IMPLICATIONS OF USING LIKERT DATA

IN MULTIPLE REGRESSION ANALYSIS

By

CHARLES OCHIENG OWUOR

M.A. (Ed). University of Ottawa (1992)

B.ED. (Sc). University of Nairobi (1980)

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF EDUCATIONAL AND COUNSELING PSYCHOLOGY AND

SPECIAL EDUCATION

We accept this thesis as conforming to

the required standard

Thesis Supervisor: Dr.Bruno. D. Zumbo...

Chair: Dr. Heather McKay..............

University Examiners: Dr. Peter Crocker......

Dr. Kim Schonert-Reichl ....

Examination Committee: Dr. Beth Haverkamp..

Examination Committee: Dr. Anita Hubley.

THE UNIVERSITY OF BRITISH COLUMBIA

July 2001

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of EDUCATIONAL PSYCHOLOGY, COUNSELING & SPECIAL EDUCATION.

The University of British Columbia
Vancouver, Canada

Date JULY 26th/2001.

DE-6 (2/88)

ABSTRACT

Many of the measures obtained in educational research are Likert-type responses on questionnaires. These Likert-type variables are sometimes used in ordinary least-squares regression analysis. However, among the key implications of the assumptions of regression is that the criterion is continuous. Little research has been done to examine how much information is lost and how inappropriate it is to use Likert variables in ordinary least-squares multiple regression. Therefore, this study examined the effect of Likert-type responses in the criterion variable and predictors for various scale points, on the accuracy of regression models using normal and skewed observed response patterns. This was done for the case of three predictors and one criterion. Similarly, eight levels of Likert-type categorization ranging from two to nine scale points were considered for both predictors and criterion variables.

It was found that the largest bias in the estimation of the model R-squared, the relative Pratt Index, and Pearson correlation coefficient occurred for two or three-point Likert scales. The bias did not substantially reduce any further beyond the four-point Likert scale. Type of correlation matrix had no effect on the model fit. However, skewed response distribution resulted in large biases in both R^2 and Pearson correlation, but not in Relative Pratt index, which was not affected by the response distribution.

Practical contribution and significance of the study is that it has provided information and insight on how much information is lost due to bias, and the extent to which accuracy is compromised in using Likert data in linear regression models in education and social science research. It is recommended that researchers and practitioners should recognize the extent of the bias in ordinary least-squares regression models with Likert data, resulting in substantial loss of information. For variable importance, the relative Pratt index should be used given that it is robust to Likert conditions and response distributions. Finally, when interpreting reported regression results in the research literature one should recognize that the reported R-squared

values are underestimated and that the Pearson correlations are also typically underestimated -- and sometimes substantially underestimated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Dr. Bruno Zumbo for the support, guidance and counsel he provided me, with great commitment, magnanimity and dedication. I also wish to thank him for his patience, encouragement, direction and insight during the preparation of this dissertation. I am grateful to the committee members, Dr. Beth Haverkamp, and Dr. Anita Hubley, and the University Departmental External Examiner, Dr. Kimberly Schonert-Reichl, for their suggestions, constructive comments, and review.

I would also like to thank Dr. Perry Leslie, head, department of Educational Counseling Psychology and Special Education who provided me with some great advice on graduate issues. Last but not least, I wish to thank my wife, Jenipher, and my children for their patience and encouragement during the preparation of this dissertation.

CHAPTER I

BACKGROUND TO THE PROBLEM

Introduction

Categorical variables used in survey and social science research often have a built-in order property except when used as labels, in which case they are referred to as nominal scales. An ordered categorical variable Y is often as a result of coarse grained measure of an underlying continuous variable η. For example, a dichotomous variable is observed as Y=1 when η exceeds some threshold value τ, and as Y = 0 otherwise (Olsson, Drasgow, & Dorans, 1982). In psychology, sociology, and biometrics, as well as econometrics, examples abound for which it is reasonable to assume that a continuous variable underlies a dichotomous or polytomous observed variable. A typical example of ordered category is the Likert scale, which has five scale points at equal interval. It is assumed that the underlying latent variable is continuous. Due to the popularity of Likert scales, similar scales based on different number of scale points, and referred to as Likert-type scales have been developed.

Many of the measures obtained in educational research are Likert-type responses on questionnaires. These Likert-type responses are considered ordered-categorical observed variables where the underlying variable is completely unobserved (i.e., latent). Furthermore, as the normally distributed latent variable increases beyond certain threshold values, the observed variable takes on higher scores, referred to as scale points. As is commonly found in the educational research literature, these variables are often referred to as Likert variables wherein, for example, such a variable with four possible observed values is commonly referred to as a "four-point Likert scale". Similarly, a variable with five observable values is referred to as a five-point Likert scale.

Likert variables are frequently used in ordinary least-squares regression analysis. Importantly, however, among the key assumptions of regression is that the criterion is continuous. Little research has been done to examine how much information is lost and how inappropriate it is to use Likert variables in ordinary least-squares regression.

Information loss

In using Likert variables, it is assumed in the literature that the larger the number of scale points the closer the estimation of the resulting statistics from the observed responses are to the population parameters of the underlying latent continuum. This is based on the premise that the underlying continuum is continuous and normally distributed. Similarly, the observed variables are also assumed to be continuous and normally distributed. However, real life response data derived from Likert scales are discrete and are not necessarily normally distributed. This results in underestimation of the population parameters of the underlying construct in the resulting statistics. The consistent underestimation of the population parameters is known as bias or in this case, loss of information. Thus, the loss of information referred to here is loss relative to the latent continuum.

