the contrary, there is no evidence that normality of distribution is a valid indicator of an interval scale (Stine, 1989). The argument of Gaito

---

measurement eludes us.”

(1959) is flawed, in that standard deviations only allow a normal curve to be divided into equal *areas* based on probabilities. The empirical distance between points on the scale is an entirely different consideration. For instance, knowing a data set has a mean of 10, standard deviation of 1, and is a perfect normal distribution only tells us that the probability of a data point falling between 8 and 9 is equal to the probability of falling between 11 and 12. It in no way informs us of the empirical equality of these distances.

Thomas (1982) illustrates a situation which further falsifies claims that normal distributions infer scale of measurement. The study he used as an example, Yuan (1933), suggested that weight be considered a lognormal variable due to the lack of negative values. Yuan graphed weight and log weight from a sample of 1000 girls and illustrated that the log weight conformed to a normal distribution much better than the untransformed weight. Since weight is a ratio scale (and consequently an interval scale), a log transform is not permissible. Thus, log weight is not an interval or ratio scale, yet it displays a normal distribution. Thomas (1982) also proves that, for any ordinal scale measuring an underlying continuous distribution, a transformation to a normal distribution exists<sup>2</sup>. He points out a startling implication of the latter proof: if we incorrectly assume normal distributions are the result of measurements using interval scales, and know that any ordinal scale can be transformed to normality, we would erroneously conclude that any ordinal scale can be transformed into an interval scale! Scholars in measurement theory have suggested there is little evidence to conclude that performance measures are linearly related to the underlying construct of interest (Krantz & Tversky, 1971). We are usually only able to show these measures to be of ordinal

---

<sup>2</sup> Proof based on Roussas (1973), pp. 185-186

level, and thus only monotonically related to the construct. Despite these findings, most psychologists still believe that normality of distribution somehow implies an interval scale.

Even among those accepting the non-interval status of measurement in psychology, there is continued use of advanced statistical methodology requiring interval scales. This use is based primarily on simulations demonstrating the robustness of parametric methods to transformations not permitted of interval scales. In one of the first, and most often cited, of these studies, Baker, Hardyck, & Petrinovich (1966) calculated  $t$ -tests on interval-level data disturbed by random ordinal transformations, finding that the sampling distributions of the resulting  $t$  statistic were very similar to the sampling distribution of the statistic with the undisturbed interval data. They therefore concluded that the  $t$ -test is robust to violations of interval assumptions, and it is adequate for use with ordinal data. Results of a Monte Carlo simulation by Gregoire & Driver (1987) suggested no clear power advantage of either parametric or non-parametric tests when testing for two group ( $t$  test vs. Mann Whitney U) or multi-group ( $F$  test vs. Kruskal Wallis H) mean differences of Likert scales under various ordinal transformations. However, reinterpretations of their results have implied a power superiority of parametric tests (see Rasmussen, 1989). Zumbo & Zimmerman (1993) demonstrated minimal power differences when  $t$ -tests are applied to ranked data, or to ranked scores with added error, and conclude that it is not necessary to use non-parametric tests on ordinal level data.

The message emerging from these investigations is that it is generally permissible to use parametric tests with ordinal data, or at the very worst, it is an arbitrary consideration. This conclusion has been challenged by Stine (1989), who notes that

simulations based on random disturbances to interval data, such as that of Baker et al. (1966) and Zumbo & Zimmerman (1993), may not be valid. According to Stine, these Monte Carlo methods are valid only if different ordinal scales are randomly selected for each use of a given statistic, i.e. that *different* ordinal representations of the data are used across replications. Stine argues that it is more likely for a single ordinal representation (or disturbance) to be in effect across many situations. For example, because of anchoring or other problems, a Likert scale may have “compressed” values near one endpoint, such that the empirical distance between ratings of ‘6’ and ‘7’ is less than that between ‘1’ and ‘2’. The behavior of a series of *t*-tests using this scale may be very different than a series conducted on a Likert scale which exhibits random “compression”. Stine concludes that if a permissible transformation exists (i.e. an *empirically* equivalent scale) such that the inferences made (or decision error probabilities of such) using the statistical method in question are altered, then the method is not robust to violations of interval level assumptions.

By advocating proxy indicators of interval scales, such as normality of distribution, and using parametric methodologies with ordinal data, psychologists have perhaps, to use an athletic analogy, both inappropriately lowered the bar and strengthened the high-jumper, with respect to measurement and statistical methodology. The next section will describe how these uncertainties in level of measurement have the potential for misinterpretations of moderated multiple regression analyses.

### Interval Scales, Ratio Scales, and Moderated Multiple Regression

Although the Monte Carlo studies discussed above have quelled most concerns about robustness of parametric methods, Stine's criticisms notwithstanding, the effects of measurement level have been further investigated in the context of moderated multiple regression. Since, for reasons discussed earlier, most psychological data is assumed to be of interval scale, much attention has been given to the effects of linear transformations on parameter values and interpretation of MMR. In addition to the standard linear transformation ( $s_t = a \cdot s_o + c$ , where  $s_t$  is the transformed scale,  $s_o$  is the original scale,  $a$  is a positive real number, and  $c$  is any real number), the special case of additive transformations ( $s_t = s_o + c$ , where  $c$  is any real number) has been addressed. Additive transformations of the form  $s_t = s_o - \bar{s}_o$  are often used to "center" the data prior to estimating the regression equation. Such centering simplifies interpretation of simple slopes and often reduces multicollinearity between the product term and its component terms (Aiken & West, 1991).

Perhaps the most comprehensive assessment of the effect of additive and linear transformations on regression equations with product terms is that of Cohen (1978). In an effort to demonstrate the invariance of results across linear transformations, Cohen illustrated the effects algebraically. Given arbitrary linear transformations of the predictors,  $X'$  and  $Z'$ , where  $X' = aX + c$ , and  $Z' = dZ + f$ , the simplified MMR equation in terms of the transformed scales becomes:

$$Y = \left( \frac{\beta_1}{a} - \frac{\beta_3 f}{ad} \right) X' + \left( \frac{\beta_2}{d} - \frac{\beta_3 c}{ad} \right) Z' + \left( \frac{\beta_3}{ad} \right) X'Z' + \left( \beta_0 - \frac{\beta_1 c}{a} - \frac{\beta_2 f}{d} + \frac{\beta_3 cf}{ad} \right) \quad (2)$$

Regarding the new regression coefficients for variables  $X'$ ,  $Z'$ , and  $X'Z'$ , several conclusions are apparent. Coefficients for  $X'$  and  $Z'$  are altered by both additive (c,f) and multiplicative (a,d) components of the transformation equations. This can easily be seen in the common situation where one wishes to standardize the predictors prior to analysis (i.e.,  $a = SD_X^{-1}$ ,  $d = SD_Z^{-1}$ ,  $c = -\bar{X}$ ,  $e = -\bar{Z}$ ), resulting in standardized regression coefficients. If there is no interaction present using the original scales ( $\beta_3 = 0$ ), then transformations which are solely multiplicative ( $c = e = 0$ ) will simply change  $\beta_1$  and  $\beta_2$  by a factor of their respective multiplicative constants, a and d. The regression weight for  $X'Z'$  is shown to be affected by the multiplicative constants a and d, but unaffected by additive components of the transformations. Thus, the “centering” operation described by Aiken & West (1991) has no effect on the value of  $\beta_3$ . Cohen (1978) also showed that under a linear transformation, the new regression weight counteracts the shift in the standard deviation of the product term created by the transformation. Because of this, significance tests on the transformed regression weight remain unchanged.

Important as they are, the examination of these effects on regression weights was not Cohen’s primary focus. Cohen demonstrated that  $R^2_{Y \bullet X, Z, XZ} = R^2_{Y \bullet X', Z', X'Z'}$ , and that  $R^2_{Y \bullet X, Z} = R^2_{Y \bullet X', Z'}$ , leaving the F-test for incremental  $R^2$  also unchanged. Cohen’s overall conclusion is that, despite the various effects on regression weights, the essential tests of interactive effects are invariant to linear transformations. According to Cohen, this demonstration renders concerns such as multicollinearity of product terms and

components (Althausen, 1971), and correlated random predictors (Sackloff, 1976) of trivial importance.

There have also been some discussions of ratio level data in MMR analyses. If one examines Equation 2 above, it can be seen that additive transformations exist ( $a = d = 1$ ;  $c = \beta_2/\beta_3$ ,  $f = \beta_1/\beta_3$ ) which equate the regression weights on  $X'$  and  $Z'$  to zero. In this situation, all predictable variance from the original equation is carried by the  $X'Z'$  product term ( $R^2_{Y \cdot X, Z, XZ} = R^2_{Y \cdot X'Z'}$ ). If one were to use the  $R^2$  as an index of fit for the model, it results in an arbitrary decision between a strictly multiplicative model ( $Y = X'Z'$ ) and an additive-multiplicative model ( $Y = X + Z + XZ$ ), depending on which interval scales are used. Schmidt (1973) illustrates this point using Vroom's (1964) Expectancy-Valence theory of motivation. Since the additive transformations which create the ambiguity between the multiplicative and additive-multiplicative models are not permissible of ratio scales, Schmidt concludes that ratio scales are necessary to make the distinction. Using ratio level measures of valence and expectancy, for instance, a researcher could only perform multiplicative transformations, which cannot convert an additive-multiplicative model to multiplicative, or vice versa.

Arnold & Evans (1979), based primarily on Cohen's (1978) previously discussed work, take issue with Schmidt (1973), and suggest that the MMR F-test of incremental  $R^2$  is the proper test of a multiplicative model. According to Arnold & Evans, the proportion of variance "carried" on the product term or its components is not relevant to testing a multiplicative relationship. In support of their point, Arnold & Evans (1979) present an example of two physicists attempting to verify the ideal gas law, which is given by:

$$V = \frac{RT}{P} \quad (3)$$

where V=volume, P=pressure, T=temperature, and R=a constant.

It is at this point where Arnold & Evans make an error. Perhaps due to a misunderstanding of physical laws, they claim the primary difference between physical laws and relationships in psychology is that physical laws must specify units of measurement (e.g. centimeters, degrees Kelvin, etc.) in order to be valid. The ideal gas law is only valid, according to Arnold & Evans, if temperature (T) is measured in units Kelvin. This statement is incorrect, as the measure of temperature only has to be a ratio scale, like the other scales in the formula. As can be seen in Equation (3), the value of the constant R can simply be adjusted to reflect a permissible ratio rescaling of T, and the law will maintain its fit to data and, more importantly, its multiplicative form.

This confusion between measurement level and measurement unit is compounded as Arnold & Evans (1979) describe their example. Physicist A uses degrees Kelvin as a measure of temperature and Physicist B uses degrees Celsius. To test their theories, these physicists set up a moderated multiple regression equation as follows:

$$V = b_0 + b_1 T + b_2 \frac{1}{P} + b_3 \left( T \cdot \frac{1}{P} \right) \quad (4)$$

Both physicists are expecting to verify the ideal gas law by finding: 1) a significant increase in  $R^2$  when the  $(T \cdot 1/P)$  term is added; 2)  $b_0$ ,  $b_1$ , and  $b_2$  all to be zero; and 3) for  $b_3$  to equal the constant R. Arnold & Evans (1979) state: “*This rather strong*



*prediction is based upon their confidence that their measures of  $T$ , and  $P$ , and  $V$  are on ratio scales.*” We will soon see that this statement cannot be true.

Both physicists run their analyses and predict 100% of the variance after the product term is added, concluding they have indeed verified a law. Physicist A finds that the increase in  $R^2$  was highly significant,  $b_0 = b_1 = b_2 = 0$ , and  $b_3 =$  the constant  $R$ . Thus, she concludes the underlying physical law is:

$$V = \frac{RT}{P},$$

or, the correct law as originally described in Equation 3.

Physicist B, who also finds an identical significant increase in  $R^2$ , finds his equation to be somewhat different. As with Physicist A,  $b_0$  and  $b_1$  are both zero, and  $b_3 =$  the constant  $R$ . However,  $b_2$  is now equal to a new constant  $K$ . Physicist B concludes the correct form of the ideal gas law is:

$$V = \frac{K}{P} + \frac{RT}{P},$$

which reduces to:

$$V = \frac{K + RT}{P}.$$

Recall that the only difference between Physicists A and B is the scale chosen for temperature, Kelvin or Celsius. Arnold & Evans use this fact to argue that neither physicist’s formula for the law is really correct. A simple change in the unit of measurement has effectively changed a law from a multiplicative form to an additive-

multiplicative form. This ostensibly supports their argument that laws are not constant without specification of measurement units.

As noted earlier, the flaw in such an argument is based on a confusion between measurement *unit* and measurement *level*. Arnold & Evans commit this error when they allow Physicist B to use a Celsius scale for temperature, and simultaneously assert that both physicists are confident their scales are of ratio level. Kelvin is a measure of temperature based on molecular activity, and thus has a true zero point. Celsius is a measure of temperature constructed by means of a non-permissible additive transformation to the Kelvin measure ( $^{\circ}\text{C} = ^{\circ}\text{K} - 273^{\circ}$ ). Thus, the Celsius scale loses the true zero point (e.g.,  $40^{\circ}\text{C}$  is not twice as much warmth or molecular activity as  $20^{\circ}\text{C}$ ), and is merely an interval scale. Physicist A can be confident her formula is the correct *form* of the ideal gas law, as permissible transformations of all its ratio measures can only change the value of the constant  $R$ , and will leave the essential multiplicative form of the law unchanged. However, Physicist B's additive-multiplicative model can be changed by permissible linear transformations of the temperature measure.

This discussion reinforces Schmidt's (1973) earlier arguments. Verification of a multiplicative theoretical model cannot be accomplished by means of moderated multiple regression when the variables are measured at the interval level. Although Cohen's (1978) work showed that the significance test for a product term is invariant to linear transformation, this test is not equivalent to a test of a purely multiplicative theory. The test of this interaction term with interval scales does, however, allow one to *reject* a purely multiplicative model based on a zero-weighted interaction term. An additive model ( $Y = X + Z$ ) cannot be made multiplicative by a linear transformation. If one

obtains a significant interaction using MMR, a further examination of the variables in question is required to verify a multiplicative model. Arnold & Evans (1979) make this point, and this author is in complete agreement. However, a clear concept of the *level* of measurement is required, not, as Arnold & Evans suggest, the *unit* of measurement. While theory and measurement in applied psychology may never reach the point of specifying standardized units, the goal of establishing constructs with theoretically meaningful zero points is more realistic, and the only necessary consideration.

#### Moderated Multiple Regression and Ordinal Scales: Criterion Issues

The use of ordinal scales with moderated multiple regression poses a more severe set of problems. It has been repeatedly noted that permissible monotonic transformations of dependent variables can completely remove non-crossing interaction effects (Cliff, 1992; Loftus, 1978; Busemeyer & Jones, 1983; Krantz & Tversky, 1971), and can often attenuate a crossing interaction to the point of potential non-significance (Busemeyer, 1980). Note that these findings also imply that data suggesting no interaction is present can be subject to a monotonic transformation which creates a significant interaction term in an MMR analysis (see Loftus, 1978). An interaction is said to be “non-crossing” when the rank orderings of Y across X are the same for all Z values, and the rankings of Y across Z are the same for all X values. Any rank order changes indicate a “crossing” interaction, as a plot of Y regressed on X (or Z) would cross at the point on Z (or X) where the order change occurred.

Unfortunately, the effects described above are not easily demonstrated algebraically, as Cohen (1978) had done with linear transformations. A few popular

monotonic transformations, such as power ( $X^n$ ), root ( $\sqrt[n]{X}$ ), and logarithm, have been discussed (see Birnbaum, 1973; Busemeyer & Jones, 1983), but these only represent a subset of monotonic transformations which have functional forms. Since monotonic transformations, as a class, cannot be expressed with a functional formula, the examples here will use small data sets and simple transformations.

First, let us examine a non-crossing interaction. Consider a situation in which we are determining whether or not slope bias exists for Black and White subgroups when predicting college achievement, indexed by GPA, from the score on a standardized SAT test. The dataset for this example, based in part on Figure 7.7 in Gregory (1996), p. 268, is shown in Table 1a. It can be seen that GPA maintains the same rank ordering within Race across levels of SAT score, and within SAT score across levels of Race, verifying a non-crossing interaction.

If one were to examine these data with moderated multiple regression, the following additive and additive-multiplicative equations would be generated:

$$GPA_A = (.006 \times SAT) + (.85 \times Race) - 1.44 \quad (5)$$

$$GPA_{AM} = (.002 \times SAT) - (.17 \times Race) + (.0026 \times SAT \times Race) + .09 \quad (6)$$

The  $R^2$  for Equations (5) and (6) are .960 and .998, respectively. The F-test for  $\Delta R^2$ , as well as the t-test of the regression weight of the (SAT  $\times$  Race) interaction term, are both significant at the .001 level. Based on these results, a researcher would conclude that slope bias exists when using a standardized SAT score to predict college

achievement. This conclusion would likely result in a more detailed examination of the test, and perhaps its discontinued use.

Table 1.

Datasets for Slope Bias Example: GPA by Race and SAT Score

---

a. Original Data

<u>Race</u>	<u>SAT Score</u>				
	<u>200</u>	<u>300</u>	<u>400</u>	<u>500</u>	<u>600</u>
White	1.1	2.0	2.5	3.25	4.0
Black	0.8	1.3	1.7	2.2	2.6

b. Transformed Criterion Data

<u>Race</u>	<u>SAT Score</u>				
	<u>200</u>	<u>300</u>	<u>400</u>	<u>500</u>	<u>600</u>
White	1.1	2.0	2.4	2.9	3.3
Black	0.4	1.3	1.7	2.2	2.6

---

Rather than conclude the problem lies with the predictor in this case, an examination of the criterion may be in order. There is no reason to believe that GPA is an interval scale of college achievement. For example, the difference in achievement between GPA's of 4.0 and 3.5 may be much greater than between 3.5 and 3.0, due to a

particular grading policy which requires students to put forth “extra effort” in order to attain very high grades. The potential variety of such grading policies casts any interpretation of GPA as an interval scale in doubt.