In linear regression model fit, the loss of information is relative to the statistics one would obtain with latent variables as compared to the observed Likert variables. For example, if it is bias in R-squared, then it is the result of comparing R-squared from Likert variables in the model to the R-squared obtained from continuous and normally distributed latent variables. The present study focuses on how different the models based on Likert variables would be, compared to the models based on the ideal latent continuum that is continuous and normally distributed.

Implications of Likert variables

In order to place the discussion of the implications of using Likert variables in regression analysis in context, let us consider a survey of students' life satisfaction Y as predicted by the following three variables namely, family relations X_1, friendships X_2, and recreational activities X_3. The questionnaire read as follows:

How **satisfied are you** with: (*Please circle your response*)

	Very Dissat.			Evenly Balanced		Very Satis.	
QL1. Your family relations, generally	1	2	3	4	5	6	7
QL2. How you feel about life as a whole	1	2	3	4	5	6	7
QL3. Your friendships	1	2	3	4	5	6	7
QL4. Your recreation activities	1	2	3	4	5	6	7

It is not uncommon for researchers to conduct an ordinary least-square regression analysis like:

Criterion (i.e., dependent) variable: QL2 (Life as a Whole)

Predictor (i.e., independent) variables: QL1 (Family Relations)

QL3 (Your Friendships)

QL4 (Your Recreation Activities)

What follows are some frequency charts from real data gathered in Northern British Columbia by the Institute for Social Research and Evaluation (N=270) shown in Figure 1 to 3.

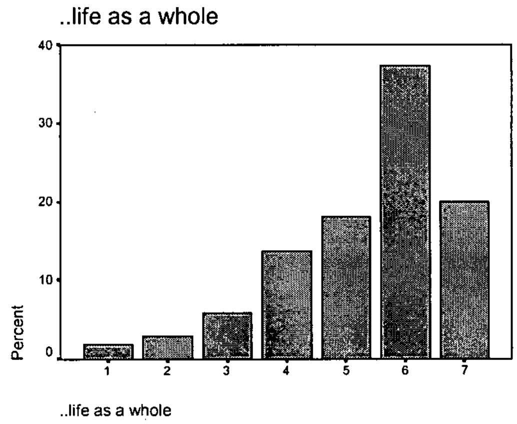

Figure 1
Distribution of responses on QL2 (Life as a whole)

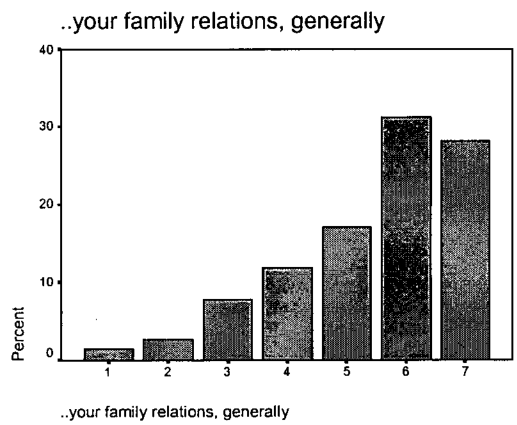

Figure 2
Distribution of responses on QL1 (family relations)

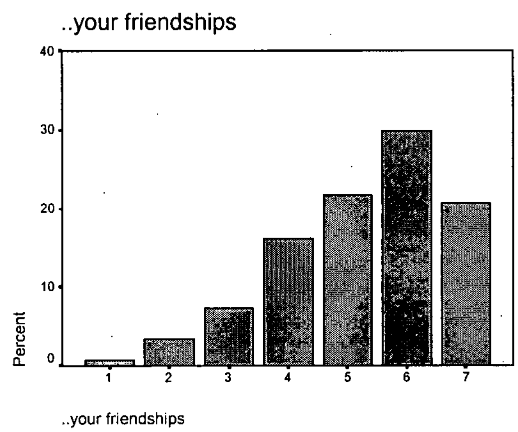

Figure 3
Distribution of responses on QL 3 (Your friendships)

From these charts it can be seen that the response patterns are, as typically found in surveys in this field, skewed. The results of conducting the ordinary least-squares regression are: R-squared = 0.467, and the standardized regression equation (with the beta-weights) is

$$ZQL2 = 0.262 * ZQL1 + 0.511 * ZQL3 + 0.048 * ZQL4, \qquad (1)$$

resulting in relative Pratt indices (Thomas, Hughes, & Zumbo, 1998) of 0.264, 0.698, and 0.038 for QL1, QL3, and QL4 respectively. Relative Pratt indices measure the proportion of the model R-squared that is attributable to each predictor – this is sometimes used as a measure of the variables importance in the model.

Speaking again more generally, if a simple additive regression model were to be fit (excluding the interaction terms), four possible combinations of continuous and Likert variables that would emerge are: (1) when the criterion variable Y is continuous and all the three predictors are also continuous, (2) when the criterion variable is continuous and the predictors are Likert, (3) when the criterion variable is a Likert and all the predictors are continuous, and (4) when the criterion variable is Likert and all the predictors are Likert. In all the four cases of combinations of continuous and Likert variables a simple regression model based on the standard ordinary least-square estimation can be written as follows:

$$Y_i = b_o + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + e_i \qquad (2)$$

where $i=1,2,....n$ number of observations or participants in the study. The error term e_i is that part of the criterion variable Y, not accounted for by the three predictors, and also includes measurement error. The key assumptions of the regression model in essence, concerns the properties of the errors and the distribution of Y conditional on the predictors. It is assumed that at each value of the predictors, there is a sub-population with a marginal distribution of Y. The population regression line passes through the mean of each sub-population in the marginal distribution. Observations within each of the sub-populations in each predictor are assumed to be independent of each other. Furthermore, for the purpose of inference, each of the sub-populations must be normally distributed with equal variance. The assumption of continuous distribution of Y is implicit in the stated conditions. The predictor X_i observation is assumed to be fixed and independent of the errors. However, the errors in Y_i are normally distributed and are uncorrelated, for every pattern of X_i observations [$N(0, \sigma_e^2)$; $Cov(e_i, e_j) = 0$, $i \neq j$].

The regression model is assumed to be linear in the coefficients of the predictors, while the predictors in turn are not a linear function of each other. In conducting regression analysis, the researcher is expected to test for the assumptions, and determine to what extent the assumptions hold in the given data. Unfortunately, most data in social science are derived from attitudinal and other affective measures that are Likert (Dawes, 1972). Consequently, researchers must often estimate and analyze relationships between continuous concepts, using the reported Likert data. This poses a measurement problem because the degree of association (i.e., R^2) and the estimation of the criterion variables (i.e., the predicted Y), whether continuous or Likert scale points, with the predictors may differ when continuous or Likert predictor variables are used. Moreover, the implicit assumption in linear regression models that the criterion variable is continuous, is often violated. Furthermore, studies have shown that when correlation coefficients based on categorized measures are used, their inaccuracy distorts the subsequent analysis and interpretation, resulting in lack of precision in measurement (Bollen & Barb, 1981; Cox, 1974). Therefore, it is important to determine the extent to which the distortions lead to imprecision, and what information is lost in the process. In other words, how misguided are researchers and practitioners in the current research practice in linear regression models, of analyzing continuous concepts with ordinal categorical measures?

It is important to note at this point, that Likert variables do occur naturally in educational and social research settings although, as a common practice researchers may also create ordinal categories for convenience in data analysis. The difference in this case with the Likert scale points is that the underlying variable is not latent, but strictly observable responses.

The general problem of correlating ordinal categorized or collapsed continuous measures have been studied as far back as the beginning of the century (Pearson, 1913), and later revisited in the middle of the 1970s by researchers (Cox, 1974; Grether, 1976; Henry, 1982; Kim, 1978; Labovitz, 1975; O'Brien, 1979). However, the main issue in the studies was the effect of ordinal categorical variables on Pearson's correlation, and whether ordinal categorical data should be treated as interval data. The impact of ordinal categorization on multiple regression models was not addressed. All the studies concentrated on the bivariate case of Pearson's correlation. No studies so far have considered the multivariate case involving multiple regression models.

Several studies have been conducted on the effect of ordinal categorical variables (herein referred to as Likert variables) on identification of factors and components in factor analysis, confirmatory maximum likelihood factor analysis, and principal component analysis (Babakus, Ferguson, & Joreskog, 1987; Bernstein & Teng, 1989; Green, Akey, Fleming, Hershberger, & Marquis, 1997; Muthen & Kaplan, 1985; Nunnally & Bernstein, 1994). Results from these studies indicate that the effect of ordinal categorization of data leads to misrepresentation of the underlying factors and wrong identification of the dimensionality of the latent variables. The distribution (skewness and kurtosis) of the observed ordinal categorical or Likert variables also result in inflated goodness of fit indices such as chi-square values, which in turn lead to increase in type I error rate (Muthen & Kaplan, 1985). Curran, West, and Finch, (1995; 1996) stated that when observed ordinal categorical variables were non-normally distributed, the chi-square values were larger than expected (i.e., inflated) leading to identification of spurious of factors. The standard errors of the correlation between the factors were also underestimated. Thus, non-normality leads to severe underestimation of standard errors of parameter estimate and renders their interpretation untrustworthy.

Other studies were conducted on the effect of Likert variables on the reliability of the scores of the criterion variable (Matell & Jacoby, 1971; Chang, 1994) with mixed and conflicting results. In each of the two studies it was found that reliability measures did not improve with the increased refinement of the Likert scale points. Thus, the size of the resulting reliability coefficient was deemed to be independent of the number of Likert scale points. Eom (1993) studied the interactive effects of ordinal categorization on measures of circularity in multivariate data and found significant effects of categorization. However, no attempt has been made in studying the effect of ordinal categorization on multiple linear regression models. It is recognized that ordinal categorization resulting in Likert scale points have implications to the accuracy of results in multivariate analysis and specifically, multiple regression, to the extent that some methods have been proposed to resolve the problem using logit and probit models (Ananth & Kleinbaum, 1997; Cliff, 1994; Long, 1999; Winship & Mare, 1984), and yet little has been done on linear regression models. In particular, the extent of information loss in these models is not known.