In light of this, consider the dataset presented in Table 1b. These GPA values have an identical rank ordering as the data in Table 1a, and thus represent a permissible monotonic rescaling of the original GPA variable. Comparing Tables 1a and 1b illustrates that the rescaling is simply slight, order-preserving changes to four GPA values. Conducting moderated multiple regression on the rescaled numbers would result in the following equations:

$$GPA_A = (.0053 \times SAT) + (.7 \times Race) - 1.18 \quad (7)$$

$$GPA_{AM} = (.0053 \times SAT) + (.7 \times Race) + (.000 \times SAT \times Race) - 1.18 \quad (8)$$

The  $R^2$  for Equations (7) and (8) are obviously the same, .976. The  $\Delta R^2$  is zero, so there is no additional variance accounted for by adding a (Race  $\times$  SAT) product term. A researcher using this dataset would conclude that no evidence of slope bias exists for this standardized test. While there is evidence of an overall mean difference in GPA across Race, the regression lines for each subgroup are otherwise identical.

The above example illustrates how a non-crossing interaction can be completely removed by applying a monotonic transformation to the dependent variable. In the specific case of assessing slope bias, it is notable that the level of measurement of the criterion may have serious implications for the future use of the predictor, when the measurement properties of the predictor are not even assessed. Using two datasets which

are *empirically* equivalent non-interval scales of achievement, one can reach two different conclusions regarding the existence of slope bias. Neither conclusion is, in fact, necessarily correct or incorrect. Using non-interval scales prevents the researcher from making any statement about slope bias, as the test for slope bias is not invariant to transformations permissible of the variables involved.

Now consider an example of a crossing interaction. A researcher is studying the relationship between social integration behaviors, previous work experience, and the length of time working for a particular organization. The dependent variable, amount of social integration behavior, is measured using summated Likert response items, resulting in a 20-point instrument. The data for this example are shown in Table 2a.

The additive and additive-multiplicative equations for the dataset in Table 2a are as follows:

$$SocBeh_A = (1.1815 \times Time) + (.46 \times WorkExp) - .095 \quad (9)$$

$$SocBeh_{AM} = (-.385 \times Time) - (5.81 \times WorkExp) + (1.045 \times Time \times WorkExp) + 9.31 \quad (10)$$

The  $R^2$  for equations (9) and (10) are .827 and .987, respectively. The  $\Delta R^2$  of .16 is significant at  $\alpha=.0001$ , which the researcher concludes to be evidence of a strong interaction effect in the determination of social integration behavior. Individuals with no previous work experience tend to exhibit more social integration behavior than individuals with previous work experience when first arriving in a new organization.

Table 2.

Datasets for Social Behavior Example: Social Behaviors by Work Experience and Time

---

a. Original Dataset

<u>Previous Work Experience</u>	<u>Time from Employment</u>				
	<u>2 mo</u>	<u>3 mo</u>	<u>4 mo</u>	<u>5 mo</u>	<u>6 mo</u>
No	5.0	6.0	7.1	9.2	10.0
Yes	1.3	4.2	7.5	12.3	14.3

b. Transformed Dataset

<u>Previous Work Experience</u>	<u>Time from Employment</u>				
	<u>2 mo</u>	<u>3 mo</u>	<u>4 mo</u>	<u>5 mo</u>	<u>6 mo</u>
No	4.4	5.3	6.1	7.9	8.9
Yes	2.9	3.7	7.0	9.9	12.1

c. Transformed Dataset

<u>Previous Work Experience</u>	<u>Time from Employment</u>				
	<u>2 mo</u>	<u>3 mo</u>	<u>4 mo</u>	<u>5 mo</u>	<u>6 mo</u>
No	4.7	6.0	7.1	9.7	10.3
Yes	3.2	4.4	7.5	10.5	12.2

---



However, as time passes, the individuals with previous work experience increase their social integration behavior more rapidly than those without such experience. The researcher concludes the initial difference in behaviors is due to the newcomers without work experience attempting to “fit in”, perhaps using social behaviors to compensate for lack of knowledge of workplace etiquette. Those with previous experience, and such knowledge, have no need to compensate, and eventually their previous workplace experiences allow them involvement in more social integration behaviors.

Suppose, however, that the 20-item scale used by this researcher is a non-interval level measure of social integration behavior. For example, this could be due to problematic anchors for Likert-type items or inclusion of items reflecting very different levels of behavior. Problems such as these might create situations where distances between any two scale points would not be a constant across the entire scale. If the scale as a whole is considered ordinal, transformations preserving rank order can result in the data presented in Table 2b. Submitting these data to a moderated multiple regression analysis results in the following equations.

$$SocBeh_A = (.905 \times Time) + (.60 \times WorkExp) + .49 \quad (11)$$

$$SocBeh_{AM} = (-.07 \times Time) - (3.3 \times WorkExp) + (.65 \times Time \times WorkExp) + 6.34 \quad (12)$$

$R^2$  for equations (11) and (12) are .866 and .976, respectively, resulting in a  $\Delta R^2$  of .11. This is significant at the  $\alpha=.01$  level. However, this  $\Delta R^2$  is smaller than the .16 obtained using the original data. While the interaction is still present, its strength has

somewhat diminished. Now consider a second transformation to the original data, presented in Table 2c. The associated regression equations for these data are:

$$SocBeh_A = (.975 \times Time) + (0 \times WorkExp) + 1.71 \quad (13)$$

$$SocBeh_{AM} = (.285 \times Time) - (2.76 \times WorkExp) + (.46 \times Time \times WorkExp) + 5.85 \quad (14)$$

$R^2$  for equations (13) and (14) are .926 and .978, respectively, with  $\Delta R^2$  equal to .052. This is significant at the  $\alpha=.05$  level. The effect size associated with the interaction is smaller than that in the previous data set, and a great deal smaller than the original data set. When treating the scale as ordinal, two researchers using empirically equivalent scales can thus reach two very different conclusions about the strength of the interaction effect. As opposed to the non-crossing type of interaction, however, the regression equation can never be rendered completely additive. It is impossible to do so without affecting the rank order of the criterion variable. This can easily be understood if one thinks of the crossing interaction graphically, in terms of intersecting regression lines. An additive equation is graphically represented by parallel regression lines. Thus, in order to transform a crossing interaction model to an additive model requires the “uncrossing” of the lines to make them parallel. Such a manipulation requires that some of the rank orders near the high or low end of the criterion be inverted. It is nevertheless possible to reduce the effect size of the interaction by minimizing the scale distance between ranks at extreme values of the criterion. Graphically, this has the effect of “compressing” the ‘X’ formed by the interaction. This increases the fit of a linear equation through the ‘X’, and subsequently reduces the amount of variance accounted for by an additional product term.

While the two examples shown above still result in a significant interaction, more severe scale transformation would cause the interaction to be statistically non-significant. However, the regression weight of (and amount of variance accounted for by) the product term can never be reduced to zero.

The above sections have highlighted the problems with interpreting interaction effects when the criterion cannot claim an interval level of measurement. When the interaction is of the non-crossing variety, transformations permissible of ordinal scales can completely remove the effect, essentially converting an additive-multiplicative equation into an additive equation. When the interaction is crossing, such transformations cannot completely remove an interaction, but can potentially attenuate the effect size to the point of statistical non-significance. Although crossing interactions can never be completely removed, the reduction to non-significance would result in a researcher concluding there is no interaction effect, and advocating the default additive model.

#### Moderated Multiple Regression and Ordinal Scales: Predictor Issues

Problems using moderated multiple regression techniques with ordinal level data are not restricted to the criterion variable. Although the measurement level of a predictor is obviously only an issue with continuous predictors, as dichotomous moderating variables are only of nominal level<sup>3</sup>, monotonic transformations of a predictor can have effects on interpretation of regression results in these situations (Busemeyer, 1980; Busemeyer & Jones, 1983). Busemeyer & Jones (1983) examine the specific case

---

<sup>3</sup> Ordinal, interval, and ratio properties cannot be assessed with only two scale points.

involving quadratic transformations of a predictor variable. If there is reason to suspect a quadratic component in the relationship between a predictor and criterion, Busemeyer & Jones (1983) suggest the inclusion of higher order terms, such as  $X^2$ , in a hierarchical regression analysis. As with interactions, tests of these trend components are interpretable if they are entered into the regression equation after lower order components (Cohen & Cohen, 1983; Cohen, 1978). Testing of cubic and higher order terms proceeds in a similar manner, with the regression equations becoming exponentially larger as an increasing number of exponential and product terms are required. Rarely, however, do psychological theories obligate an assessment of trends beyond the quadratic form (Cohen, 1983).

The methods outlined by Cohen (1978) allow us to examine nonlinear trend components and their interactions within the context of moderated multiple regression, but only in the case where the functional form of the nonlinear transformation (or nonlinear relationship) is suspected or known. In a situation where we have no reason to suspect a predictor is related to a criterion via a logarithmic or polynomial function, yet also have no reason to believe the predictor is of interval level, inclusion of specific functions of predictors in a regression equation offers us little more than a “hit and miss” method of finding a critical functional transformation, if one even exists. This method is useless when faced with a non-interval predictor with unknown distances between scale points.

The general case of monotonic predictor transformations can, however, be examined in a similar fashion as that of criterion transformations, i.e., finding the transformation of  $X$  which renders  $X$ - $Y$  regression lines parallel across levels of  $Z$ .

Determining whether a transformation of  $X$ , say  $X'$ , exists, such that  $Y = b_0 + b_1X' + b_2Z$ , is equivalent to determining whether a criterion transformation exists for  $X' = [Y + b_2(-Z) - b_0] / b_1$ . Since linear transformations are a subset of monotonic transformations, we can dispose with the  $b_0$ ,  $b_1$ , and  $b_2$  terms in the equation, leaving  $X' = Y + (-Z)$ . In other words, if  $Y$  is shown to be an additive function of  $X$  and  $Z$ , then  $X$  is also an additive function of  $Y$  and  $Z$ , albeit with an inverted ordering on the  $Z$  variable. Any transformation of  $X$  which achieves  $X' = Y + (-Z)$  also achieves  $Y = X' + Z$ .

We can see from the previous data examples that, when  $Y$  values are tabled by  $X$  and  $Z$ , the transformation of a  $Y$  value is an operation on a cell, or set of equally valued cells, in the table. The transformation is order-preserving as long as the new number is less than the next highest original number and greater than the next lowest original number. For instance, in the slope bias example data in Table 1a and Table 1b, four GPA values were changed, which preserved rank ordering and completely removed the moderating effect. However, consider what a monotonic change in a predictor represents in the data tables presented to this point. No longer is one altering a single cell, but “shifting” an entire column or row. The regression equation at any level of  $Z$  is affected by a change in an  $X$  value, provided a  $Y$  value exists at the level of  $X$  and  $Z$ . For example, in Table 1a, changing any SAT value would result in changes to both White and Black regression lines, and their slopes could never be equated. Thus, it appears that the monotonic transformation of a predictor cannot remove a moderating effect.

A predictor transformation examined in this manner can, however, attenuate the interaction effect. Consider the very simple dataset in Table 3.  $R^2_A$  and  $R^2_{AM}$  for these data are .9375 and 1.00, respectively, resulting in a  $\Delta R^2$  of .0625. After one performs the

simple monotone predictor transformation ( $2=2.5$ ) on  $X$ , the  $R^2_A$  and  $R^2_{AM}$  become .894 and .952, respectively. The  $\Delta R^2$  in this case is .058, slightly lower than the original value. If one performs a further transformation ( $1=1.5$ ) on  $X$ , the  $R^2_A$  and  $R^2_{AM}$  become .917 and .978, respectively. The  $\Delta R^2$  increases from the last situation to .060. While this dataset represents a non-crossing interaction, similar effects are likely to be observed with a crossing interaction. Monotonic transformation of a predictor appears to have the potential to attenuate an interaction effect but not remove one.

Table 3.

Dataset for Predictor Rescaling Example: Y by X and Z

<u>Z</u>	<u>X</u>		
	<u>1</u>	<u>2</u>	<u>3</u>
<u>1</u>	1	2	3
<u>2</u>	2	4	6

However, this still is not the entire story. Because the labelings of “predictor” and “criterion” in these tables have already been shown to be arbitrary if a transformation to an additive model exists<sup>4</sup>,  $X$  could be considered the criterion, with  $Y$  and  $Z$  as predictors. We’ve already shown that  $Y$  can be transformed so that  $Y = b_1X + b_2Z$ , which is equivalent to saying  $b_1X = Y - b_2Z$ . If  $X$  is considered the criterion, it looks as if a

---

<sup>4</sup>  $Y = b_1X + b_2Z + b_0 \iff b_1X = Y - b_2Z - b_0$

monotonic transformation to “predictor” Y exists in the data from Table 1a to create this additive model. This is contrary to what was found when X was considered a predictor. The reason for this can be seen by examining the same data in Table 1a, but with SAT values in the cells at different levels of GPA and Race. This arrangement of the data can be found in Table 4.

Table 4.

Dataset in Table 1a presented as SAT by Levels of Race and GPA

<u>Race</u>	<u>GPA</u>									
	<u>0.8</u>	<u>1.1</u>	<u>1.3</u>	<u>1.7</u>	<u>2.0</u>	<u>2.2</u>	<u>2.5</u>	<u>2.6</u>	<u>3.25</u>	<u>4.0</u>
White		200			300		400		500	600
Black	200		300	400		500		600		

The difference between Table 4 and Table 1a is clear. Monotonic transformations of GPA in Table 4 still involve the shifting of entire columns, but since only one SAT value is in each column, we can effectively alter the regression line for one group without affecting the other. Thus, the regression slopes *can* be equated by an order-preserving rescaling of the GPA “predictor”. The cause of this phenomenon is primarily the design of the dataset. While SAT score is likely a continuous distribution in the larger sample, the table represents a dataset in which a pair of observations were selected from six levels

of SAT score, one observation for each Race. As such, the table represents a completely crossed design of Race  $\times$  SAT. Monotonic predictor changes in a fully crossed design will, by definition, alter regression equations across all levels of the other design factor. The data neither are, nor were designed to be, fully crossed in GPA  $\times$  SAT, and the empty “predictor” cells created by this crossing allow a monotonic transformation to have an effect.

The issue of predictor transformations affecting moderating effects thus reduces to the question of what situations involve empty cells, rows, or columns in the data matrix. An obvious situation is one in which X and Z are correlated. If X and Z are completely uncorrelated, as in a completely crossed design, no predictor transformation can remove the interaction, as it necessarily affects all regressions across Z. Conversely, if X and Z are perfectly correlated, Z is a linear function of X, and the moderated multiple regression equation reduces to:

$$Y = b_0 + b_1X + b_2(kX+m) + b_3(X)(kX+m) \quad (15)$$

$$Y = (b_0 + b_2m) + [X \times (b_3m + b_2k)] + [X^2 \times (b_3k)] \quad (16)$$

In this case, it can be seen that Y becomes an additive function of X and  $X^2$ . Since the quadratic term itself is a monotonic function of X, a monotonic rescaling of X can easily remove any variance accounted for by  $X^2$ , leaving only a main effect for X.

We have now seen that when predictors X and Z are perfectly orthogonal and uncorrelated, monotonic transformations cannot remove moderating relationships, but can attenuate them. In the trivial demonstration of perfect correlation, predictor



transformations can completely remove nonlinear effects. This suggests predictor intercorrelation may have an important role in determining the “robustness” of moderated multiple regression when predictor variables are not of interval level. Dunlap & Kemery (1988) have noted that increases in intercorrelation between X and Z result in the increased probability of detecting an interaction effect, and, when X and Z are measured with error, a higher reliability for the XZ product term. However, the above discussions suggest that, when predictors are of non-interval level, the increases in detection of interactions (due to predictor intercorrelation) noted by Dunlap & Kemery (1988) may paradoxically be accompanied by an increasing lack of precision when interpreting them.

The problems associated with a single predictor measured at the ordinal level are compounded when both X and Z predictors are ordinal scales. Typically, in these cases, the researcher is not interested in examining differences in regression slopes across a third variable, but in evaluating a theory which predicts a multiplicative combination of the two predictors. In addition to considering the issues related to interval and ratio scales raised by Schmidt (1973), the researcher is advised to be wary of ordinal level data. The issues raised above now apply to both X and Z variables, and one must consider the effects of monotone transformations of both simultaneously.

When faced with two predictors and a criterion of ordinal level, the researcher can make very few confident statements about the form of the relationships between the variables. Birnbaum (1973, 1974) notes that, in this situation, the multiplicative equation  $Y = a \times X^b \times Z^c$  can be rendered additive by permissible logarithmic transformations,  $\log(Y) = \log(a) + b \log(X) + c \log(Z)$ . Thus, when we lose confidence that any of our

variables are measured at an interval level, we have come full circle to the point of being unable to distinguish a multiplicative model from an additive model.

### Moderated Multiple Regression and Level of Measurement: Summary and Implications

The previous sections have described the behavior of moderated multiple regression when the predictor and criterion variables are defined by scales at various levels of measurement. Three types of models emerging from moderated regression were discussed, the *additive* model, the *additive-multiplicative* model, and the *multiplicative* model. It was shown that when our predictor and criterion data are ratio scaled, we can accurately select one of these models as providing the best fit to data. When all variables are measured at the interval level, we can confidently reject the multiplicative model, but, if failing to reject it, cannot confirm or reject an additive-multiplicative model. When the criterion is measured at the ordinal level, we can attenuate or even eliminate interaction effects, thereby making a choice between additive and additive-multiplicative models arbitrary, or choosing an additive model in default due to lack of statistical significance of an additive-multiplicative model. The extent of potential attenuation is primarily a function of whether the interaction is of the crossing or non-crossing variety. Similar effects are potentially observed when a predictor is measured at the ordinal level, though the effects may themselves be moderated by the degree of intercorrelation between the predictor variables. Further ambiguity between the additive and additive-multiplicative models can arise when both predictors are measured at the ordinal level. Finally, when all variables are ordinal, we cannot make a distinction between additive, multiplicative, or additive-multiplicative models, since transformations exist which can transform any of

the three models into any other of the three models. The relevance of these problems is borne out by the earlier discussion of measurement in psychology. As our confidence in the interval nature of psychological measurement decreases, the interpretation of moderated multiple regression results becomes more difficult.

A lack of scale precision may be an important factor reducing the “power” of moderated regression tests. While not related to *statistical* power, per se, the probability of detecting an interaction effect, when one exists, may be reduced when scales are not of interval level. Consider a situation where one thousand researchers are testing a moderating effect. Five hundred use interval scales and five hundred use non-interval scales. Since all interval scales are related by linear transformation, and we know linear transformations cannot remove moderating effects, we then know that if one of the five hundred researchers using interval scales finds a moderating effect, all of the researchers will find the effect<sup>5</sup>. The situation is bleaker for the researchers using non-interval scales. These five hundred scales, related only by monotonic transformation (perhaps slight), wouldn’t necessarily show the same moderator effect size, and some might not even show the moderating effect at all. Examining the studies using the interval scales, the scientific field as a whole would likely decide they have found a robust and important moderating effect. Using the non-interval scales, the field might argue the interaction is difficult to detect, statistically unreliable, or perhaps not even to exist. These arguments should ring familiar, as they are those currently made regarding moderator effects in applied psychology.