Problem Statement

Given the results of the studies on the effect of ordinal categorization resulting in Likert scale points, on the observed variables in linear regression, and the occurrence of errors in the measures, what information is lost in using Likert data in ordinary least-square linear regression models? To address this problem, and the implications of the associated variables, the effect of the distribution of the responses, number of Likert scale points, and the underlying latent variable were investigated.

Rationale of the Study

A common practice in educational and social science research is to use ordinal categorized scale such as Likert scales, to measure observable variables, and to infer a continuous latent variable. The inference is based on analysis of data using multivariate methods such as linear regression models, which implicitly assume that the observed criterion variable is continuous and meets the basic assumptions for linear regression models. However, the analysis of ordinal categorical variables is conducted as if they were continuous. While studies have demonstrated that measurement errors emerge from this practice, the extent of information loss and therefore the levels of imprecision are not known. This impacts on the accuracy of interpretation of results and generalizability of findings.

Several Likert-type scales have been adapted for use, ranging from two to ten scale points. However, the number of Likert scale points has been hypothesized to influence the accuracy of the criterion measures (Chan, 1994). Bollen and Barb (1981) found that the Pearson correlation between two continuous variables was higher than the correlation between the same variables when they were divided into ordinal categories i.e. transformed to Likert scales. The greatest attenuation occurred when few ordinal categories were employed for either variable in the correlation. Findings indicated that coarse ordinal categorization (i.e. fewer Likert scale points) of continuous variables resulting in Likert variables, was found to have a negative impact on parameter estimates and standard errors. Likert variables with more Likert scale points provided more accurate measures than those with fewer scale points. Based on the previous studies on correlations, it was hypothesized that the use of Likert variables in multiple regression analysis would result in inaccuracy and information loss. Several studies on Likert scales have focused on Pearson correlation, factor analysis and structural equation modeling but little has been done on the impact of Likert data on multiple regression analysis. Therefore, there was a need to study the extent of the inaccuracy for better measurement precision.

Other than the number of scale points in Likert variables, the distribution of the observed variables and that of the underlying latent variable have been shown in previous studies to influence the accuracy of parameter estimates, and model fit in factor analysis (Babakus, et.al, 1987; Muthen & Kaplan, 1985; Muthen & Kaplan, 1992). Given that multiple regression models have not been investigated, the present study investigates the extent of the impact of the distribution of responses in Likert variables on the accuracy of the ordinary least-square regression models, which are frequently used in education and social science research, and also addresses, in particular normal and non-normal (positive and negatively skewed) distributions.

In the light of the need to investigate the use of Likert variables in multiple regression analysis, the research literature is first reviewed in chapter two leading to the methodology and study design in chapter three. The results are presented in chapter four and the discussion as well as the implications of the findings is presented in chapter five.

CHAPTER II

LITERATURE REVIEW

Introduction

Concern for the effect of ordered categorization on correlations and its interpretation goes as far back as 1913 in an exposition in which Pearson discussed the analysis and interpretation of Pearson's r correlation from coarsely ordered categorized variables. Although studies abound on the effects of ordinal categorization of continuous variables on some multivariate analysis results, little has been done in linear regression models. Studies on Likert variables have tended to concentrate on factor analysis using generalized least-square methods and maximum likelihood, and structural equation modeling (Green et.al, 1997). An additional variable other than the number of Likert scale points is the effect of the distribution of the underlying latent variable, and the distribution of responses of ordinal categorized variables (Muthen & Kaplan, 1985, 1992). Both the distribution and number of Likert scale points affected the correct identification of factors and model fit resulting in spurious factors. The main concern in the studies was the imprecision resulting from ordinal categorization and therefore Likert scales. However, the concern in the present study is the loss of information in fitting regression models with Likert data, as most studies tended to concentrated on Pearson correlation and factor analysis. Relevant studies were reviewed for theoretical and methodological contributions to address the model fit in linear regression models using Likert data.

Likert Data and Pearson Correlation

Collapsing measures in continuous variables to create Likert scales introduces errors, which presents potential problems in the measures of association among the variables (Bollen & Barb, 1981). An earlier study by Labovitz (1975) demonstrated that ordinal categorical variables resulting into Likert data, could be analyzed as continuous variable with interval scale with equal accuracy. The basis of the claim was that Likert data are linear monotonic transformations of the underlying continuous variables. Thus, Likert data could be treated as continuous data and analyzed as if continuous. O'Brien (1979) reexamined the study by Labovitz (1975) and found contrary results with respect to the Pearson's correlation. The main focus of O'Brien's study was how the collapsed or ordered variables correlated with its true continuous or interval measure. The correlation between continuous or interval variable and its categorized version was not a monotonic function of the number of categories as previously claimed. This implied that Likert data could not be analyzed as if continuous, as proposed earlier. However, O'Brien did not address the correlation between two different categories and variables. Moreover, only normal and uniformly underlying distributions were considered. Non-normal distributions were not considered.

As a follow-up to the O'Brien (1979) study, and using simulated data, Bollen and Barb (1981) investigated how the correlations between continuous variables were affected by analyzing collapsed or categorized scales of the same continuous variables. In addition to this analysis, they investigated whether the number of categories affected the accuracy of the Pearson's correlation between the categories and the continuous variables. Continuous variables with a normal distribution were generated and collapsed into a number of categories ranging from 2 to 10. Collapsing the scale resulted in variables that were symmetric and approximately normal. Each pair of the variables, collapsed and continuous, was constructed to correlate at one of five magnitudes: 0.2,0.4, 0.6, 0.8, and 0.9. Fifty samples of 500 observations each were generated for each correlation. The variables were collapsed into smaller number of categories and the resulting correlation between the collapsed variables compared to that of the original continuous variables.

Findings indicated that differences in the correlations of the continuous and the categorical variables decreased with the increase in the number of categories. The standard error of the collapsed correlation was greater than that of the continuous variable correlations. Results of the study were in support of Labovitz's (1975) study and contrasted with O'Brien's (1979) study. While the O'Brien study found that the correlation of a collapsed variable with itself was a non-monotonic function of the number of categories, the Bollen and Barb (1981) study found that the correlation between two different collapsed variables and that of each its own versions were monotonic functions of the number of categories in each case. The study concluded that the more categories used, the more accurate the measure of correlation. It was recommended in the study that further investigations be conducted to determine the effect of the distributions, in particular, non-normal distributions. Furthermore, the question of correlated measurement error as a result of collapsing the scales needs to be investigated, as correction for attenuation is not appropriate when measures are correlated. The study recommended that implications of ordinal categorical variable to correlation estimates involving more than one variable be explored, as this was not addressed in the study as well as in the previous ones.

Because Pearson correlation is often used in the computation of reliability coefficients, bias in the estimation of the Pearson correlation due to Likert scale points have psychometric implications which has led to pertinent questions with regard to the effect of Likert data on the reliability estimates of instruments.

Likert Data and Reliability Estimates

Use of Likert data has been shown to result in underestimation of reliability estimates. These findings has psychometric implications and therefore, research concerns. Cicchetti, Showalter and Tyler (1985) conducted a simulation study of the effect of the number of Likert scale points on the reliability of measures of a clinical scale. The aim of the study was to investigate the extent to which the inter-rater reliability of a clinical scale is affected by number of categories or Likert scale points. Results indicated that reliability increased up to the 7-point Likert scale and thereafter remained unchanged when the Likert points were increased.

The study investigated how inter-rater reliability under different conditions compare for dichotomous, Likert and continuous scales of measurement with categories ranging from 2 to 100 scale points. The average levels of inter rater agreement were 30%, 50%, 60% and 70%, across the main diagonal of a rater 1 by rater 2 contingency table. These levels were chosen to be consistent with clinical applications. The sample size for each condition simulated was 200. Given the large number of rater pairings as the categories approached 100 scale points, 10 000 replications were used in each condition. Selection criterion of the reliability measure was based on the fact that the criterion would (1) measure levels of inter rater agreement rather than similarity in ranking, (2) correct for amount of agreement expected by chance, and, (3) validly be applied to all the three types of scales being investigated. Thus, the intra-class correlation coefficient identical to Cohen's kappa was chosen.

To compare the extent to which inter-rater levels were affected by categories of scale points, difference in the size of intra-class R (weighted kappa with quadratic weights) were computed at each level of the categories. Significance of the data was based on the guidelines proposed by Cicchetti and Sparrow (1981) in which clinical significance of intra-class reliability values were rated as follows: those less than 0.40 as poor, 0.40 to 0.59 as fair, 0.60 to 0.70 as good, and 0.75 to 1.00 as excellent.

Level of inter-rater agreement was lowest for 2-point Likert scale and highest at the 100-point Likert scale. The reliability measure increased as the number of Likert scale points increased with most dramatic increases being at 2 and 3-point Likert scales. Beyond the 7-point Likert scale, increase in reliability measure was not substantial. The result from this study differed from the previous ones on reliability in that the methodologies used varied across studies, while some of the results such as those by Komorita and Graham (1965) were found to be sample specific. The reliability measures used in previous studies were Cronbach's alpha and Pearson correlation which are well known measures of internal consistency and not inter rater reliability, and so the Cicchetti et al (1985) study differ from the rest of the studies on Likert scale and reliability in this regard. The study demonstrated how inter rater reliability measures were affected by number of Likert scales. Increase in scale points up to 7-point Likert scale improved the reliability ratings. However, beyond the 7-point scale there was no significant increase and improvement of the reliability coefficient.

Krieg (1999) examined bias induced by coarse measurement scale in Pearson correlation and reliability coefficients. Equations for calculating the induced bias were derived, based on the probability density function of the scale and the rule for assigning values to the scale. Expectations of the value of the mean, variance, covariance, correlation coefficient and reliability coefficient were also derived. The study demonstrated that bias in the correlation coefficient, and reliability coefficient vary depending on factors such as the mean and variance of the quantities being measured, the number of scale points, the rule for assigning quantities to scale points and number of items in the instrument. Equations for the limits of the biases were provided.

Krieg (1999) also extensively reviewed previous studies on the effect of categorization or coarseness of scale on reliability measures of scores and Pearson correlation, starting with the Symond (1924) study on the number of scale points on reliability. The Symond study found that using a 7-point Likert scale was the most suitable and optimal as the relative gain in using more scale points beyond seven was not worthwhile. This was followed by a review of Champney and Marshall's (1939) study that tested the Symond findings. Champney and Marshall showed that the Symond study produced larger than expected bias in the estimation of reliability coefficient and correlation coefficient. This was because bias in the covariance and variance were not previously considered by earlier studies. They suggested that coarseness of a scale affected the covariance in the numerator and the standard deviations in the denominator in the correlation coefficients, and should be considered when computing bias.