---

<sup>5</sup> For the sake of argument, this assumes all other research factors are the same.

Lack of scale precision may also be a factor explaining the common observation that moderator effects are more often detected in controlled, experimental settings than in applied field settings (McClelland & Judd, 1993). Several lines of reasoning point to measurement level playing an important role. First, experimental studies are more likely to use categorical predictors and test interactions via cell mean comparisons. Such tests are only affected by the scaling of the dependent measure, as predictors are merely of nominal level. Applied studies often use continuous scales for both predictor and criterion variables. In these situations, permissible transformations are given greater latitude to affect tests of moderation.

Second, the independent variables involved in experimental studies are typically under sufficient control to allow the complete crossing of factors. Even in cases where one predictor is polychotomous, this prevents predictor rescalings from removing interactions, for the reasons discussed earlier. Conversely, applied studies usually sample both predictor variables, having very little control over their intercorrelation. This intercorrelation potentially leads to a greater likelihood of predictor rescalings affecting the test of moderation.

Third, McClelland & Judd (1993) note that experimentalists typically predict crossing interactions, whereas field researchers usually predict only non-crossing interactions. We have seen that monotonic rescalings can attenuate a crossing interaction, but not remove it, and can completely remove a non-crossing interaction. For these reasons, it is possible that the use of non-interval data in applied field research poses much more of a threat to empirical meaningfulness of results than using such data in experimental settings.

While the above issues relate to the potential effects of scale misspecification on detection of interaction effects, there are also important implications for interpreting interactions that are found. Currently, applied psychologists lament that many interactions that are found account for a very small portion of overall variance. Field researchers have indeed noted that observed interactions usually account for between 1% and 3% of total variance (Champoux & Peters, 1987). The frustrating search for moderating effects has also led some authors to go so far as claiming that interactions accounting for 1% of the variance should be deemed important (Evans, 1985). In light of the demonstrations earlier in this thesis showing that minor changes to data can create large changes in moderator effects, it is possible that the meaningful interpretation of interactions accounting for 1% of variance would require measurement precision beyond the status of most psychological scales.

### Rationale and Overview for the Study

Thus far, this thesis has demonstrated that using non-interval data with moderated multiple regression procedures can have a variety of harmful effects on a researcher's ability to interpret results. These harmful effects have important implications for theory verification in applied psychology, and may especially be relevant to issues distinguishing experimental and field detection of moderators.

However, two important issues remain. First, under what conditions will these harmful effects manifest themselves? The simple demonstrations presented earlier in this thesis are not representative of the wide variety of moderator effects found in research settings. To address this issue, the study presented in this thesis examined interaction

effects in situations defined by a variety of factors, including the baseline  $R^2$  prior to adding a product term, the  $\Delta R^2$  incremental percentage of variance accounted for by the product term, the intercorrelation of predictors, and the measurement properties of all variables involved in the moderated regression equation. Results obtained from this study can assist researchers by determining what situations are most susceptible to interpretation problems when the precision of measurement is uncertain.

A second important issue is reconsidering what exactly constitutes a monotonic transformation. Some researchers might defend their scales of measurement - which cannot be proven to be interval level - by suggesting that just because a scale is not verified to be interval level does not mean we can conclude it is *merely* a rank ordering of the attribute. In this sense, scales commonly used in psychology may be thought to lie somewhere on a continuum between ordinal level and interval level. Advocates of this position might argue that violent monotonic rescalings, though technically permissible of purely ordinal data, are not reasonable with most psychological scales. This author would agree that the majority of psychological scales likely represent more than ordinal information, and lie somewhere on the continuum between ordinal and interval level, it may also be true that the “reasonableness” of the transformation may be inversely related to the strength of the observed moderating effect. Interactions with large effect sizes may require drastic rescaling to remove or attenuate the interaction to non-significance, but moderators with small effect sizes may require only slight alterations of the scales used.

Given the earlier discussion on interpreting moderators which account for very small percentages of variance (Evans, 1985), it is important to clarify this issue. If a monotonic transformation results in a scale which has measurement properties very

similar to the original data, most of the information present in the scale has been preserved, and an argument suggesting we have somehow destroyed the scale is less tenable. If such transformations remove or attenuate moderating effects with very small effect sizes, interpreting such effect sizes is likely a fruitless endeavor when one lacks very precise interval scales. This study will examine this issue by attempting to place specific monotonic transformations on the continuum between pure rank-order preserving transformations and linear transformations permissible of interval scales. This will be done by calculating the Pearson correlation coefficient between pre-transformation and post-transformation variables. Since a value of 1.00 denotes a linear transformation, I argue that very high correlations in the .8-.9 range are “reasonable” and similar to the original scale. In these cases, the transformation is not drastic, and any changes of interpretation based on the transformation should be of serious concern.

Answering the two general research questions presented above requires both a means of determining whether a moderating effect can be removed or attenuated by a monotonic transformation, and a means of generating a transformation which accomplishes such a feat. Two approaches have generally been used. The first, *simultaneous conjoint measurement*, examines the extent to which conditions are met in the dataset such that an additive, non-interactive representation is possible. Generation of such a representation is not required. The second method, Multiple Optimal Regression by Alternating Least Squares (MORALS) (Young, de Leeuw, & Takane, 1976), is an iterative algorithm for generating numerical transformations which maximize the  $R^2$  between sets of independent and dependent variables. Each of these methods is discussed in more detail below.

### Simultaneous Conjoint Measurement

Recall that one of the criticisms of Stevens' (1946) measurement paradigm was the arbitrary nature of scale assignment. Measurement level was not determined by consistent empirical relationships, but by the judgment of the investigator. The operations described by Campbell were not possible in psychology, so demonstrating the ratio or even interval level nature of data within Stevens' framework was extremely difficult. As noted earlier, much of the desire to use parametric methods with ordinal data may have been due to the imposing conditions necessary to verify an interval scale. Some psychologists, however, chose to develop alternatives to both Stevens' and Campbell's theories, attempting to loosen the restraints imposed by the latter without accepting the investigator-centered aspects of the former.

This recent avenue was spearheaded by the work of Luce & Tukey (1964). Rather than relax Campbell's requirements of empirical additivity, as Stevens did, Luce & Tukey relaxed only the requirement that the basis of additivity be physical concatenation. Luce & Tukey demonstrated that empirical addition can be based on non-physical operations and did not require use of subunits placed side-by-side. In other words, psychologists could develop what amounted to interval-level scales, in Stevens' framework, without formal extensive measurement. Luce & Tukey called this new type of measurement *simultaneous conjoint measurement*. As its name suggests, simultaneous conjoint measurement considers the combined effects of variables, rather than treat them independently. Simultaneous conjoint measurement can potentially be applied in any instance where two or more variables are thought to determine an empirical outcome



variable. Only categorical (nominal) and rank order (ordinal) properties need to be present in the determining variable set and outcome variable, respectively. If the relationships among the variables conform to a number of axioms (a series of if-then rules), necessary and sufficient conditions are met to define interval scales (termed *standard sequences*) on the set of determining variables, and subsequently on the determined variable. Essentially, the theory says that given variables  $X$  and  $Z$  which determine  $Y$ , if the necessary axioms are met, then monotonic transformations of ordered variables  $a$ ,  $b$ , and  $c$  exist, such that  $a(X) + b(Z) = c(Y)$ . These scales are defined by transforming the original component variables, producing an additive representation for all components. The work of Luce & Tukey (1964) was expanded upon in a three volume series entitled *Foundations of Measurement* (Krantz, Luce, Suppes, & Tversky, 1971; Suppes, Krantz, Luce, & Tversky, 1989; Luce, Krantz, Suppes, & Tversky, 1990), in which other types of conjoint measurement other than additive (e.g. polynomial, difference, geometric) are discussed in detail.

The development of simultaneous conjoint measurement provided exactly what early 20<sup>th</sup> century psychologists were looking for in response to Campbell's theory. If the axioms were successfully applied to psychological measures, it could provide standard sequences, i.e., interval scales, for psychological variables. The construction of these scales should also have interested adherents to Stevens' theory, as it allowed more advanced statistical techniques to be used, avoiding debates about permissible statistics. More importantly, the central concept in Campbell's original theory, additivity, had been preserved, and shown to be possible with non-physical variables. Thus, one of the problems motivating the development of Stevens' theory had, in effect, been solved.

Despite the dramatic ability of conjoint measurement to potentially produce interval-level scales and additive relationships, these are not its most important implications. Taken as a whole, the essence of the theory is that these scale definitions were produced by examining empirical relationships between one or more variables. Scales were not constructed or arbitrarily determined in isolation. Measurement, according to conjoint measurement theory, is the assignment of numbers to empirical components, such that the *relationship between* the numerical assignments adequately represents the *relationship between* the empirical components. In this fashion, it stands in contrast with Stevens' theory, which proscribed only permissible transformations for scale types, which themselves could be selected in isolation by the researcher.

Partly because of the abstract nature of the theory's presentation, demonstrations of its utility have been few in number. Early uses of the theory include areas such as animal behavior (Campbell & Masterson, 1969) and psychophysics (Levelt, Riemersma, & Bunt, 1971). Recent recognition of the relationship between conjoint measurement and use of the Rasch model in item response theory (Perline, Wright, & Wainer, 1979) has generated some linkages between the two areas, specifically in the assessment of interactions with classical vs. IRT ability estimates (Embretson, 1996).

The axioms of simultaneous conjoint measurement can perhaps be best understood with an example. Consider a researcher investigating the combined effects of ability and motivation on performance. In order for the conjoint measurement method to be applied, it is assumed that the dependent measure (in this case, performance) is of ordinal level. The determining factors (ability and motivation) are probably assumed

ordinal by the researcher, but only need to be of nominal level for conjoint measurement to be used.

It is helpful to view this situation in terms of a matrix, similar to the data tables presented earlier in the thesis. If we let  $a_1 \dots a_k$  denote different classifications (values) of Ability, and  $m_1 \dots m_n$  represent different values of Motivation, then Performance at any combination of Ability and Motivation can be denoted by  $a_a m_b$ . Thus, performance (P) when ability (A) is at level  $i$  and motivation (M) is at level  $j$ , can be represented by  $a_i m_j$ , which simply means that these levels of the variables combine in some way to result in a certain level of P. Recall that we *only* assume ordinal properties on P, with both A and M treated as nominal variables. This is shown graphically in Table 5.

Table 5.

Performance by Levels of Motivation and Ability

		<u>Ability</u>		
		<u>Level 1</u>	<u>Level 2</u>	<u>Level 3</u>
<u>Motivation</u>	<u>Level 1</u>	$m_1 a_1$	$m_1 a_2$	$m_1 a_3$
	<u>Level 2</u>	$m_2 a_1$	$m_2 a_2$	$m_2 a_3$
	<u>Level 3</u>	$m_3 a_1$	$m_3 a_2$	$m_3 a_3$

Once the data are arranged in this manner, with each cell containing the mean or value of performance at the appropriate levels, the researcher can begin testing the axioms of conjoint measurement.

The most important axiom of conjoint measurement is *double cancellation*. Essentially, the double cancellation axiom tests whether the order of certain P values implies the ordering of other P values. The axiom is stated as follows (Krantz, Luce, Suppes, & Tversky, 1971):

*For any three values of M,  $m_a, m_b, m_c$ , and any three values of A,  $a_d, a_e, a_f$ :*

*if  $m_a a_e \geq m_b a_d$  and  $m_b a_f \geq m_c a_e$  then  $m_a a_f \geq m_c a_d$ .*

The double cancellation axiom is essential to determining whether M and A have an additive relationship with P, as it assumes one exists. To illustrate this, replace each  $m_i a_j$  term above with “ $m_i + a_j$ ”. This is equivalent to stating that any given level of performance is an additive function of ability and motivation. Note that we do not invoke any concept of weights on M and A, as they are still assumed to be of nominal level, and we make no assumptions regarding their ordering. Replacing gives:

*if  $(m_a + a_e \geq m_b + a_d)$  and  $(m_b + a_f \geq m_c + a_e)$  then  $(m_a + a_f \geq m_c + a_d)$*

Summing across the left side and subtracting common terms (denoted by ~~strikeout~~ text):

*if  $(m_a + a_e + \cancel{m_b} + a_f) \geq (\cancel{m_b} + a_d + m_c + a_e)$  then  $(m_a + a_f \geq m_c + a_d)$ ,*

Thus, if there is an additive relationship between A and M in determining P, the double cancellation axiom will hold true for all values of A, M, and P (i.e.  $m_i a_j$ ) in the data set. Typically, the extent to which a data set satisfies the double cancellation axiom is indicated by the percentage of independent axiom tests which support the double cancellation axiom (Nickerson & McClelland, 1984). As this number would grow exponentially with the number of categories of A and M and number of levels in P, these tests are usually done using computer algorithms.

A second axiom of conjoint measurement is the *solvability* axiom:

*Given any three of:  $m_a$ ,  $m_b$ ,  $a_d$ ,  $a_e$ , the fourth must exist such that  $m_a a_d = m_b a_e$*

This means that values for M and A must exist such that all feasible values of P *can* be generated. In terms of the data matrix illustrated above, this simply means that for any combination of ability and motivation, there must be a level of performance, i.e. there are no structurally empty cells in the data matrix. For this example (and for nearly all psychological data), this axiom is trivial and usually assumed true. This axiom seems to suggest that ability and motivational components need to be uncorrelated to use the conjoint measurement methodology, but this isn't necessarily the case. A sample correlation between two predictors tells us nothing about which cells are impossible to exist, but only which combinations of predictors are more likely to occur.

There is another axiom of conjoint measurement, the Archimedian axiom.

Although methods have been devised to test it indirectly (Scott, 1964), it is generally considered technical in nature, and usually not tested in finite data sets (Luce et al, 1990).

Michell (1990) notes that if both solvability and double cancellation axioms are established, two additional important properties of M and A are also verified: *order* and *independence*. Independent ordering implies the following statements:

*Given  $m_1, m_2$ .  $m_2 \geq m_1$  if  $m_2a \geq m_1a$  for any  $a$  in  $A$ .*

*Given  $a_1, a_2$ .  $a_2 \geq a_1$  if  $ma_2 \geq ma_1$  for any  $m$  in  $M$ .*

Or, Level 2 of motivation ( $m_2$ ) is greater than Level 1 of motivation ( $m_1$ ) if, for any ability level ( $a_i$ ), individuals with motivation level 2 have higher performance ( $m_2a_i$ ) than individuals with motivation level 1 ( $m_1a_i$ ). This observation is identical to the notion of equal ordering of Y on X across levels of Z and Y on Z across levels of X described in the previous example data sets.

Since successful tests of the aforementioned axioms imply an ordering of P, M, and A which create an additive equation  $P = M + A$ , it is sufficient evidence that monotonic transformations exist which can eliminate any multiplicative component present in the original data. The researcher can successfully construct scales of performance, motivation, and ability, and, consistent with their theory, ability and motivation will have to be compensatory in determining performance, i.e. a given change in Motivation will result in a specific change in Performance, and be offset by a specific

change in Ability. All these changes will be constant across all scale points, thus defining interval scales for all three constructs.

Due to the nature of this study, the aforementioned axioms of simultaneous conjoint measurement proved untestable in the generated data for several reasons which will be described in more detail later in the thesis.

### The MORALS Algorithm

The goal of the MORALS algorithm is similar to that of conjoint measurement, in that an additive representation is sought. However, rather than verify conditions which permit an additive representation, the MORALS algorithm attempts to generate actual scales conforming to such a representation. The algorithm uses a least squares convergent procedure, in which least squares estimations of regression weights are alternatively performed on a matrix of transformation parameters and a matrix of regression parameters. The least squares estimates for one matrix are used in the next iteration for the other matrix, until a convergent solution is reached. Further mathematical details on the procedure can be found in Appendix A. de Leeuw, Young, & Takane (1976) and Young, de Leeuw, & Takane (1976) contain detailed conceptual and procedural discussions of the MORALS algorithm.

Having now stated the rationale and reasoning behind the study, a formal research design using statistical simulation methods will be described.

## Research Design

In order to assess the effect of measurement precision on estimation of interaction effects, the statistical simulation will examine six factors. These factors are: 1) Incremental  $R^2$  of the XZ product term; 2) Baseline  $R^2$  of the additive model prior to adding an XZ term; 3) Interrelation between predictors X and Z; 4) Measurement level of predictors and criterion, 5) quantitative/qualitative nature of X and Z; 6) Crossing or non-crossing nature of the interaction. Each is now specified in more detail.

## Study Independent Variables

Incremental/Baseline  $R^2$  and Predictor Interrelation. The first two factors in the design involve the strength of the interaction effect and the predictability of the additive model prior to adding an interaction term. The strength of an interaction effect is often indexed by the  $\Delta R^2$  after addition of a product term. This study used incremental  $R^2$  values at three levels: .05, .15, and .25. The baseline  $R^2$  of the additive model was also varied with three levels: .2, .4, .6. This resulted in a 3 x 3 crossing of factors, with maximal and minimal  $R^2$  of .85 and .25, respectively, for an interactive model. The correlation between predictors X and Z in the simulated datasets was varied at three levels: .1, .3, and .5.

The levels for the preceding three factors were selected to create a wide coverage of potential  $R^2$  values (.25 to .85), a range of additive  $R^2$  representative of those found in psychological research, as well as a wide enough spread in interrelation levels to detect small differences across level. Complete orthogonality ( $r_{x,z} = 0$ ) was omitted due to its potential qualitative difference from situations where interrelation was non-zero.



Measurement Level of Predictors and Criterion. Measurement level for predictors and criterion was fixed at five possible levels, representing all possible combinations of ordinal or interval continuous variables, and assuming the criterion is always continuous:

- 1) Non-Interval Y, Interval X, Interval Z;
- 2) Non-Interval Y, Non-Interval X, Interval Z;
- 3) Non-Interval Y, Non-Interval X, Non-Interval Z;
- 4) Interval Y, Non-Interval X, Interval Z; and
- 5) Interval Y, Non-Interval X, Non-Interval Z.

The assignment of this independent variable determines which of the variables are permitted to undergo monotone transformations. If a variable is Non-Interval, monotone transformations are permitted. If a variable is Interval, no transformations are permitted. It is important to note that the level of this independent variable does not change anything about the variables themselves, but only what transformations are permitted to them. The numerical values generated in the simulation have no inherent interval or non-interval status. This is only determined when they are used in the scaling algorithm.