Other studies such as Bendig (1953) and Komorita and Graham (1965) gave mixed results. Bendig (1953) found that reliability coefficient based on Hoyt's analysis of variance remained unaffected by the number of scale points for individual ratings and intra-class correlation methods. This was contrary to findings from other studies on reliability and Likert scales.

In the Komorita and Graham (1965) study, respondents were asked to use either 2 or 6-point Likert scales. The number of items and the homogeneity of the items were also varied. Increase in the number of items increased reliability of the items with low homogeneity but not for those with high homogeneity. Also reviewed was the Matell and Jacob (1971) study in which respondents used 2 to 19–point Likert scale. The study found that internal consistency as measured by coefficient alpha, and test-retest reliability coefficients were not affected by number of scale points.

Nine simulation studies relating effects of Likert scales on reliability coefficients and correlation coefficients were also reviewed, starting with the Lissitz and Green (1975) study. In the Lissitz and Green study, 10 items with three levels of covariance were simulated. Numbers of scale points were varied by transforming normally distributed data to discrete uniform distribution. Coefficient alpha, test retest reliability and correlation between observed and true scores were computed. It was found that all the coefficients increased as the number of scale points increased at all levels of covariance.

The second simulation study was the Jenkins and Taber (1971) study, in which the number of items, covariance between items, and measurement error were varied. The study found that under all the stated conditions, coefficient alpha, test retest reliability and the correlation between observe and true scores increased as the number of scale points increased from 2 to 5, but leveled off thereafter.

Other simulation studies that followed concentrated on the effect of coarse measurement on Pearson correlation. In particular, studies by Wylie (1976), Martin (1973, 1978) and Bollen and Barb (1981) showed that using coarse measurement (i.e. Likert scales) reduced the Pearson correlation between two variables. Cohen (1983) demonstrated through simulations that dichotomized variables reduce statistical power for detecting a relationship between two variables. Dichotomization of the variables lowered the Pearson correlation coefficient between two variables. In the simulation study by Cicchetti et al (1985), it was shown that inter rater reliability increased as the number of scale points increases up to 7-point scale. In the case of regression models, Russel, Pinto, and Bobbko (1991) as well as Russel and Bobbko (1992) showed through simulations that effect sizes in moderated regression analysis vary according to number of scale points. Effect of coarse measurement on the estimation and testing of structural equation models was also demonstrated by Bollen (1989) using simulated data.

Based on the classical test theory model, bias in correlation coefficients resulting from coarse measurement is provided in Krieg (1999). The measurement model is given by:

$$X = T + E \qquad (3)$$

where T is a real-valued constant or variable that represents the quantity to be measured, and E represents random independent measurement errors. X is the observed score. In coarse measurement (e.g. Likert scales), the observed score is transformed to a measurement scale by a function $Y = f(X)$, where function f may vary depending upon the scaling procedure. Conversion of response data to a scale that is coarser than the observed quantities to be measured results in bias. Biases are expressed as the difference between the moments of X and the moments Y. Moments are determined by the distribution of X. The values Y results from making measurements with an N-point scale which is represented as the weighted sum of N variables, Z_i that are mutually exclusive, taking on values 0 to 1, and are multinomially distributed such that:

$$Y = \sum_{i=1}^{N} S_i Z_i \tag{4}$$

S_i is the value of the *ith* scale point. The expected value of Y is:

$$E(Y) = \sum_{i=1}^{N} S_i P_i \tag{5}$$

P_i is the probability that S_i occurs and is dependent on the distribution of X and the function $f(X)$. Bias of Y is defined as:

$$\beta = Y - X, \tag{6}$$

such that the bias of the expected value of Y is $\beta_\mu = E(\beta) = E(Y) - E(X)$. *(7)*

Bias for the variance is defined as:

$$\beta_v = Var[Y] - Var[X] \tag{8}$$

where $Var[Y]$ is the variance of Y and $Var[X]$ is the variance of X. Bias of the covariance is derived from the definition of covariance as follows:

$$Cov\,[Y_1 Y_2] = Cov\,[\sum_{i=1}^{N} Z_{1i} S_{1i} \sum_{j=1}^{N} Z_{2j} S_{2j}] \tag{9}$$

Therefore bias of the covariance is as follows:

$$\beta_c = Cov\,[Y_1, Y_2] - \sigma_{12} \tag{10}$$

Bias of the correlation coefficient is a function of the bias of variance and covariance used in the calculations. Thus, Krieg (1999) shows that the bias for the correlation coefficient is:

$$\beta_p = \rho_y - \rho_{12} = \frac{\sigma_{12} + \beta_c}{\sqrt{(\sigma_1^2 + \beta_{v1})(\sigma_2^2 + \beta_{v2})}} - \frac{\sigma_{12}}{\sigma_1 \sigma_2} \qquad (11)$$

where the correlation of Y_1 and Y_2 is given by ρ_y while ρ_{12} is the correlation between X_1 and X_2, and σ_1, and σ_2 are the standard deviation of X_1 and X_2. To calculate the bias of the correlation coefficient, one would need to calculate the *Cov* $[Y_1, Y_2]$, *Var* $[Y_1]$ and *Var* $[Y_2]$ and then ρ_Y. Similarly, ρ_{12} is calculated from σ_1, σ_2 and σ_{12}. The value of σ_{12} is the covariance of X_1 and X_2. Bias in reliability coefficient is treated as bias in correlation coefficient in that reliability coefficient is regarded as a correlation measure between two parallel variables.

In the study by Krieg (1999), bias in the correlation coefficient was demonstrated for population correlation $\rho = 0.8$ across scale points ranging from 2 to 10. The number of scale points for each variable was set to be equal. The minimum scale was –3, and the maximum was +3. Simulation was done as a check for the calculated biases in the correlation coefficient. A sample size of 1000 was used in the simulation. Bias in the correlation coefficient was negative, implying that coarse measures attenuated the correlation coefficient. Bias obtained from computational formulae (11) stated, those obtained from simulation results, and those from Bollen and Barb study, as well as Sheppard's formula were compared for a population correlation $\rho = 0.8$. For scale points greater than three, the absolute amount of bias decreased as the number of scale points increased. The difference in the amount of bias across the methods, decreased with increase in number of scale points. This observation was more evident for scales beyond the 6-point Likert scale. While the study showed a systematic decrease in bias in the correlation across the methods with equal interval between scale points and stated conditions, little is known of the effect of unequal interval on the degree of bias. This was investigated in the present study.

Krieg recommended the use of the correction formula as the degree of bias is almost equivalent to those observed in simulation result suggesting that the formulae is equally accurate. However, a limitation of using the formula is the fact that the data must meet the assumptions made and the stated conditions, otherwise the corrections may not be accurate. This is not the case in practice. Often, the response data is not normally distributed. It is also suggested in the study to use polychoric correlations for data when two continuous variables have been transformed into coarse measures as recommended by Olsson (1979) or to use polyserial correlation when only one continuous variable is transformed into a coarse measure while the other one is not.

Pearson Correlation and Non-normal Likert Data

Wiley (1976) investigated the effect of coarse grouping and skewed (observed) marginal distributions on the Pearson correlation coefficient. While most studies assumed normal distributions for both variables in the correlation, the effect of skewed observed distribution was previously not addressed. Twelve non-normal bivariate distributions were constructed to investigate the effects that coarsely grouped skewed marginal distributions have on the correlation coefficient, and the accuracy of the correction method used in grouping. The study showed that the correlation coefficient is severely limited in the values it can assume when the marginal distributions are highly oppositely skewed and one or both of the marginal distributions has only two intervals. The study showed that the correction formula is accurate for normal distributions only. The study concluded that the use of asymmetrical, dichotomous variables should be avoided in correlation analysis and that the correction for coarse grouping should be applied on experimental basis to correlations computed for joint distributions of coarse categorized variables.

The three research questions were as follows: first, whether the marginal asymmetry of coarsely categorized variables were likely to have the same restrictive effect on the correlation coefficient as do asymmetrical dichotomous variables. Second, whether there is a minimum number of class intervals that a researcher should employ in the construction of variables to be used in the correlation analysis, and third, how accurate is the correction formulae used for coarse grouping when applied to correlation coefficients computed for such variables.

To answer the research questions, the following procedure was used: three different types of bivariate distributions with the following pairs of marginal distribution were constructed (1) extreme negative skewed and extreme positive skewed (2) moderate negative skew with moderate positive skewed and (3) slight negative skew and slight positive skew. It was hypothesized that skewed marginal distributions lower the correlation between two variables.

Four correlation coefficients namely, 0.4, 0.5, 0.6, and 0.8 were chosen because the first three frequently occur in social science research, while the fourth correlation coefficient of 0.8 was chosen in order to examine the effect of reduced scale intervals on the extreme values of Pearson correlation. Each marginal distribution was constructed to have 24 class intervals. The intervals were then collapsed into 12, 8, 6,4, 3, and 2 categories by combining interval so that the correlation coefficient could be computed for every combination of interval grouping for the two variables.

The study compared the actual and corrected values of the correlation coefficient computed for each combination of collapsed class intervals for a given bivariate distribution. The computed values of the correlation coefficient dropped below the theoretical values of the correlation coefficients for fewer intervals among the variables. This was more pronounced at the 2 interval scales and extremely skewed distributions. Although corrected correlation coefficients were consistently below the actual values of the correlation coefficient, they were more accurate than the computed correlation coefficients and were least affected by skewness of the marginal distributions.

An implication of the study is that correction formula for coarse grouping could be used for intervals greater than three with a reasonable limit of accuracy, for skewed and normal marginals, as long as the underlying assumptions are met. The correction formula for corrected correlation coefficient used in the study was that proposed by Peters and Van Voorhis (1940) given as follows:

$$\hat{\rho}_{ij} = \frac{\rho_{ij}}{\rho_{cx}\,\rho_{cy}} \qquad (12)$$

Where ρ_{ij} is the actual computed value of the correlation coefficient for a bivariate distribution whose marginals had i and j class intervals respectively ($\rho_{24,24}$ was used to represent the value of population correlation in the study). The Pearson correlation ρ_{cx} is the correlation of coarsely grouped variable X with its continuous form of X, while ρ_{cy} is the correlation of coarsely grouped variable Y with its continuous form of Y. The formula was shown by Peters and Van Voorhis (1940) to yield larger biases in the estimation of Pearson correlation for intervals less than 15.