Note that this design factor could not be completely crossed with the qualitative / quantitative nature of predictor variables, as monotone rescalings of binary X or Z variables are impossible. Thus, when Z is binary, only levels 1,2, and 4 of this factor are possible. When both X and Z are binary, only level 1 is possible. This will affect the “sample” sizes in the cells associated with these combinations.

Quantitative and Qualitative Nature of Variables. The type of variables involved was varied according to the three possible combinations of predictor variables: Continuous X and Continuous Z, Continuous X and Binary Z, and Binary X and Binary

Z. Binary variables were defined as a 50% proportion in each qualitative category.

Continuous variables were “true” continuous numerical values with precision of eight decimals.

Crossing vs. Non-crossing Interaction. The form of the interactive effects was manipulated by fixing the crossing point for X-Y regression lines across levels of Z, and the crossing point for Z-Y regression lines across levels of X. The formulas for these crossing points are  $X_c = -b_z/b_{xz}$  and  $Z_c = -b_x/b_{xz}$ , respectively (Aiken & West, 1991). Values for  $X_c$  and  $Z_c$  were set at  $-2.00$  for non-crossing interactions in the case when both variables were continuous distributions with variances of 1.0 and means of zero.<sup>6</sup> In the case of binary variables scored  $(-1, 1)$ , crossing points for non-crossing interactions were set at  $-1.1$  for both X and Z. Crossing interactions were set to cross at the mean of X  $(0.0)$ .

### Study Dependent Variables

The dependent measures in the study were selected to assess the effects of optimal monotonic rescalings of the simulated data. The nature of these transformations depends primarily on the factors outlined above. Two variables were examined: 1) the effect size differences between interaction effects assessed in the pre-transformation data and those assessed using the post-transformation data, after transformation by a MORALS

---

<sup>6</sup> Since continuous variables were random standard normal distributions, the determination of a crossing point outside a variable's range was probabilistic. For this, and other reasons, non-crossing interactions were excluded from later analyses.

algorithm; 2) the Pearson correlation between pre-transformation and post-transformation variables.

The effectiveness in reducing interaction effects in the above conditions will be indexed by the difference in effect size between the post-transformation interaction effect and the original effect size in the pre-transformed data ( $\Delta f^2$ ). In both cases, the appropriate calculation of effect size is:

$$f^2 = \frac{r^2_{Y.MI} - r^2_{Y.M}}{1 - r^2_{Y.MI}}$$

In this formula,  $r^2_{Y.MI}$  refers to the squared multiple correlation of a model including both additive effects of X and/or Z, and the product term XZ.  $r^2_M$  is the squared multiple correlation of a model including only the additive effects of X and/or Z. As can be seen, the overall effect size of the interaction depends on both components. As already described,  $r^2_M$  and  $r^2_{MI}$  will be manipulated as experimental factors in this study. Since each factor has 3 levels, 9 distinct  $f^2$  values will be present in the pre-transformation data. The two values for the post-transformation data,  $r^2_{M'}$  and  $r^2_{M'I'}$ , were evaluated by conducting a moderated multiple regression analysis on the data. These values were then used to calculate the transformed effect size,  $f_t^2$ . Since the effect size of the moderation was expected to be larger in the original data, the index ( $f_t^2 - f^2$ ) was used as a standardized indicator of attenuation.

The double cancellation axiom of conjoint measurement was to be examined using a computerized testing procedure. All possible double cancellation tests were to be

conducted, and the proportion that are true was to be used as an index of additive representability. However, due to both the lack of predictors with three levels (for qualitative predictors) and the sampled nature of predictors (for quantitative predictors), the assumptions underlying the double cancellation tests of simultaneous conjoint measurement were not met in any of the design cells. The lack of testability of these axioms does not necessarily translate into unimportance. The axioms do hold for additive relationships between quantitative variables, and thus would be relevant to the examination of such relationships in psychology. Although formal examination is not possible in this study, it is probable that the manipulated independent variables would have effects in situations which did not violate axiomatic assumptions.

### Hypotheses

Main Effects of Baseline  $R^2$ . Since  $1 - R^2$  represents error in predicting the additive model  $Y = X + Z$ , it was expected that decreases in this  $R^2$  would have effects on the double cancellation axiom tests of conjoint measurement. Recall that conjoint measurement assumes all component variables are measured without error. However, as noted previously, this dependent variable was not assessed. In contrast with the axiom tests, the MORALS algorithm has been shown by Young, de Leeuw, and Takane (1976) to perform well, even when error is present. Also, when one examines the  $f^2$  effect size equation for moderating effects,

$$f^2 = \frac{r^2_{Y.MI} - r^2_{Y.M}}{1 - r^2_{Y.MI}}$$

it is apparent that effect size calculations are more sensitive to  $\Delta R^2$  ( $R^2_{Y.MI} - R^2_{Y.M}$ ) as the  $R^2$  of the additive model ( $R^2_{Y.M}$ ) increases, i.e., as the total  $R^2$  approaches 1.00. A given  $\Delta R^2$  is a stronger effect at a high baseline  $R^2$  as opposed to a low baseline  $R^2$ . Thus, it is expected that the MORALS algorithm will be more successful at attenuating interaction effect sizes observed at low baseline  $R^2$ , and, that such transformations will be less severe.

*H<sub>1</sub>: As the  $R^2$  of the additive model decreases, the proportion of true double cancellation axioms will also decrease.*

*H<sub>2</sub>: As the  $R^2$  of the additive model decreases, the attenuation of the moderator effect size will increase.*

*H<sub>3</sub>: As the  $R^2$  of the additive model decreases, the transformed data will exhibit lesser deviation from the original variables, thus higher pre-post transformation correlation coefficients.*

Main Effects of  $\Delta R^2$ .  $\Delta R^2$  indexes the amount of additional variance predicted by the XZ product term when it is added to the regression equation. The only main effect predicted for this factor involves the correlation between pre- and post- transformation variables. Regardless of whether the transformation attenuates or completely removes an interaction, more severe transformations are expected to be required in order to affect stronger multiplicative components.

*H<sub>4</sub>: As  $\Delta R^2$  increases, the transformed data will exhibit greater deviation from the original variables, thus lower pre-post transformation correlation coefficients.*

Main Effects of Predictor Interrelation. Earlier in the thesis, it was suggested that the rescaling of predictors may attenuate an interaction effect, depending on the intercorrelation of the predictors. Recall the earlier discussion of Equations [15] and [16], where it was shown that perfect ( $r_{x,z}=1.0$ ) correlation between predictors X and Z resulted in a model identical to a quadratic model involving either predictor. In this situation, monotone transformation can render the model completely additive. It is predicted that similar effects will occur with intercorrelations less than 1.0. Specifically, it is predicted that, as the correlation between predictors increases, the MORALS algorithm will be more effective at attenuating or removing the interaction. In addition, it is also expected that as predictor intercorrelations increase, the severity of transformation attenuating an interaction effect will be reduced.

*H<sub>5</sub>: As predictor intercorrelation increases, the attenuation of the moderator effect size will increase.*

*H<sub>6</sub>: As predictor intercorrelation increases, the size of the transformation required to attenuate interaction effects will be smaller, and thus higher pre-post transformation correlation coefficients will be observed.*

Main Effects of Measurement Level. This factor reflects different levels of uncertainty about predictor and criterion measures. This uncertainty gives the MORALS

scaling algorithm more potential for rescaling, and, it is expected, a greater ability to attenuate or remove moderating effects. It is expected that more variables presumed to be non-interval (i.e., between ordinal and interval level) will reduce the severity of transformation (necessary to attenuate the interaction) in any single variable, as the additive model can be generated by changing more variables. In this sense, the transformations necessary to attenuate a moderating effect are “spread” across multiple variables, with each individual variable carrying less of the necessary transformations. This factor is not relevant to the axiom tests of conjoint measurement, since the tests assume a criterion measured at the ordinal level and predictors measured at the nominal level.

*H<sub>7</sub>: As the number of variables submitted to monotone transformation increases, the effect size of the interaction will be attenuated to a greater extent.*

*H<sub>8</sub>: As the number of variables submitted to monotone transformation increases, the transformed data will exhibit lesser deviation from the original variables, thus higher pre-post transformation correlation coefficients.*

Main Effects for Form of Interaction. As discussed and demonstrated in this thesis, non-crossing interactions can be completely removed by criterion rescalings, and crossing interactions can often be attenuated. It was also implied that predictor rescalings can potentially attenuate any form of interaction. Given this, it is expected that the attenuation of moderator effects will be greater for all non-crossing interactions. For

crossing interactions, the attenuation can never be a complete removal, so the attenuation will necessarily be lower.

*H<sub>9</sub>: Non-crossing interactions will exhibit greater attenuation than crossing interactions.*

Interaction Effects. In addition to the previously listed main effects, two interaction effects are expected: Predictor Intercorrelation  $\times$  Measurement Level, and Measurement Level  $\times$  Form of Interaction.

Predictor intercorrelation may have a greater effect when two predictors are measured at the ordinal level rather than only one, as the potential effects of intercorrelation on rescalings now applies to rescalings of both X and Z.

We also know that *any* non-crossing interaction can be removed by a criterion rescaling, so it can be said with certainty that attenuation in ordinal criterion / non-crossing interaction conditions will necessarily result in complete removal of interaction effects. However, the same is not true for crossing interactions. These interactions can be attenuated, but never removed. We have also demonstrated that predictor rescalings can eliminate an interaction effect, but only in the trivial case of perfect intercorrelation between predictors. In the range of intercorrelation used in this study, and present in most data, this will never happen. Predictor rescalings may, however, attenuate an interaction effect at many levels of predictor intercorrelation. Thus, it is expected that the main effect for form of interaction will be stronger for conditions with criterion rescalings



than for conditions with predictor rescalings, since criterion rescalings can completely remove an interaction, but predictor rescalings can only, in most cases, attenuate them.

*H<sub>10</sub>: As predictor intercorrelation increases, the attenuation of the moderator effect size will increase to a greater extent with two predictors submitted to monotone transformation than with one predictor submitted to monotone transformation.*

*H<sub>11</sub>: Differences in attenuation of non-crossing and crossing interactions will be greater when the criterion is submitted to monotone transformation than when predictors are submitted to monotone transformation.*

## METHOD

The baseline and incremental  $R^2$  values, predictor intercorrelation, and qualitative / quantitative nature of variables were all manipulated during dataset generation. The basic moderated multiple regression equation can be expressed as:

$$y = b_0 + b_1x + b_2z + b_3xz + b_e e, \quad (17)$$

where  $Y$  is a continuous variable and  $X/Z$  are continuous or binary variables. The  $b_e e$  term represents error variance uncorrelated with  $X$ ,  $Z$ , or the  $XZ$  product term.  $b_e$  controls the total  $R^2$  of the model assuming error distribution  $e$  has a mean of zero and variance of one, and is uncorrelated with variables  $X$ ,  $Z$ , or the product term  $XZ$ .

The  $R^2$  values for an additive model ( $R^2_{Y \bullet X, Z}$ ) and additive-multiplicative model ( $R^2_{Y \bullet X, Z, XZ}$ ) can be calculated from the complete correlation matrix of  $\{Y, X, Z, XZ\}$ , based on the matrix determinant formulations of McNemar (1969):

$$R^2_A = 1 - \frac{\begin{vmatrix} 1 & r_{y,x} & r_{y,z} \\ r_{y,x} & 1 & r_{x,z} \\ r_{y,z} & r_{x,z} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{x,z} \\ r_{x,z} & 1 \end{vmatrix}} \quad (18)$$

$$R^2_{AM} = 1 - \frac{\begin{vmatrix} 1 & r_{y,x} & r_{y,z} & r_{y,xz} \\ r_{y,x} & 1 & r_{x,z} & r_{x,xz} \\ r_{y,z} & r_{x,z} & 1 & r_{z,xz} \\ r_{y,xz} & r_{x,xz} & r_{z,xz} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{x,z} & r_{x,xz} \\ r_{x,z} & 1 & r_{z,xz} \\ r_{x,xz} & r_{z,xz} & 1 \end{vmatrix}} \quad (19)$$

The correlations used in the above calculations are themselves functions of the variance-covariance matrix of  $\{Y, X, Z, XZ\}$ . Under the assumption of bivariate normality<sup>7</sup> of  $X$  and  $Z$ , and given knowledge of  $E(x)$ ,  $E(z)$ ,  $\text{var}(x)$ ,  $\text{var}(z)$ ,  $\text{cov}(x,z)$ ,  $b_1$ ,  $b_2$ ,  $b_3$ , and  $b_e$ , the remaining two variances and five covariances are derived as follows:

$$\begin{aligned} \text{var}(y) &= b_1^2 \text{var}(x) + b_2^2 \text{var}(z) + 2b_1b_2 \text{cov}(x,z) \\ &+ b_3^2 [\text{var}(z)E(x)^2 + \text{var}(x)E(z)^2 + 2\text{cov}(x,z)E(x)E(z) + \text{var}(x)\text{var}(z) + \text{cov}(x,z)^2] \\ &+ 2[b_1b_3(\text{var}(x)E(z) + E(x)\text{cov}(x,z)) + b_2b_3(\text{var}(z)E(x) + E(z)\text{cov}(x,z))] + b_e^2 \end{aligned} \quad (20)$$

$$\begin{aligned} \text{var}(xz) &= \text{var}(z)E(x)^2 + \text{var}(x)E(z)^2 + 2\text{cov}(x,z)E(x)E(z) \\ &+ \text{var}(x)\text{var}(z) + \text{cov}(x,z)^2 \end{aligned} \quad (21)$$

$$\text{cov}(x, y) = b_1 \text{var}(x) + b_2 \text{cov}(x, z) + b_3 (E(z) \text{var}(x) + \text{cov}(x, z)E(x)) \quad (22)$$

$$\text{cov}(z, y) = b_2 \text{var}(z) + b_1 \text{cov}(x, z) + b_3 (E(x) \text{var}(z) + \text{cov}(x, z)E(z)) \quad (23)$$

$$\text{cov}(xz, y) = b_1 [E(z) \text{var}(x) + \text{cov}(x, z)E(x)] + b_2 [E(x) \text{var}(z) + \text{cov}(x, z)E(z)] + b_3 [E(x)^2 \text{var}(z) + 2E(x)E(z) \text{cov}(x, z) + E(z)^2 \text{var}(x) + \text{var}(x) \text{var}(z) - \text{cov}(x, z)^2] \quad (24)$$

$$\text{cov}(x, xz) = \text{var}(x)E(z) + \text{cov}(x, z)E(x) \quad (25)$$

$$\text{cov}(z, xz) = \text{var}(z)E(x) + \text{cov}(x, z)E(z) \quad (26)$$

Calculation of correlations can proceed from these variances and covariances. These correlations determine  $R^2_A$  and  $R^2_{AM}$  as described above. Substitution of these correlation formulas into the  $R^2$  determinant formulas presented earlier produces large and unwieldy expressions. In order to facilitate use of these formulas in later analyses and discussion, they were programmed using a Microsoft Excel spreadsheet. This spreadsheet was created to generate solutions for  $b_1$ ,  $b_2$ ,  $b_3$ , and  $b_e$ , given a desired  $R^2_A$ ,  $R^2_{AM}$ ,  $E(x)$ ,  $E(z)$ ,  $\text{var}(x)$ ,  $\text{var}(z)$ , and  $r_{x,z}$ . Further constraining the crossing points on X and Z axes by setting minimums and/or maximums could generate a crossing or non-crossing interaction. Solving for  $b_1$ ,  $b_2$ ,  $b_3$ , and  $b_e$  completely specified  $R^2_A$ ,  $R^2_{AM}$ , and the crossing points on X and Z.  $R^2_A$  was set to .2, .4, or .6.  $R^2_{AM}$  was set to  $R^2_A$  plus .05, .15, or .25. The qualitative vs. quantitative nature of predictors X and Z, and their intercorrelation, are the remaining factors constrained in data generation, and are now described in detail.

---

<sup>7</sup> Calculations will also hold, with slight modification of squared covariance terms, when X or Z is a binary variable coded -1/1 with  $E(x)=E(z)=0$  and  $\text{var}(x)=\text{var}(z)=1$ .

### Structure of Predictor Variables and Error Variance

The study required the generation of X and Z variables which had a given intercorrelation ( $r_{x,z}$ ) and a given qualitative or quantitative nature (binary or continuous). The intercorrelation of X and Z was set to one of three levels: .1, .3, or .5. The nature of variables X and Z was set to one of three combinations: Continuous X – Continuous Z, Continuous X – Binary Z, or Binary X – Binary Z. An error vector (E) was also generated to be correlated zero with either predictor vector. The specific procedures for generating observations in each of the above conditions are detailed below.

In the case of two continuous predictors, three vectors of standard normal deviates ( $n=10,000$ ) were generated and submitted to a principal components analysis (SAS procedure PRINCOMP). The resulting orthogonal components ( $P_1, P_2$ ) were used to construct scores for predictors X and Z using the following equations:

$$\begin{aligned} Z &= P_1 \\ X &= P_2 r_{x,z} + P_1 \sqrt{1 - r_{x,z}^2}, \end{aligned} \tag{27}$$

where  $r_{x,z}$  is the desired Pearson correlation between continuous predictors X and Z. This resulted in an exact  $r_{x,z}$  correlation, and an uncorrelated error vector based on the third principal component ( $P_3$ ). All variables (X, Z, E) were standardized to means of 0 and variances of 1. Product term XZ was constructed from the standardized scores. Note that the intercorrelation between product term XZ and X, Z, or E, is theoretically zero, as X, Z, and E are constructed to be bivariate normal, and with  $E(X)$  and  $E(Z)$  set to zero:

$$\begin{aligned}\text{cov}(XZ, X) &= \text{var}(X)E(Z) + \text{cov}(X, Z)E(X) \\ \text{cov}(XZ, X) &= 0.\end{aligned}\tag{28}$$

However, since the vectors are only *sampled* from a bivariate normal distribution, the assumption will never perfectly hold in the observed vectors. Without the assumption:

$$\begin{aligned}\text{cov}(XZ, X) &= E\left[(\Delta X)^2(\Delta Z)\right] + \text{var}(X)E(Z) + \text{cov}(X, Z)E(X) \\ \text{cov}(XZ, X) &= E\left[(\Delta X)^2(\Delta Z)\right],\end{aligned}\tag{29}$$

where  $\Delta X = X - E(X)$  and  $\Delta Z = Z - E(Z)$ . Similar equations hold for  $\text{cov}(XZ, Z)$  and  $\text{cov}(XZ, E)$ . Although this issue proves to be somewhat of a limitation to generating datasets with exact study parameters, the deviations from bivariate normality are likely slight enough to have little effect on the final outcome of analyses.

In the case of one continuous predictor and one binary predictor, the desired intercorrelation parameter is a point-biserial correlation between continuous vector  $X$  and binary vector  $Z$ . The binary predictor ( $Z$ ) was constructed to represent a qualitative binary variable, and not simply a split of an underlying quantitative variable. Each qualitative category of  $Z$  had equivalent sample sizes. Thus, it was also assumed that  $X$  is normally distributed in each of the qualitative categories. Note that this creates a bimodality in the total distribution of  $X$  across both categories of  $Z$ , the degree of which is dependent on the magnitude of  $r_{x,z}$ .

Two vectors of standard normal deviates ( $n=10,000$ ) were generated and submitted to a principal components analysis (SAS procedure PRINCOMP). Each set of

resulting orthogonal components ( $P_1, P_2$ ) were used as uncorrelated vectors  $X$  and  $E$  – one for each  $Z$  category. Error vector  $E$  was standardized to a mean of zero and variance of 1 within each  $Z$  category, which created an  $r_{z,e}$  of zero. The standardization of  $X$  was based on the desired intercorrelation between  $X$  and  $Z$ . Given the equation for the point biserial correlation,

$$r_{x,z} = \frac{\bar{X}_2 - \bar{X}_1}{S_X} \sqrt{p_1 p_2} \quad (30)$$

which, in the case of  $p_1=p_2=.5$ , and for overall  $S_X=1$  and  $\bar{X} = 0$ , reduces to

$\bar{X}_2 = \bar{X}_1 = r_{x,z}$ . The variances of  $X$  within each category of  $Z$  are determined using the formula for the variance of a mixture of two distributions ( $X_{tot}$ ), and solving for the variance of components  $X_1$  and  $X_2$ .

$$\begin{aligned} \text{var}(X_{tot}) &= \text{var}(X_1)p_1 + \text{var}(X_2)p_2 + (\bar{X}_1 - \bar{X})^2 p_1 + (\bar{X}_2 - \bar{X})^2 p_2 \\ p_1 &= p_2 = .5; \\ \text{var}(X_{tot}) &= 1; \\ \bar{X}_1 &= -r_{x,z} \\ \bar{X}_2 &= r_{x,z} \end{aligned} \quad (31)$$

$$\text{var}(X_1) = \text{var}(X_2) = 1 - r_{xz}^2$$

Categories of variable  $Z$  were set to  $-1$  and  $1$  to force a mean of zero and variance of one.

The  $X$  subset in subgroup  $Z=-1$  was standardized to a mean of  $-r_{x,z}$  and variance of  $(1-$

$r_{x,z}^2$ ). The X subset in subgroup Z=1 was standardized to a mean of  $r_{x,z}$  and variance of  $(1-r_{x,z}^2)$ .

An XZ product term was calculated based on X and Z. Due to Z's binary nature, there were no bivariate normality issues, as in the previous case with two continuous predictor variables. All intercorrelations between product term XZ, components X and Z, and error vector E, are exactly zero.

In the case of two binary predictors, an exact phi coefficient was desired. Marginal proportions for both predictors X and Z were .5, and the following proportions generate the desired correlation:

$$\begin{aligned} p_{-1,-1} &= .25r_{x,z} + .25 \\ p_{1,1} &= .25r_{x,z} + .25 \\ p_{-1,1} &= .5 - p_{1,1} \\ p_{1,-1} &= .5 - p_{1,1} \end{aligned} \tag{32}$$

X and Z values are determined according to each (p\*N, where N=10,000). Within each cell, a normal deviate is generated for the error distribution. It is standardized to a mean of zero and variance of one within each cell, forcing a zero correlation with X and Z. The product term XZ was created based on X and Z values. As in the previous case, there are no bivariate normality issues to consider. All intercorrelations between product term XZ, predictors X and Z, and error vector E are exactly zero.

As noted earlier, the formulas for the X-axis and Z-axis crossing points are  $X_c = -b_2/b_3$  and  $Z_c = -b_1/b_3$ , respectively (Aiken & West, 1991). Values for  $X_c$  and  $Z_c$  were set at -2.00 in the case when both variables were continuous distributions with variances of



1.0 and means of zero. In the case of binary variables scored (-1, 1), crossing points were set at -1.1. Crossing interactions were set to cross at the mean of X (0.0).  $X_c$  and  $Z_c$  were set as constraints in solving for  $b_1$ ,  $b_2$ ,  $b_3$ , and  $b_4$  in the Excel spreadsheet.

Because the Pearson, point biserial, and phi correlations described above are all product-moment correlations, they can be used in the previous presented formulas to obtain multiple regression parameters by means of determinant analysis. Using the  $b_1$ ,  $b_2$ ,  $b_3$ , and  $b_4$  values obtained from the Excel spreadsheet, criterion scores (Y) were calculated for the generated X and Z distributions based on Equation [17]. These datasets were then submitted to the MORALS algorithm for rescaling.

The manipulation of measurement level for predictor and criterion variables was accomplished in the MORALS algorithm itself. If a variable is ordinal (non-interval) in a given condition, the MORALS algorithm was allowed to perform monotone transformations to the variable. The algorithm was only allowed to perform identity transformations (i.e., no transformation) to interval level variables. In terms of allowable transformations within the MORALS algorithm, the five possible measurement level combinations become: 1) Monotone Y; 2) Monotone Y, X; 3) Monotone Y, X, Z; 4) Monotone X; 5) Monotone X, Z.

### Dataset Generation

The above design factors required the generation of 162 data sets - 3 ( $R^2_A = .2, .4, .6$ ) x 3 ( $\Delta R^2 = .05, .15, .25$ ) x 3 ( $r_{X,Z} = .1, .3, .5$ ) x 3 (2 continuous predictors, 2 binary predictors, 1 of each) x 2 (crossing vs. non-crossing interaction). It was discovered during data generation that some of the factor combinations in specific design cells were

mathematically impossible to create. It was impossible to create a non-crossing interaction between two continuous predictors at additive  $R^2$  levels of .2 and .4, and, even then, datasets could only be generated for an incremental  $R^2$  of .05. Similar problems were found in the case of two binary predictors and the case with one of each type, with slightly more conditions being possible in these situations. The end result is 63 of the 162 cells being impossible, leaving 99 cells for analysis. Within each of these design cells, up to five levels of the measurement level design factor can be fixed, limited by the potential qualitative nature of predictors described earlier.

For each of these cells, a single 10,000 observation dataset of variables Y, X, and Z was generated per the above descriptions. This sample size was chosen due to the behavior of the MORALS algorithm at smaller sample sizes. Pilot tests conducted by the author showed that, across several sample sizes of 300, the MORALS algorithm converged to additive  $R^2$  values varying in a roughly .07 range within the same design cell. When the sample size was increased to 10,000, the fluctuations only had a range of .01. Larger sample sizes, such as 50,000, resulted in minimal gains of convergence stability at the expense of exponential increases in processing time, primarily as the number of non-interval variables increased.

## RESULTS

### Hypothesis 1

Hypothesis 1 predicted that increasing values of additive  $R^2$  would result in larger proportions of true double cancellation axiom tests. This hypothesis could not be formally tested due to the nature of the generated data and the assumptions of the double cancellation axiom. Double cancellation tests assume levels of predictors are fixed. This assumption is violated in design conditions with continuous random predictors. Double cancellation tests also assume at least three levels of a fixed predictor. The situations examined in this study only involved fixed predictors with two levels. Thus, all conditions violated double cancellation assumptions in some manner.

### Hypothesis 2

Hypothesis 2 predicted that moderator effect sizes would be attenuated to a greater degree at lower levels of additive  $R^2$ . Degree of attenuation is indexed by  $\Delta f^2$ , the  $f^2$  effect size statistic of the post-transformation dataset minus that of the pre-transformation dataset. Negative values indicate an interaction effect is being attenuated. Mean  $\Delta f^2$ , design cell frequencies, and standard deviations for the five measurement level combinations, collapsed across all levels of predictor intercorrelation and types of predictor, are shown in Table 6.

The pattern of attenuation is opposite that predicted by Hypothesis 2. Across all combinations of monotone transformation of variables, optimal transformations resulted in greater attenuation at higher levels of additive  $R^2$ . The greatest mean attenuation (-.743) occurred when monotone transformations were permitted to predictors X and Z, and

Table 6. Mean  $\Delta R^2$  by  $R^2$  of Additive Model and Measurement Level of Variables

$R^2_A$	<u>Monotone Y</u>			<u>Monotone Y,X</u>			<u>Monotone Y,X,Z</u>			<u>Monotone X</u>			<u>Monotone X,Z</u>			Simple Mean
	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	
.2	.013	.060	27	-.037	.171	18	-.020	.223	9	-.104	.101	18	-.221	.154	9	-.074
.4	-.003	.089	27	-.154	.180	18	-.268	.268	9	-.153	.161	18	-.317	.236	9	-.179
.6	-.235	.252	27	-.482	.498	18	-.632	.615	9	-.344	.412	18	-.743	.658	9	-.487
Simple Mean	-.075			-.224			-.307			-.200			-.427			

Note:  $R^2_A$  = additive  $R^2$ ;  $\Delta R^2$  = post-transformation effect size minus pre-transformation effect size; n = # of design datasets in cell

the additive  $R^2$  was 0.6. The lowest mean attenuation (.013, actually a slight enhancement of interaction effects) occurred when monotone transformations were only permitted of the criterion Y, at an additive  $R^2$  of 0.2.

### Hypothesis 3

Hypothesis 3 predicted that the attenuations of interaction effects presented for Hypothesis 2 would be achieved with less severe monotone transformations at lower levels of additive  $R^2$ . Mean correlation coefficients, design cell frequencies, and standard deviations for the five measurement level combinations, collapsed across all levels of predictor intercorrelation and types of predictor, are shown in Table 7. The obtained pattern is opposite that predicted by Hypothesis 3. The severity of monotone transformations attenuating interaction effects generally decreased as additive  $R^2$  increased. There also appears to be a pattern related to the number of variables for which monotone transformation was permitted. As the number of non-interval variables increased (from 1 to 3), average severity of individual transformations was greater.

Despite the fact that results for Hypothesis 2 and 3 were opposite that predicted, the observed patterns are internally consistent, i.e., greater attenuation of interaction effects was achieved with less severe monotone transformations.

Table 7. Mean Correlation by R<sup>2</sup> of Additive Model and Measurement Level of Variables

R <sup>2</sup> <sub>A</sub>	<u>Monotone Y</u>		<u>Monotone Y, X</u>		<u>Monotone Y, X, Z</u>		<u>Monotone X</u>		<u>Monotone X, Z</u>		Simple Mean
	$\Gamma_{Y,IV}$	$\Gamma_{Y,IX}$	$\Gamma_{Y,IV}$	$\Gamma_{X,IX}$	$\Gamma_{Y,IV}$	$\Gamma_{X,IX}$	$\Gamma_{X,IX}$	$\Gamma_{Z,tz}$	$\Gamma_{X,tz}$	$\Gamma_{Z,tz}$	
.2	.968	.658	.658	.617	.573	.531	.942	.410	.951	.565	.691
.4	.974	.722	.722	.670	.792	.744	.965	.548	.972	.599	.775
.6	.970	.743	.743	.694	.895	.849	.977	.651	.979	.567	.814
Simple Mean	.971	.684	.684		.666		.961		.772		.760

Note: y, x, z = pre-transformation variables; ty, tx, tz = post-transformation variables

#### Hypothesis 4

Hypothesis 4 predicted that in order to attenuate interaction effects, variables would undergo transformations of greater severity at higher levels of  $\Delta R^2$ . Transformations of greater severity are indicated by lower correlations between pre- and post-transformed variables. Table 8 shows mean pre-post transformation correlations broken down by  $\Delta R^2$  and combinations of measurement level.

As can be seen in Table 8, the pattern of mean correlation across levels of incremental  $R^2$  is different for variables Y, X, and Z, depending on which of Y, X, and Z are subject to monotone transformations. Predictor X demonstrates generally decreasing correlations with larger incremental  $R^2$  values, regardless of which other variables are transformed. Criterion Y and predictor Z show no such consistent pattern.

Although no specific hypotheses were made regarding degree of attenuation at different levels of incremental  $R^2$ , the pattern of results are worthy of presentation. Mean  $\Delta f^2$ , design cell frequencies, and standard deviations for the five measurement level combinations, collapsed across all levels of predictor intercorrelation and types of predictor, are shown in Table 9. It can be seen that overall, larger pre-transformation interaction effects were attenuated to a greater extent. Further discussion of this finding will be presented in a later section.

#### Hypothesis 5

Hypothesis 5 predicted that higher levels of intercorrelation in the pre-transformation predictors would be associated with greater attenuation of interaction effects. Mean  $\Delta f^2$ , design cell frequencies, and standard deviations for the five

Table 8. Mean Correlation by Incremental  $R^2$  of Interaction Effect and Measurement Level of Variables

$\Delta R^2$	<u>Monotone Y</u>		<u>Monotone Y,X</u>		<u>Monotone Y,X,Z</u>		<u>Monotone X</u>		<u>Monotone X,Z</u>		Simple Mean
	$r_{y,ty}$	$r_{y,tx}$	$r_{y,ty}$	$r_{y,tx}$	$r_{y,ty}$	$r_{y,tx}$	$r_{x,tx}$	$r_{x,tz}$	$r_{x,tx}$	$r_{x,tz}$	
.05	.982	.795	.795	.771	.811	.793	.509	.983	.985	.583	.801
.15	.975	.691	.691	.640	.715	.665	.561	.960	.967	.572	.750
.25	.955	.637	.637	.570	.734	.665	.540	.941	.950	.577	.730
Simple Mean	.971	.684	.684		.666			.961	.772		.760

Note: y, x, z = pre-transformation variables; ty, tx, tz = post-transformation variables



Table 9. Mean  $\Delta f^2$  by Incremental  $R^2$  of Interaction Effect and Measurement Level of Variables

$\Delta R^2$	<u>Monotone Y</u>			<u>Monotone Y,X</u>			<u>Monotone Y,X,Z</u>			<u>Monotone X</u>			<u>Monotone X,Z</u>			Simple Mean
	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	
.05	-.001	.020	27	-.015	.086	18	-.039	.079	9	-.036	.028	18	-.083	.028	9	-.035
.15	-.034	.093	27	-.165	.157	18	-.179	.250	9	-.147	.116	18	-.331	.152	9	-.171
.25	-.190	.292	27	-.492	.514	18	-.702	.608	9	-.418	.379	18	-.867	.550	9	-.534

Note:  $R^2_A$  = additive  $R^2$ ;  $\Delta f^2$  = post-transformation effect size minus pre-transformation effect size; n = # of design datasets in cell

measurement level combinations at each level of predictor intercorrelation, collapsed across all levels of  $R^2_A$ ,  $\Delta R^2$ , and types of predictor, are shown in Table 10. Table 10 illustrates a pattern generally consistent with Hypothesis 5. Across all combinations of measurement level, higher correlations between predictors X and Z resulted in greater attenuation of interaction effects. However, the patterns within each category of measurement level are worthy of further discussion. As they directly relate to Hypothesis 10, these issues will be discussed in more detail later in the thesis.

### Hypothesis 6

Hypothesis 6 predicted that the greater attenuations at higher levels of predictor intercorrelation (presented in Table 10) would be achieved with less severe monotone transformations. Table 11 shows mean correlations between pre-transformation and post-transformation variables at each level of pre-transformation predictor intercorrelation and measurement level combinations, collapsed across all levels of  $R^2_A$  and  $\Delta R^2$ .

Overall, there is no clear pattern relating the severity of transformation and predictor intercorrelation, although the pattern for any given variable appears to depend on which other variables are also submitted to monotone transformation. For instance, pre-post correlations for criterion Y decrease with increasing predictor intercorrelation when it is the only variable submitted to monotone transformation, and when *both* predictors X and Z are additionally submitted to monotone transformation. When only predictor X is also submitted, pre-post correlations for criterion Y increase with predictor intercorrelation. Predictor X shows a consistent decrease in pre-post correlations as predictor intercorrelation increases, except in the case where criterion Y is also submitted

Table 10. Mean  $\Delta f^2$  by Pre-Transformation Predictor Intercorrelation and Measurement Level of Variables

$r_{X,Z}$	<u>Monotone Y</u>			<u>Monotone Y,X</u>			<u>Monotone Y,X,Z</u>			<u>Monotone X</u>			<u>Monotone X,Z</u>			Simple Mean
	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	<u>Mean</u>	<u>SD</u>	<u>n</u>	
.1	-.085	.225	27	-.195	.299	18	-.213	.454	9	-.053	.061	18	-.413	.490	9	-.192
.3	-.077	.190	27	-.230	.398	18	-.286	.522	9	-.210	.243	18	-.422	.450	9	-.245
.5	-.063	.168	27	-.247	.415	18	-.421	.458	9	-.337	.366	18	-.446	.494	9	-.303

Note:  $R^2_A$  = additive  $R^2$ ;  $\Delta f^2$  = post-transformation effect size minus pre-transformation effect size; n = # of design datasets in cell

Table 11. Mean Pre-Post Transformation Correlation by Pre-Transformation Predictor Intercorrelation and Measurement Level of

Variables

$\underline{I_{X,Z}}$	<u>Monotone Y</u>		<u>Monotone Y,X</u>		<u>Monotone Y,X,Z</u>				<u>Monotone X</u>		<u>Monotone X,Z</u>		Simple Mean
	$\underline{I_{Y,tY}}$	$\underline{I_{Y,tY}}$	$\underline{I_{Y,tY}}$	$\underline{I_{X,tX}}$	$\underline{I_{Y,tY}}$	$\underline{I_{X,tX}}$	$\underline{I_{X,tX}}$	$\underline{I_{Z,tZ}}$	$\underline{I_{X,tX}}$	$\underline{I_{X,tX}}$	$\underline{I_{Z,tZ}}$	$\underline{I_{Z,tZ}}$	
.1	.980	.661	.653	.653	.788	.782	.992	.550	.992	.992	.516	.516	.768
.2	.971	.716	.669	.669	.770	.719	.964	.582	.964	.966	.594	.594	.772
.3	.962	.746	.658	.658	.702	.622	.929	.478	.929	.944	.621	.621	.740

Note: y,x, z = pre-transformation variables; ty, tx, tz = post-transformation variables

to transformation, where no trend is apparent. Predictor Z shows a generally increasing pre-post transformation correlation when only predictor X is also transformed, but no consistent pattern when criterion Y is also transformed.

### Hypothesis 7

Hypothesis 7 predicted that the extent of interaction attenuation would vary as a function of the number of variables submitted to monotone transformation. Specifically, it was predicted that greater degrees of attenuation would occur when more variables underwent transformation. The average  $\Delta f^2$  for each category of variable transformation can be found in Table 6. Two results are apparent from examining these means.

First, there is not a simple relationship between the number of variables undergoing monotone transformation and the degree of interaction attenuation. The average attenuation for a one-variable-transformed situation (monotone Y or monotone X) is -.1375. The average attenuation for a two-variable-transformed case (monotone Y,X and monotone X,Z) is -.3255. The attenuation for the three-variable-transformed case (only monotone Y,X,Z) is -.307.

Second, it can be seen that the incremental attenuation resulting from monotone transformation of any single variable depends on which variable is considered. Specifically, monotone transformations of predictors X or Z attenuate interaction effects more than transformations to criterion Y. Table 12 lists the incremental attenuation ( $\Delta\Delta f^2$ ) for each variable, which is defined as the difference between the  $\Delta f^2$  when the variable wasn't subject to monotone transformation and the  $\Delta f^2$  after it was subject to such transformation. Table 12 shows that the average incremental attenuation of adding

Table 12. Incremental Attenuation ( $\Delta\Delta f^2$ ) by Variable and Variables Already Transformed

<u>Effects of Including Criterion Y</u>	<u>(<math>\Delta\Delta f^2</math>)</u>	<u>Mean</u>
No Transformations $\rightarrow$ Monotone (Y)	-.075	
Monotone (X) $\rightarrow$ Monotone (Y,X)	-.024	
Monotone (X,Z) $\rightarrow$ Monotone (Y,X,Z)	+.120	+.007
<u>Effects of Including Predictor X</u>		
Monotone (Y) $\rightarrow$ Monotone (Y,X)	-.149	
No Transformations $\rightarrow$ Monotone (X)	-.200	-.175
<u>Effects of Including Predictor Z</u>		
Monotone (Y,X) $\rightarrow$ Monotone (Y,X,Z)	-.083	
Monotone (X) $\rightarrow$ Monotone (X,Z)	-.227	-.155

predictors X or Z (-.1745 and -.155, respectively) was greater than that of adding criterion Y (+.007).

### Hypothesis 8

Hypothesis 8 predicted that any given variable would be transformed less violently when the number of variables being transformed was higher. The relevant mean pre-post transformation correlations can be found in Table 7. Results suggest a pattern opposite that predicted. The highest mean correlation occurred when only one variable was transformed (.971 and .961 for Monotone Y and Monotone X, respectively). The lowest mean occurred in the case when all three variables underwent monotone transformation (.666).

### Hypothesis 9

Hypothesis 9 predicted that non-crossing interactions would be attenuated to a greater extent than crossing interactions. This hypothesis, however, cannot be adequately evaluated, given the mathematical impossibility of several cells in the study design. This is apparent from examining Table 13, which contains mean attenuation, standard deviations, and cell frequencies for crossing and non-crossing interactions, collapsed across all other design factors. It would appear that crossing interactions were generally subject to greater amounts of attenuation than non-crossing interactions, contradictory to the prediction of Hypothesis 9, and much of the literature cited earlier in the thesis. There are several potential explanations for these results, some of which have important

Table 13. Mean  $\Delta f^2$  by Non-Crossing / Crossing Interaction and Measurement Level of Variables

Type	Monotone Y			Monotone Y,X			Monotone Y,X,Z			Monotone X			Monotone X,Z			Simple Mean
	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	
Non-Crossing	.182	.420	18	-.027	.041	8	-.003	.056	3	-.044	.035	8	-.075	.054	3	.007
Crossing	-.075	.193	81	-.224	.368	54	-.307	.469	27	-.200	.277	54	-.427	.460	27	-.247
Simple Mean	.054			-.126			-.155			-.122			-.251			

Note:  $R^2_A$  = additive  $R^2$ ;  $\Delta f^2$  = post-transformation effect size minus pre-transformation effect size; n = # of design datasets in cell



theoretical implications for the mathematics underlying interaction effects. These issues will be discussed in greater detail in a later section.

#### Hypothesis 10

Hypothesis 10 predicted an interaction between predictor intercorrelation and the number of variables submitted to monotone transformation, in determining the attenuation of the interaction. Specifically, it was predicted that the difference in attenuation between using two non-interval predictors (monotone X and Z) and using one non-interval predictor (monotone X) would increase as the intercorrelation between pre-transformation X and Z increased. Table 10 gives  $\Delta f^2$  for Monotone X and Monotone X,Z categories at all levels of  $r_{x,z}$ . Subtraction yields  $\Delta\Delta f^2$  values of -.360, -.212, and -.109, for  $r_{x,z}$  of .1, .3, and .5, respectively. The pattern is opposite that predicted by Hypothesis 10.

#### Hypothesis 11

Hypothesis 11 predicted an interaction between the non-crossing / crossing nature of an interaction and the measurement status of predictor and criterion variables in determining the degree of attenuated effect. It was predicted that differences in attenuation between non-crossing and crossing interactions would be greater when the criterion is subject to monotone transformation than when a predictor is subject to monotone transformation. The relevant summary information for this hypothesis is found in Table 13, in the columns for Monotone Y and Monotone X. Subtraction yields a  $\Delta\Delta f^2$  of -.257 for Monotone Y, and a  $\Delta\Delta f^2$  for -.156 for Monotone X. Although this pattern is

consistent with Hypothesis 11, it should not necessarily be considered empirical support for the hypothesis, given the previously mentioned problems generating certain cells for non-crossing interactions.

## DISCUSSION

The general goal of the study was to examine the relationships between the attenuation of interaction effects, the severity of monotonic transformations required by such attenuations, and a variety of factors commonly associated with moderated multiple regression analysis. Discussion of these relationships will be organized in two sections. The first section summarizes the results of each design factor, and offers potential explanation of observed results. The second section provides a general discussion on the impact of these results for the future study of interaction effects in psychology.

### Effects of Baseline $R^2$

Effects were predicted for the baseline additive  $R^2$  in determining both the extent of, and severity of transformations required by, attenuation of interaction effects. The obtained results were opposite that predicted. Interaction effects were attenuated to a greater extent at higher levels of additive  $R^2$ , and with less severe transformation. The reason for this inverted pattern of results is unclear, but may be related to the amount of error variance present in the criterion (Y). At higher levels of additive  $R^2$ , the ordering on Y is constrained to a greater degree by predictors X and Z than by random error. In a situation where additive  $R^2$  is very low, most of the predicted variance is carried by the XZ product term. Since the ordering on Y is minimally constrained by the ordering of X and Z (separate from their product term), transformations to X or Z will minimally increase the additive  $R^2$ . Thus, interactions in situations where additive  $R^2$  is low (.2) are more difficult to attenuate (mean  $\Delta f^2 = -.074$ ) than those at high (.6) additive  $R^2$  (mean

$\Delta f^2 = -.487$ ). The facility with which interactions were removed or attenuated in the simple examples presented earlier in this document is easily understood when one recalls that the examples all had total  $R^2$  values in the high .90's.

In the process of optimizing total  $R^2$ , the MORALS algorithm was, in part, optimizing the fit to error variance in  $Y$ . By definition, the *minimum* pre-post transformation correlation will be obtained when a variable is monotonically transformed to maximize its Pearson correlation with uncorrelated random error. At lower levels of additive  $R^2$ , the overall pre-post transformation was thereby suppressed. Consider the differences between an additive  $R^2$  of .2 and .6. The average pre-post transformation correlation at  $R^2_A = .2$  is .691, compared to the .814 at  $R^2_A = .6$ . Comparing these values for individual columns in Table 7 indicates the largest difference occurred when  $Y$ ,  $X$ , and  $Z$  were all subject to monotonic transformation ( $Y: .573 \rightarrow .895$ ,  $X: .531 \rightarrow .849$ ,  $Z: .410 \rightarrow .651$ ). Supporting this reasoning is the fact that the highest average pre-post transformation correlation across all levels of additive  $R^2$  (.971) was obtained when only  $Y$  was subject to monotonic transformation.

### Effects of Incremental $R^2$

It was predicted that the attenuation of interaction effects would require more severe transformation of variables at higher levels of incremental  $R^2$ . This prediction received general support in the case of predictor  $X$ , which displayed the predicted pattern of decreasing pre-post transformation correlations as incremental  $R^2$  increased, regardless of which other variables were permitted monotone transformation. This can be seen by examining the pre-post transformation correlations for predictor  $X$  ( $r_{x,tx}$ ) in Table 8. The

correlations for predictor Z ( $r_{z,tz}$ ) displayed no consistent ordinal relationship with incremental  $R^2$ , with an order inversion when Z was transformed along with X and Y (.509  $\rightarrow$  .561  $\rightarrow$  .540, as  $\Delta R^2$  goes from .05 to .15). Transformations to criterion Y were generally consistent with the predicted ordering, except in the case when all three variables were permitted monotone transformation (Monotone Y,X,Z), which it displayed an order inversion (.811  $\rightarrow$  .715  $\rightarrow$  .734, as  $\Delta R^2$  goes from .05 to .15). These results suggest a more complex relationship between the severity of transformation required to attenuate an interaction, and the number of variables undergoing such transformation. The exact nature of this relationship can likely be understood via a careful examination of the underlying mathematics, which is beyond the scope of the current study. It is also possible that the differing behavior of predictor Z may be due to its being a qualitative binary variable in 2/3 of the design cells, whereas predictor X was a quantitative continuous variable in 2/3 of the cells.

Although no effects were predicted relating the degree of attenuation and original interaction effects size, it was discovered that greater attenuation occurred with larger effect sizes. Table 9 shows an average  $\Delta f^2$  of -.035 at  $\Delta R^2$  of .05, and an average  $\Delta f^2$  of -.534 at  $\Delta R^2$  of .25. On one hand, this finding may be tautological, i.e., interactions with small effect sizes cannot be attenuated to a large extent. This result may also have a similar explanation as did the effects of baseline  $R^2$ . Across all levels of baseline  $R^2$ , larger values of incremental  $R^2$  implies a larger total  $R^2$ , which implies less error variance in Y. If error variance in Y were a factor working against the attenuation of interaction effects, one would see greater attenuation at higher levels of both baseline  $R^2$  and incremental  $R^2$ , which is consistent with study results.

### Effects of Predictor Intercorrelation

The study predicted effects for predictor intercorrelation in determining both the extent of, and severity of transformations required by, attenuation of interaction effects. Results shown in Table 10 generally supported the prediction that greater attenuation of interaction effects would occur at higher levels of intercorrelation. The only exception to this pattern occurred when only criterion Y was subject to monotone transformation, in which case increases in predictor intercorrelation slightly *decreased* attenuation of the interaction ( $-.085 \rightarrow -.077 \rightarrow -.063$ , as  $r_{x,z}$  goes from .1 to .5). Given that the predicted pattern occurred in other conditions in which Y was transformed (Monotone Y,X,Z and Monotone Y,X), it is possible that the effects of predictor transformation in these conditions compensated for the slight opposite effect of criterion transformation.

The explanation of these results is straightforward, and consistent with the theoretical reasoning presented earlier in the thesis. As  $r_{x,z}$  increases, the product term XZ behaves more like a quadratic function of either predictor. As discussed earlier, in the extreme case of perfect redundancy ( $r_{x,z} = 1.0$ ), the moderated regression formula reduces to a quadratic function of X, and any interaction between X and Z is completely removed by monotone transformation.

The pattern of pre-post transformation correlations presented in Table 11 is not entirely consistent with the patterns of attenuation in Table 10. In a few cases, increasing mean attenuation in Table 10 occurs with *decreasing* average pre-post correlation in Table 11. For example, the Monotone Y,X,Z case shows increases in attenuation as  $r_{x,z}$  increases ( $-.213 \rightarrow -.286 \rightarrow -.421$ ), yet both the Y and X pre-post correlation decreases

(.788 → .702 and .782 → .622). The same pattern occurs, though less dramatically, in the Monotone X case, where mean attenuation increases (-.053 → -.210 → -.337) as mean pre-post correlation decreases (.992 → .929). As noted earlier, the pattern for any given variable depends on which other variables also underwent monotone transformation. There are no theoretical reasons for expecting this pattern, and its cause remains uncertain.

### Effects of Measurement Properties of Variables

It was predicted that the extent of interaction attenuation would be higher when a greater number of variables underwent monotonic transformation. This generally proved to be true, as shown in Table 6, except for the case of monotonic transformations to predictors X and Z, which served to attenuate interaction effects (mean  $\Delta f^2 = -.427$ ) to a greater degree than monotonic transformations to Y, X, and Z (mean  $\Delta f^2 = -.307$ ). The reasons for this inversion are not clear, but it may be related to the interaction-enhancing effects of monotonic transformations to criterion Y. Table 12 illustrates that across all situations, the addition of monotonic transformation to Y had minimal effect on interaction attenuation, and, in fact, slightly enhanced the effects (+.007). The addition of transformation to predictors X and Z had similar, and much greater, effects on attenuation (-.175 and -.155, respectively). These large differences in average attenuation between predictors and criteria may be due to the fact that all interactions evaluated in the study were crossing interactions. As noted in earlier discussions, crossing interactions cannot be removed by monotonic transformation to the criterion, but can be attenuated by such transformation of predictors.

### Crossing vs. Non-Crossing Interactions

As noted earlier, the incompatibility of the selected study parameters and generation of non-crossing interactions precluded an adequate examination of this issue. However, the difficulty in generating non-crossing interactions at particular levels of baseline  $R^2$  and incremental  $R^2$  is itself an important issue. It has been lamented that the non-crossing interactions involving continuous variables typically predicted in field settings have been notoriously difficult to find (McClelland & Judd, 1993). The difficulty in finding these interactions may be due to their mathematical impossibility. In attempting to generate these interactions for this study, it was found that non-crossing interactions were more feasible (i.e., crossing points farther away from variable means) when the additive  $R^2$  was high and the incremental  $R^2$  was low. It was also possible to generate greater numbers of non-crossing interactions for binary-continuous and binary-binary predictor pairings than for continuous-continuous predictor pairings. This suggests that even if non-crossing interactions exist for the empirical constructs under study, it may not even be mathematically possible to discover them until a reduction of error variance in predicting the criterion can be established. It also suggests that the effect size of an interaction and statistical evidence of it may be related in a complex manner with the qualitative vs. quantitative nature of the predictors.

### Design Interaction Effects

The results from Table 10 cited earlier suggested that the difference in attenuation between using two non-interval predictors and using one non-interval predictor decreases



as the intercorrelation between pre-transformation X and Z increases. This result may have a simple explanation analogous to the effect of adding predictors in regression equations. Given a constant  $r_{y,z}$ , adding Z to a Y-X regression model will result in larger increases in  $R^2$  when  $r_{x,z}$  is lower. As  $r_{x,z}$  increases, X and Z have more common variance, and the addition of one to the other results in less unique variance predicting Y. The attenuation of interaction effects via monotonic transformation may work in a similar manner. If X and Z are highly correlated, monotone transformation of Z will add less incremental attenuation over the transformation of X alone. Just as X and Z share predictability of Y via their correlation in the regression situation, they may share the potential for attenuation in the transformation situation.

The interaction effects for non-crossing and crossing interactions will not be discussed, as the generated data did not support an adequate means of comparison.

### Measurement, Interaction Effects, and Psychology

This study examined the effects of measurement imprecision on the interpretation of statistical results in one particular methodology - moderated multiple regression. The simple examples presented earlier in the document suggested dire consequences of measurement imprecision on the interpretation of interaction effects in MMR. Interaction effects were severely attenuated and completely removed based on innocuous changes to the predictor and criterion variables. Although the results of the actual study suggest that these effects aren't nearly as clear-cut when realistic data are considered at realistic levels of predictability ( $R^2$ ), there are real and dramatic effects of measurement imprecision.

Consider the effect of monotone transformation to predictor X, observed in Table 6. Average reductions in interaction effect size of .104, .153, and .344, were obtained with average pre-post transformation correlations of .942, .965, and .977. This, to me, remains a striking finding. The correlations between the original data and the transformed data (representing a lack of measurement precision) are higher than those seen in virtually any reliability situation, and are representative of very slight changes in the original scales. Two researchers studying the same interaction could obtain extremely high correlations between their separate predictor measures of the same construct, and yet arrive at very different estimates of interaction effect size. Extend this situation to several researchers studying the same interaction, and you arrive at the situation commonly lamented in psychology - that interaction effects are difficult to detect and unreliable. The nature of this difficulty may be rooted in the measurement precision of predictors, and to a lesser extent, criterion variables.

Even considering the findings, several factors may have contributed to an underestimation of attenuation effect in this study. First, the continuous variables used in this study were assumed to be quantitative variables with infinite resolution, i.e., any difference represented an actual empirical difference. Thus, the number of realizable states was extremely large.<sup>8</sup> In real psychological data, this is rarely the case. Typically, what psychologists consider “continuous” variables are Likert scale items or composites of such items. Rarely do such single scales exceed 7 realizable states, and the composites almost never attain over 100 states. The implications of this difference for interpreting

---

<sup>8</sup> This fact would have made the axiom tests of conjoint measurement (if they were not excluded on theoretical grounds) an extremely computer-intensive process, as 100,000,000 (for  $n=10,000$ ) paired comparisons would be needed for each cell.

the results of this study are unclear, and future research needs to examine this issue. It would seem, at first blush, that even if the deleterious effects of transformation on interaction effects can be shown to be less severe with Likert-type items, the question of empirical representability can still be raised. That is to ask, does the limited number of realizable states in a Likert scale adequately represent realizable states in the attribute under study? The question can perhaps best be answered by further progress in both measurement theory and substantive psychological theory on qualitative vs. quantitative judgment.

Second, the primary set of study results were the result of analyzing only crossing interactions. It is known on theoretical grounds that these interactions are less susceptible to attenuation than non-crossing interactions, so overall effects of transformation across all design cells may be underestimates. Further study of the issue at levels of study parameters where non-crossing and crossing interactions can be examined in a fully-crossed design will shed light on this aspect of the problem.

Third, the MORALS algorithm used in this study only optimizes additive fit to data, and does not minimize severity of transformation to do so. Thus, many of the pre-post transformation correlations presented in this study may not be maximum, and may over-estimate the severity of transformation necessary to attenuate an interaction. There is no immediate solution to this problem, as such additional optimization criteria would somehow have to be integrated into the MORALS algorithm.<sup>9</sup>

Despite the dramatic findings described above, several of the results were in direct opposition to proposed hypotheses. As discussed earlier, much of this may involve the

---

<sup>9</sup> Based on discussion with the developer of the MORALS implementation in the SAS package, this

unclear role of error variance. In general, the message emerging from this study may be that the less we know about our criterion (in terms of overall  $R^2$ ), the *more* we are able to interpret observed interaction effects (due to our decreased ability to attenuate them via transformation). To this author, this seems an odd conclusion, despite its consistency with study findings. It would seem that the advocates of Stevens' theory, who would make interval assumptions of their measurement systems, and interpret interaction effects without scaling concerns, and those in Michell's "purist" camp, who might claim additive models more parsimonious, and perform rescalings rather than interpret multiplicative effects, in the end are fighting against the same enemy - error variance.

For the purposes of this study, error variance in Y was simply uncorrelated variance added after the effects of X, Z, and XZ were calculated. In real situations, this error variance may be due to simple unpredictability of Y, unreliability in Y, or even unreliability in predictors X or Z. The effects of predictor unreliability on detecting interaction effects are well known (Dunlap & Kemery, 1988), but the role of measurement error in the context of measurement theory are only beginning to be examined (Falmagne 1979).

The role of error variance in Y also implies a larger paradox. As theoretical models in psychology become more complex, and better at predicting human behavior, the confidence we place in the interactive effects present in such models decreases, as monotonic transformation is more able to attenuate interactions at higher  $R^2$  levels. The ultimate solution lies in the simultaneous development of both psychological theory and

psychological measurement, so that advances in understanding the relationships between psychological phenomena are accompanied by the requisite advances in measuring them.

### Practical Implications of the Study

Although the discussions above highlight the potential complex nature of several obtained results, an overall examination of the effects of each study design can provide cautions for everyday psychological research. Table 14 lists the overall extent of attenuation of crossing interaction effects at all levels of study design factors. As can be gleaned from the table, the greatest problem estimating interaction effects would occur in a situation with high additive  $R^2$ , a large observed interaction effect, high predictor intercorrelation, and two continuous non-interval predictors. Conversely, the least problem occurs at low levels of additive  $R^2$ , a small observed interaction effect, low predictor intercorrelation<sup>10</sup>, a non-interval criterion<sup>11</sup>, and two binary predictors.

This pattern of results points to a clear role of measurement imprecision in explaining the difference in detecting interactions in experimental settings and field settings. The optimal situation of orthogonal qualitative predictors virtually defines the experimental design, where stimulus control allows the random assignment of observations to factors in crossed designs to force orthogonality. The measurement status of experimental stimuli are not at issue, as they are, by definition, controlled stimuli. In the case of a binary predictor, the experimenter is simply controlling a single qualitative difference.

---

<sup>10</sup> Although orthogonal predictors weren't included in the study, we can presume it to be the optimal situation

<sup>11</sup> Optimal within the context of *some* measurement imprecision. Ideally, all variables would be interval.

Table 14. Mean  $\Delta f^2$  Values for Study Design Factors

<u>DESIGN FACTOR</u>	<u>MEAN <math>\Delta f^2</math></u>
<u>Additive <math>R^2</math></u>	
.2	-.074
.4	-.179
.6	-.487
<u>Incremental <math>R^2</math></u>	
.05	-.035
.15	-.171
.25	-.534
<u>X-Z Interrelation</u>	
.1	-.192
.3	-.245
.5	-.303
<u>Measurement Status of Variables</u>	
Non-Interval Y	-.075
Non-Interval Y,X	-.224
Non-Interval Y,X,Z	-.307
Non-Interval X	-.200
Non-Interval X,Z	-.427
<u>Qualitative/Quantitative Nature of Variables</u>	
Continuous X and Z	-.221
Continuous X, Binary Z	-.189
Binary X, Binary Z	-.133

In contrast, a field study does not have the luxury of experimental control, and therefore, in most cases, cannot force orthogonality of predictors or control the levels of predictors. These are typically observed values which naturally occur in the field setting. While the experimentalists need only fear the effects of non-interval measurement in their dependent measures, the field researchers must also concern themselves with scaling issues in random, continuous predictors.

This thesis has also examined the issue of measurement imprecision by assessing reductions in interaction effect size. An alternative approach would have been to examine changes in decisions made on *statistical* grounds. That is, the impact of measurement precision would not be problematic if only effect sizes were reduced, but if these reductions in effect size also resulted in statistical non-significance of the interaction effect. Although effect size is theoretically independent of any given statistical test, the decisions resulting from inferential methods are based on finite, often small, samples. This consideration invariably raises the issue of statistical power. Several factors related to statistical power and moderated multiple regression were discussed earlier in the thesis, but kept distinct from the focus on measurement properties of data. Extending the argument that statistical decisions be the criterion against which measurement imprecision is judged would suggest that the measurement properties of an instrument become more important as sample sizes increase, thereby making any given change in  $R^2$  more statistically significant. This logic is problematic, as the measurement properties are inherent in the instrument, and the identical instrument is used whether a sample size is large or small. Whether a given change in  $R^2$  is statistically significant or not can be solely a function of sample size and unimportant in evaluating the precision of

the measurement instrument involved. The magnitude of the change in effect size is the only relevant consideration. Regardless of its effect on statistical decision-making, measurement imprecision may still have effects on the underlying moderator effect size.

Although this study has examined how measurement imprecision may cause researchers to interpret non-existent interaction effects, it is important to note that the converse is also true, i.e., measurement imprecision may also contribute to problems in not finding interaction effects which *are* empirically present. An effect size observed when using non-interval scales may actually be an underestimate of the actual effect size, and in situations where this underestimate is great enough to result in statistical non-significance of a moderator effect, the researcher has missed a potentially important scientific finding. In this thesis, the choice was made to focus on the reduction of observed moderating effects, based on the scientific principle of parsimony. Typically, when an additive and additive-multiplicative model have equal statistical viability, scientists choose the simpler, additive model. The focus on removal or attenuation of interaction effects was also designed to critically address the interpretation of interactions with extremely small effect sizes. The difficulties associated with finding interaction effects in applied psychology may contribute to their rarity, but in no way increase the scientific value of the interactions we *do* find. Rather, both the observation and non-observation of moderator effects should be evaluated from a measurement framework. Once we establish a certain level of precision in our measurement, we may better understand which of our observed interaction effects are “real”. If weak measurement is also contributing to applied psychologists *not* finding significant moderator effects, then increasing the quality and precision of measurement can only improve the situation.



## APPENDICES

## APPENDIX A: MORALS Algorithm

The MORALS algorithm maximizes the canonical correlation coefficient between two sets of variables,  $X$  and  $Y$ , by transforming the variables according to specified constraints. The specification of the model follows below. More detail can be found in Young, de Leeuw, & Takane (1976).

1. Let  $X$  be a matrix of  $k$  observations of  $n$  variables.  
Let  $Y$  be a matrix of  $k$  observations of  $m$  variables.
2. Each  $x_i$  and  $y_j$  assumed to be measured at a specified measurement level (nominal, ordinal, interval, ratio).
3. Two parameter vectors,  $\alpha$  and  $\beta$  are defined to have  $n$  and  $m$  elements, respectively.
4. Two matrices,  $X^*$  and  $Y^*$ , are defined to have the same dimensions as  $X$  and  $Y$ .
5. The columns of the  $X^*$  and  $Y^*$  matrices,  $x_i^*$  and  $y_j^*$ , have two properties. 1) They are defined at the interval level of measurement. 2) They are related to the corresponding columns in  $X$  and  $Y$ ,  $x_i$  and  $y_j$ , by transformations permissible of the specific variable. So:

$$x_i^* = \mathfrak{I}_i(x_i)$$

$$y_j^* = \mathfrak{I}_j(y_j)$$

6.  $\mathfrak{I}_i$  and  $\mathfrak{I}_j$  above represent *measurement transformations* of observed variables  $X$  and  $Y$ .

The goal of the algorithm is to find transformations  $\mathfrak{I}_i$  and  $\mathfrak{I}_j$  and regression weights  $\alpha$  and  $\beta$ , so the canonical correlation between  $\mathbf{X}^*$  and  $\mathbf{Y}^*$  is maximized. This is equivalent to minimizing the sum of squared differences between composite variables  $\mathbf{a}$  and  $\mathbf{b}$ , defined as:

$$\mathbf{a} = \mathbf{X}^* \alpha$$

$$\mathbf{b} = \mathbf{Y}^* \beta$$

subject to minimization criteria:  $\lambda^2 = (\mathbf{a} - \mathbf{b})' (\mathbf{a} - \mathbf{b})$

7. The minimization is constrained by allowable forms of the  $\mathfrak{I}$  functions. These depend on the level of measurement of the variable in question and the processes by which the distributions are generated. The constraints on  $\mathfrak{I}$  fall into three types: order ( $\mathfrak{I}^0$ ), linear ( $\mathfrak{I}^1$ ), and polynomial ( $\mathfrak{I}^p$ ):

$$\mathfrak{I}^0: (x_{ai} < x_{bi}) \rightarrow (x_{ai}^* \leq x_{bi}^*)$$

$$\mathfrak{I}^1: x_{ai}^* = \delta_0 + \delta_1 x_{ai}$$

$$\mathfrak{I}^p: x_{ai}^* = \sum_{q=0}^p \delta_p x_{ai}^q$$

APPENDIX B1:  $r^2$  Values for Crossing Interactions with Two Continuous Predictors

$R^2_A$	$\Delta R^2$	$r_{X,Z}$	$r^2_{\underline{1}}$	$r^2_{\underline{2}}$	$r^2_{\underline{3}}$	$r^2_{\underline{4}}$	$r^2_{\underline{5}}$
0.20	0.05	0.10	0.08	0.08	0.01	0.06	0.02
0.20	0.05	0.30	0.08	0.06	0.22	0.04	0.01
0.20	0.05	0.50	0.09	0.36	0.00	0.02	0.00
0.20	0.15	0.10	0.24	0.01	0.53	0.16	0.02
0.20	0.15	0.30	0.28	0.31	0.24	0.13	0.02
0.20	0.15	0.50	0.32	0.33	0.01	0.08	0.03
0.20	0.25	0.10	0.54	0.51	0.60	0.41	0.05
0.20	0.25	0.30	0.57	0.64	0.43	0.23	0.05
0.20	0.25	0.50	0.60	0.54	0.00	0.12	0.04
0.40	0.05	0.10	0.09	0.09	0.05	0.08	0.02
0.40	0.05	0.30	0.10	0.08	0.00	0.05	0.01
0.40	0.05	0.50	0.12	0.06	0.09	0.03	0.01
0.40	0.15	0.10	0.33	0.36	0.33	0.29	0.07
0.40	0.15	0.30	0.35	0.07	0.05	0.19	0.04
0.40	0.15	0.50	0.36	0.26	0.01	0.10	0.03
0.40	0.25	0.10	0.72	0.68	0.10	0.65	0.21
0.40	0.25	0.30	0.77	0.62	0.34	0.45	0.08
0.40	0.25	0.50	0.77	0.61	0.00	0.18	0.06
0.60	0.05	0.10	0.13	0.13	0.01	0.13	0.03
0.60	0.05	0.30	0.13	0.15	0.10	0.08	0.02
0.60	0.05	0.50	0.13	0.07	0.03	0.05	0.01
0.60	0.15	0.10	0.52	0.45	0.23	0.58	0.09
0.60	0.15	0.30	0.51	0.43	0.25	0.37	0.08
0.60	0.15	0.50	0.48	0.30	0.13	0.18	0.03
0.60	0.25	0.10	1.17	0.99	0.56	1.53	0.11
0.60	0.25	0.30	1.12	0.77	0.08	0.82	0.18
0.60	0.25	0.50	1.12	0.57	0.21	0.40	0.05

Note:  $r^2$  subscripts - 1=Montone Y; 2=Monotone Y,X; 3=Monotone Y,X,Z; 4=Monotone X; 5=Monotone X,Z;

APPENDIX B2:  $f^2$  Values for Crossing Interactions with One Continuous and One Binary Predictor

$R^2_A$	$\Delta R^2$	$r_{X,Z}$	$f^2_{\underline{1}}$	$f^2_{\underline{2}}$	$f^2_{\underline{3}}$	$f^2_{\underline{4}}$	$f^2_{\underline{5}}$
0.20	0.05	0.10	0.07	0.07	n/a	0.06	n/a
0.20	0.05	0.30	0.07	0.04	n/a	0.04	n/a
0.20	0.05	0.50	0.08	0.02	n/a	0.02	n/a
0.20	0.15	0.10	0.26	0.20	n/a	0.21	n/a
0.20	0.15	0.30	0.26	0.01	n/a	0.15	n/a
0.20	0.15	0.50	0.25	0.06	n/a	0.08	n/a
0.20	0.25	0.10	0.48	0.35	n/a	0.40	n/a
0.20	0.25	0.30	0.46	0.18	n/a	0.24	n/a
0.20	0.25	0.50	0.41	0.06	n/a	0.16	n/a
0.40	0.05	0.10	0.10	0.01	n/a	0.08	n/a
0.40	0.05	0.30	0.10	0.07	n/a	0.06	n/a
0.40	0.05	0.50	0.11	0.00	n/a	0.03	n/a
0.40	0.15	0.10	0.37	0.15	n/a	0.29	n/a
0.40	0.15	0.30	0.38	0.20	n/a	0.21	n/a
0.40	0.15	0.50	0.41	0.12	n/a	0.12	n/a
0.40	0.25	0.10	0.84	0.29	n/a	0.61	n/a
0.40	0.25	0.30	0.81	0.16	n/a	0.41	n/a
0.40	0.25	0.50	0.78	0.18	n/a	0.22	n/a
0.60	0.05	0.10	0.13	0.05	n/a	0.13	n/a
0.60	0.05	0.30	0.13	0.09	n/a	0.09	n/a
0.60	0.05	0.50	0.16	0.06	n/a	0.05	n/a
0.60	0.15	0.10	0.48	0.06	n/a	0.55	n/a
0.60	0.15	0.30	0.51	0.37	n/a	0.37	n/a
0.60	0.15	0.50	0.57	0.24	n/a	0.19	n/a
0.60	0.25	0.10	1.07	0.64	n/a	1.46	n/a
0.60	0.25	0.30	1.09	0.19	n/a	0.86	n/a
0.60	0.25	0.50	1.22	0.29	n/a	0.47	n/a

Note:  $f^2$  subscripts - 1=Montone Y; 2=Monotone Y,X; 4=Monotone X; subscripts 3 and 5 not applicable when Z is binary

APPENDIX B3:  $f^2$  Values for Crossing Interactions with Two Binary Predictors

$R^2_A$	$\Delta R^2$	$f^2_{XZ}$	$f^2_{\underline{1}}$	$f^2_{\underline{2}}$	$f^2_{\underline{3}}$	$f^2_{\underline{4}}$	$f^2_{\underline{5}}$
0.20	0.05	0.10	0.07	n/a	n/a	n/a	n/a
0.20	0.05	0.30	0.07	n/a	n/a	n/a	n/a
0.20	0.05	0.50	0.07	n/a	n/a	n/a	n/a
0.20	0.15	0.10	0.23	n/a	n/a	n/a	n/a
0.20	0.15	0.30	0.23	n/a	n/a	n/a	n/a
0.20	0.15	0.50	0.23	n/a	n/a	n/a	n/a
0.20	0.25	0.10	0.45	n/a	n/a	n/a	n/a
0.20	0.25	0.30	0.45	n/a	n/a	n/a	n/a
0.20	0.25	0.50	0.45	n/a	n/a	n/a	n/a
0.40	0.05	0.10	0.09	n/a	n/a	n/a	n/a
0.40	0.05	0.30	0.09	n/a	n/a	n/a	n/a
0.40	0.05	0.50	0.09	n/a	n/a	n/a	n/a
0.40	0.15	0.10	0.33	n/a	n/a	n/a	n/a
0.40	0.15	0.30	0.33	n/a	n/a	n/a	n/a
0.40	0.15	0.50	0.33	n/a	n/a	n/a	n/a
0.40	0.25	0.10	0.71	n/a	n/a	n/a	n/a
0.40	0.25	0.30	0.71	n/a	n/a	n/a	n/a
0.40	0.25	0.50	0.71	n/a	n/a	n/a	n/a
0.60	0.05	0.10	0.14	n/a	n/a	n/a	n/a
0.60	0.05	0.30	0.14	n/a	n/a	n/a	n/a
0.60	0.05	0.50	0.14	n/a	n/a	n/a	n/a
0.60	0.15	0.10	0.60	n/a	n/a	n/a	n/a
0.60	0.15	0.30	0.60	n/a	n/a	n/a	n/a
0.60	0.15	0.50	0.60	n/a	n/a	n/a	n/a
0.60	0.25	0.10	1.67	n/a	n/a	n/a	n/a
0.60	0.25	0.30	1.67	n/a	n/a	n/a	n/a
0.60	0.25	0.50	1.67	n/a	n/a	n/a	n/a

Note:  $f^2$  subscripts - 1=Montone Y; remaining subscripts not applicable when both predictors are binary

APPENDIX B4: Pre-Post Correlation Coefficients for Crossing Interactions with Two Continuous Predictors

$R^2_A$	$\Delta R^2$	$r_{x,z}$	$r^2$	Monotone Y		Monotone Y X		Monotone Y X Z		Monotone X		Monotone X Z	
				$I_{IV,Y}$	$I_{IV,X}$	$I_{IV,Y}$	$I_{IX,X}$	$I_{IV,Y}$	$I_{IX,X}$	$I_{IX,X}$	$I_{IX,X}$	$I_{IX,X}$	$I_{IX,Z}$
0.20	0.05	0.10	0.07	0.99	0.96	0.96	0.96	0.97	0.97	0.99	0.99	0.99	0.62
0.20	0.05	0.30	0.07	0.98	0.85	0.83	0.83	0.77	0.77	0.98	0.98	0.98	0.60
0.20	0.05	0.50	0.06	0.97	0.57	0.52	0.52	0.59	0.57	0.95	0.96	0.96	0.61
0.20	0.15	0.10	0.21	0.98	0.17	0.14	0.14	0.17	0.13	0.99	0.99	0.99	0.42
0.20	0.15	0.30	0.22	0.98	0.60	0.53	0.53	0.53	0.47	0.95	0.95	0.95	0.59
0.20	0.15	0.50	0.23	0.96	0.65	0.54	0.54	0.68	0.61	0.90	0.92	0.92	0.62
0.20	0.25	0.10	0.47	0.98	0.47	0.44	0.44	0.33	0.32	0.98	0.98	0.98	0.44
0.20	0.25	0.30	0.45	0.97	0.63	0.53	0.53	0.64	0.54	0.91	0.91	0.91	0.58
0.20	0.25	0.50	0.45	0.96	0.45	0.33	0.33	0.48	0.39	0.86	0.88	0.88	0.62
0.40	0.05	0.10	0.08	1.00	0.95	0.95	0.95	0.95	0.95	1.00	1.00	1.00	0.54
0.40	0.05	0.30	0.08	1.00	0.86	0.84	0.84	0.89	0.87	0.99	0.99	0.99	0.60
0.40	0.05	0.50	0.10	0.99	0.64	0.57	0.57	0.46	0.41	0.97	0.97	0.97	0.61
0.40	0.15	0.10	0.31	0.99	0.92	0.92	0.92	0.88	0.88	0.99	0.99	0.99	0.55
0.40	0.15	0.30	0.32	0.99	0.56	0.48	0.48	0.64	0.57	0.97	0.97	0.97	0.61
0.40	0.15	0.50	0.31	0.99	0.80	0.69	0.69	0.84	0.76	0.94	0.95	0.95	0.62
0.40	0.25	0.10	0.71	1.00	0.86	0.86	0.86	0.97	0.96	0.99	0.99	0.99	0.64
0.40	0.25	0.30	0.74	0.99	0.84	0.76	0.76	0.80	0.74	0.96	0.96	0.96	0.59
0.40	0.25	0.50	0.72	0.99	0.66	0.48	0.48	0.68	0.55	0.90	0.92	0.92	0.63
0.60	0.05	0.10	0.13	1.00	0.99	0.99	0.99	1.00	0.99	1.00	1.00	1.00	0.50
0.60	0.05	0.30	0.13	1.00	0.82	0.78	0.78	0.81	0.78	0.99	0.99	0.99	0.56
0.60	0.05	0.50	0.14	1.00	0.90	0.85	0.85	0.86	0.82	0.98	0.99	0.99	0.61
0.60	0.15	0.10	0.62	1.00	0.97	0.97	0.97	0.96	0.96	1.00	1.00	1.00	0.49
0.60	0.15	0.30	0.60	1.00	0.90	0.85	0.85	0.88	0.83	0.98	0.98	0.98	0.61
0.60	0.15	0.50	0.56	0.99	0.88	0.78	0.78	0.86	0.76	0.96	0.96	0.96	0.64
0.60	0.25	0.10	1.74	0.99	0.88	0.89	0.89	0.87	0.88	0.99	0.99	0.99	0.46
0.60	0.25	0.30	1.67	0.99	0.88	0.81	0.81	0.96	0.90	0.97	0.97	0.97	0.61
0.60	0.25	0.50	1.70	0.99	0.90	0.76	0.76	0.86	0.72	0.93	0.93	0.93	0.62

Note: y, x, z denote pre-transformation variables; ty, tx, tz denote post-transformation variables; tabled correlations of 1.00 are not exact

**APPENDIX B5: Pre-Post Correlation Coefficients for Crossing Interactions with One Continuous and One Binary Predictor**

$R^2_A$	$\Delta R^2$	<u>Monotone Y</u>		<u>Monotone Y,X</u>		<u>Monotone Y,X,Z</u>		<u>Monotone X</u>		<u>Monotone X,Z</u>	
		$I_{Y,Y}$	$r^2$	$I_{Y,Y}$	$I_{Y,X}$	$I_{Y,Y}$	$I_{Y,X}$	$I_{Y,Z}$	$I_{X,X}$	$I_{X,Z}$	$I_{Z,Z}$
0.20	0.05	0.99	0.07	0.98	0.98	n/a	n/a	n/a	0.99	n/a	n/a
0.20	0.05	0.98	0.07	0.92	0.91	n/a	n/a	n/a	0.98	n/a	n/a
0.20	0.05	0.97	0.07	0.91	0.88	n/a	n/a	n/a	0.95	n/a	n/a
0.20	0.15	0.97	0.23	0.87	0.89	n/a	n/a	n/a	0.99	n/a	n/a
0.20	0.15	0.96	0.23	0.14	0.12	n/a	n/a	n/a	0.94	n/a	n/a
0.20	0.15	0.94	0.23	0.83	0.76	n/a	n/a	n/a	0.88	n/a	n/a
0.20	0.25	0.94	0.45	0.50	0.52	n/a	n/a	n/a	0.98	n/a	n/a
0.20	0.25	0.92	0.45	0.63	0.59	n/a	n/a	n/a	0.91	n/a	n/a
0.20	0.25	0.90	0.45	0.71	0.63	n/a	n/a	n/a	0.85	n/a	n/a
0.40	0.05	1.00	0.09	0.46	0.45	n/a	n/a	n/a	0.99	n/a	n/a
0.40	0.05	1.00	0.09	0.95	0.94	n/a	n/a	n/a	0.99	n/a	n/a
0.40	0.05	0.99	0.09	0.40	0.37	n/a	n/a	n/a	0.97	n/a	n/a
0.40	0.15	0.99	0.33	0.55	0.53	n/a	n/a	n/a	0.99	n/a	n/a
0.40	0.15	0.99	0.33	0.78	0.73	n/a	n/a	n/a	0.97	n/a	n/a
0.40	0.15	0.98	0.33	0.84	0.76	n/a	n/a	n/a	0.93	n/a	n/a
0.40	0.25	0.99	0.71	0.55	0.52	n/a	n/a	n/a	0.99	n/a	n/a
0.40	0.25	0.97	0.71	0.57	0.50	n/a	n/a	n/a	0.95	n/a	n/a
0.40	0.25	0.95	0.71	0.81	0.71	n/a	n/a	n/a	0.89	n/a	n/a
0.60	0.05	1.00	0.14	0.28	0.26	n/a	n/a	n/a	1.00	n/a	n/a
0.60	0.05	1.00	0.14	0.94	0.92	n/a	n/a	n/a	0.99	n/a	n/a
0.60	0.05	1.00	0.14	0.92	0.88	n/a	n/a	n/a	0.98	n/a	n/a
0.60	0.15	1.00	0.60	0.24	0.23	n/a	n/a	n/a	1.00	n/a	n/a
0.60	0.15	0.99	0.60	0.87	0.82	n/a	n/a	n/a	0.98	n/a	n/a
0.60	0.15	0.99	0.60	0.87	0.77	n/a	n/a	n/a	0.95	n/a	n/a
0.60	0.25	0.99	1.67	0.28	0.26	n/a	n/a	n/a	0.99	n/a	n/a
0.60	0.25	0.98	1.67	0.15	0.11	n/a	n/a	n/a	0.97	n/a	n/a
0.60	0.25	0.97	1.67	0.70	0.57	n/a	n/a	n/a	0.93	n/a	n/a

Note: y,x, z denote pre-transformation variables; ty, tx, tz denote post-transformation variables; tabled correlations of 1.00 are not exact



APPENDIX B6: Pre-Post Correlation Coefficients for Crossing Interactions with Two Binary Predictors

$R^2_A$	$\Delta R^2$	Monotone Y		Monotone Y,X		Monotone Y,X,Z		Monotone X		Monotone X,Z	
		$I_{Y,Y}$	$f^2$	$I_{Y,Y}$	$I_{Y,X}$	$I_{Y,Y}$	$I_{Y,X}$	$I_{X,X}$	$I_{X,Z}$	$I_{X,X}$	$I_{X,Z}$
0.20	0.05	0.99	0.07	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.20	0.05	0.98	0.07	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.20	0.05	0.99	0.07	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.20	0.15	0.99	0.23	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.20	0.15	0.99	0.23	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.20	0.15	0.98	0.23	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.20	0.25	0.99	0.45	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.20	0.25	0.96	0.45	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.20	0.25	0.92	0.45	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.40	0.05	0.97	0.09	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.40	0.05	0.97	0.09	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.40	0.05	0.97	0.09	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.40	0.15	0.97	0.33	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.40	0.15	0.96	0.33	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.40	0.15	0.95	0.33	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.40	0.25	0.95	0.71	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.40	0.25	0.91	0.71	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.40	0.25	0.89	0.71	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.60	0.05	0.94	0.14	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.60	0.05	0.94	0.14	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.60	0.05	0.94	0.14	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.60	0.15	0.94	0.60	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.60	0.15	0.93	0.60	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.60	0.15	0.92	0.60	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.60	0.25	0.93	1.67	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.60	0.25	0.90	1.67	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.60	0.25	0.89	1.67	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Note: y,x, z denote pre-transformation variables; ty, tx, tz denote post-transformation variables; tabled correlations of 1.00 are not exact

## REFERENCES

- Achenbach, T. M. (1978). Research in developmental psychology: Concepts, strategies, methods. New York: Free Press, 1978.
- Aguinas, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. Journal of Applied Psychology, 82, 192-206.
- Aiken L. S., & West, S. G. (1991). Multiple regression: Testing and interpreting interactions. Newbury Park, CA: Sage.
- Alexander, R. A., & DeShon, R. P. (1994). Effect of error variance heterogeneity on the power of tests for regression slope differences. Psychological Bulletin, 115, 308-314.
- Althauser, R. P. (1971). Multicollinearity and non-additive regression models. In H. M. Blalock, Jr. (Ed.), Causal models in the social sciences. Chicago: Aldine-Atherton.
- Anderson, N. H. (1961). Scales and statistics: Parametric and nonparametric. Psychological Bulletin, 58, 305-316.
- Arnold, H. J., & Evans, M. G. (1979). Testing multiplicative models does *not* require ration scales. Organizational Behavior and Human Performance, 24, 41-59.
- Baker, B. O., Hardyk, C. D., & Petrinovich, L. F. (1966). Weak measurement vs. strong statistics: An empirical critique of S. S. Stevens' proscriptions on statistics. Educational and Psychological Measurement, 26, 291-309.
- Birnbaum, M. H. (1973). The devil rides again: Correlations as an index of fit. Psychological Bulletin, 79, 239-242.

Birnbaum, M.H. (1974). Reply to the devil's advocates: Don't confound model testing with measurement. Psychological Bulletin, 81, 854-859.

Bridgman, P. (1922). Dimensional analysis. New Haven: Yale University Press.

Burke, C. J. (1953). Additive scales and statistics. Psychological Review, 60, 73-75.

Busemeyer, J. R. (1980). Importance of measurement theory, error theory, and experimental design for testing the significance of interactions. Psychological Bulletin, 88, 237-244.

Busemeyer, J. R., & Jones, L. E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. Psychological Bulletin, 93, 549-562.

Campbell, N. R. (1920). Physics, the elements. Cambridge: Cambridge University Press.

Campbell, N. R. (1928). An account of the principles of measurement and calculation. London: Longmans, Green.

Campbell, B. A. & Masterson, F. A. (1969). Psychophysics of punishment. In B. A. Campbell & R. M. Church (Eds.), Punishment and Aversive Behavior. New York: Appleton-Century-Crofts, 3-42.

Champoux, J. E., & Peters, W. S. (1987). Form, effect size, and power in moderated regression analysis. Journal of Occupational Psychology, 60, 243-255.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. Journal of Educational Measurement, 5, 115-124.

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. Psychological Science, 3, 186-190.

Cohen, J. (1978). Partialled products *are* interactions; partialled powers *are* curve components. Psychological Bulletin, 85, 858-866.

Cohen, J., & Cohen, P. (1983). Applied multiple regression / correlational analysis for the behavioral sciences. Hillsdale, N.J.: Erlbaum.

Cronbach, L. J. (1987) Statistical tests for moderator variables: Flaws in analyses recently proposed. Psychological Bulletin, 102, 414-417.

de Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. Psychometrika, 41, 471-503.

Dunlap, W. P., & Kemery, E. R. (1988). Effects of predictor intercorrelation and reliabilities on moderated multiple regression. Organizational Behavior and Human Decision Processes, 41, 248-258.

Embretson, S. E. (1996). Item response theory models and spurious interaction effects in factorial ANOVA designs. Applied Psychological Measurement, 20, 201-212.

Evans, M. G. (1985). A monte carlo study of the effects of correlated method variance in moderated multiple regression analysis. Organizational Behavior and Human Decision Processes, 36, 305-323.

Falmagne, J. C., Iverson, G., & Marcovici, S. (1979). Binaural "loudness" summation: Probabilistic theory and data. Psychological Review, 86, 25-43.

Ferguson, A., Meyers, C. S. (Vice Chairman), Bartlett, R. J. (Secretary), Banister, H., Bartlett, F. C., Brown, W., Campbell, N. R., Craik, K. J. W., Drever, J., Guild, J.,

Houstoun, R. A., Irwin, J. O., Kaye, G. W. C., Philpott, S. J. F., Richardson, L. F., Shaxby, J. H., Smith, T., Thouless, R. H., & Tucker, W. S. (1940). Quantitative estimates of sensory events. The advancement of science. Report of the British Association for the Advancement of Science, 2, 331-349.

Gaito, J. (1959). Nonparametric methods in psychological research. Psychological Reports, 5, 115-125.

Gaito, J. (1960). Scale classification and statistics. Psychological Review, 67, 277-278.

Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. Psychological Bulletin, 87, 564-567.

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

Gregoire, T. G., & Driver, B. L. (1987). Analysis of ordinal data to detect population differences. Psychological Bulletin, 101, 159-165.

Gregory (1996). Psychological Testing: History, Principles, and Applications. Allyn & Bacon.

Helmholtz, H. V. (1887). Numbering and measuring from an epistemological viewpoint. (Reprinted in Hermann von Helmholtz: Epistemological writings, P. Hertz & M. Schlick (Eds.), Boston Studies in the philosophy of science, 37, 72-113.) Dordrecht-Holland, Reidel, 1977.

Holder, O. (1901). Die axiome der qualitat und die lehre vom mass. Berichte der Sachsischen Gesellschaft der Wissenschaften, Mathematische-Physicke Klass, 53, 1-64.

Jensen, A. R. (1974). Cumulative deficit: A testable hypothesis? Developmental Psychology, 10, 996-1019.

Jensen, A. R. (1980). Bias in mental testing. New York: Free Press.

Johnson, H. M. (1936). Pseudo-mathematics in the mental and social sciences. American Journal of Psychology 48, 342-351.

Kanfer, R. & Ackerman, P. (1989). Motivation and cognitive abilities: An integrative/aptitude - treatment interaction approach to skill acquisition. Journal of Applied Psychology, 74, 657-690.

Krantz, D. H., & Tversky, A. (1971). Conjoint-measurement analysis of composition rules in psychology. Psychological Review, 78, 151-169.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A.. (1971). Foundations of measurement: Vol. I. Additive and polynomial representations. New York: Academic Press.

Levelt, W. J. M., Riemersma, J. B., & Bunt, A. A. (1971). Binaural additivity of loudness. Technical Report NR: HB-71-70EX, R. U. Groningen, Netherlands, Heymans Bulletins Psychologische Instituten.

Loftus, G. A. (1978). On the interpretation of interactions. Memory & Cognition, 6, 312-319.

Lord, F. M. (1953). On the statistical treatment of football numbers. American Psychologist, 8, 750-751.

Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious “moderator effects”: Illustrated substantively with the hypothesized (“synergistic”) relation between spatial and mathematical ability. Psychological Bulletin, 107, 385-393.

Luce, R. D. & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. Journal of Mathematical Psychology, 1, 1-27.

Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). Foundations of measurement: Vol. III. Representation, axiomatization, and invariance. San Diego: Academic Press.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. Psychological Bulletin, 114, 376-390.

McGregor, D. (1935). Scientific measurement and psychology. Psychological Review, 42, 246-266.

Michell, J. (1990). An introduction to the logic of psychological measurement., Hillsdale, N.J.: Erlbaum.

Morris, J. H., Sherman, J., & Mansfield, E. R. (1986). Failures to detect moderating effects with ordinary least squares moderated-regression: Some reasons and a remedy. Psychological Bulletin, 99, 282-288.

Nickerson, C. A., & McClelland, G. H. (1984). Scaling distortion in numerical conjoint measurement. Applied Psychological Measurement, 8, 183-198.

Perline, R., Wright, B. D., & Wainer, H. (1979). The Rasch model as additive conjoint measurement. Applied Psychological Measurement, 3, 237-255.

Pierce, J. L., Gardner, D. G., Dunham, R. B., Cummings, L. L. (1993). Moderation by organization-based self-esteem of role condition-employee response relationships. Academy of Management Journal, 36, 271-288.

Rasmussen, J. L. (1989). Analysis of Likert-scale data: A reinterpretation of Gregoire and Driver. Psychological Bulletin, 105, 167-170.

Roussas, G. G. (1973). A first course in mathematical statistics. Reading, Mass: Addison-Wesley.

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

Schmidt, F. L. (1973). Implications of a measurement problem for expectancy theory research. Organizational Behavior and Human Performance, 10, 243-251.

Scott, D. (1964). Measurement models and linear inequalities. Journal of Mathematical Psychology, 1, 233-247.

Smith, B. O. (1938). Logical aspects of educational measurement. New York: Columbia University Press.

Sockloff, A. L. (1976). The analysis of nonlinearity via linear regression with polynomial and product variables: An examination. Review of Educational Research, 46, 267-291.

Stevens, S. S. (1946). On the theory of scales of measurement. Science, 103, 667-680.

Stevens, S. S., & Davis, H. (1938). Hearing: Its psychology and physiology. New York: Wiley.

Stine, W. W. (1989). Meaningful inference: The role of measurement in statistics. Psychological Bulletin, 105, 147-155.

Stone-Romero, E. F., & Anderson, L. E. (1994). Relative power of moderated multiple regression and the comparison of subgroup correlation coefficients for detecting moderating effects. Journal of Applied Psychology, 79, 354-359.



Thomas, H. (1982). IQ, interval scales, and normal distributions. Psychological Bulletin, 91, 198-202.

Townsend, J. T., & Ashby, F. G. (1984). Measurement scales and statistics: the misconception misconceived. Psychological Bulletin, 96, 394-401.

Vroom, V. H. (1964). Work and motivation. New York: Wiley.

Wise, S. L., Peters, L. H., & O'Connor, E. J. (1984). Identifying moderator variables using multiple regression: A reply to Darrow and Kahl. Journal of Management, 10, 227-233.

Young, F. W., de Leeuw, J., & Takane, Y. (1976). Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. Psychometrika, 41, 505-529.

Yuan, P. T. (1933). On the logarithmic frequency distribution and semilogarithmic correlation surface. Annals of Mathematical Statistics, 4, 30-74.

Zedeck, S. (1971). Problems with the use of "moderator" variables. Psychological Bulletin, 76, 295-310.

Zumbo, B. D. & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement?, Canadian Psychology, 34, 390-400.

MICHIGAN STATE UNIV. LIBRARIES



31293018017032