The second implication of the study is that the use of dichotomous variables in correlation analysis can be problematic in cases where the underlying continuous distributions of such variables are even moderately skewed. Dichotomous variables yield very poor estimates of the true Pearson correlation between the continuous forms of there variables. However, because of the non-continuous nature of many dichotomous variables such as sex, ethnicity, religious affiliation, it is unrealistic to exclude dichotomous data in correlation analysis. Often, such variables are included in multiple regression analysis as predictors of practical criteria like academic and job performance, absenteeism, grade point average etc. It is therefore important to be cognizant of the extent of bias that occurs as a result of coarse measurement and Likert scales both as criterion and predictors. This provides a justification for determining the extent of bias in multiple regression analysis with Likert data.

Likert Data and Scale Intervals

Cohen (1983) provided an exposition on the cost of dichotomization of continuous variables for Pearson correlation under the bivariate normality condition. He considered the case of dichotomizing one variable only at the mean, and when both variables are dichotomized at the mean. The effect of dichotomization on effect size and power were assessed for Pearson correlation of r = 0.2, 0.4 and 0.5 under different sample sizes.

A common practice in social science research is to dichotomize continuous variables in order to simplify the data analysis. The decision is not prudent as the loss of information is substantial, especially in small sample sizes and for data with skewed distribution. Cohen (1983) criticized dichotomization arguing that it results in underestimating effect sizes and reduces the power of statistical hypotheses tests. Specifically, dichotomization results in proportions of variance accounted for that are approximately 0. 64 times as large as when it is performed on one of the two variables being correlated.

Cohen (1983) showed that, assuming a bivariate normal population with a Pearson correlation of r, if one variable is dichotomized at the mean so that two equal intervals result, then the observed correlation between the dichotomized variable, X_d and the continuous variable, Y was equal to 0.798 of Pearson correlation r. This procedure resulted in X_d accounting for 0.637 as much variance in Y as the original continuous X. When dichotomization was done one standard deviation away from the mean, the resulting correlation was 0.66 of Pearson correlation r, while dichotomization at one and a half standard deviation away from the mean resulted in 0.52 of Pearson correlation r, with the explained variance dropping to 0.27 of as much variance in Y explained by the original X. The further away from the mean the dichotomization, the worse the estimation of Pearson correlation between the two variables got.

It was also observed that dichotomization lowered the t-values associated with the resultant Pearson correlation, and lowered the latter's chance of being statistically significant. This was more apparent in small sample sizes. For large sample sizes, t-values were lowered to three fifths of their expected values in the original continuous *X* variable. It was evident that, for Pearson correlation values ranging from 0.3 to 0.6 typically occurring in social science research, dichotomization was very costly in terms of information loss.

Effects of dichotomization on statistical power and effective sample size were also addressed. Dichotomization of *X* variable at the mean results in the reduction of the population Pearson correlation to 0.798 times the original *r*-value i.e. 0.798 of *r*. For example, for $r = 0.3$, the correlation dropped to 0.239. The power is reduced from 0.78 (for a sample size of $n = 80$) to 0.57. However, for the non-dichotomized *X*, a Pearson correlation *r* of 0.3 produces a power of 0.57 with a sample size of n= 50. From this example, dichotomization is equivalent to discarding 30 of the 80 subjects for a correlation coefficient of 0.3. The loss in effective sample size when *X* is dichotomized at the mean is about 38% for Pearson correlation of coefficient of 0.2 to 0.5, at alpha of 0.05. It was further demonstrated that, for dichotomization at one standard deviation from the mean there, was a loss of 55% of the subjects over a range of population correlation ranging from 0.2 to 0.5.

To appreciate the costly effect of dichotomization, an alternative consideration was the number of subjects needed to offset dichotomization for a given population correlation. For a population correlation coefficient of 0.3, the sample size needed for power to be equal to 0.8 for a two-tailed test is 84, at alpha of 0.05. For optimal dichotomization at the mean the required sample size is 133. For a Pearson correlation of 0.2 the sample size increases from 193 to 304. Thus, the cost of dichotomization in terms of power and additional sample size is substantial and needs to be addressed.

The effect of dichotomizing both variables, X and Y, results in a four-fold table. The Pearson correlation transforms to a phi-coefficient. The dichotomization of X_d and Y_d results in a reduced correlation of 0.637 of the original correlation r. This is because 0.637 is actually the square of 0.798, the result of applying 0.798 correction twice, in X_d and Y_d. In terms of variance reduction, the effect of double dichotomization at the mean is the reduction of the population correlation $0.6372r^2$ to $0.405r^2$. The consequence to power of reducing correlation coefficient for double dichotomization can be translated into reduction in effective sample size. For a correlation of 0.3 and a sample size of n = 40, the reduction in power is from 0.47 to 0.21 at an alpha of 0.05, attainable from original data with n = 16. Thus, double dichotomization is equivalent to discarding 60% of the cases. Evidently, dichotomization results in a costly loss of information produced by a drop in correlation coefficient.

The loss in information as a result of dichotomization is not due to attenuation that is typical of random measure error, but rather an outcome of systematic loss of measurement information. An instance where dichotomization occurs is in blocking in analysis of variance. Since blocking is equivalent to partialling out, the reduction in the Pearson correlation produces the same kind of distortion that occurs when unreliable variables are partialled out. The situation is worsened when the blocking variable is dichotomized and when it is unreliable (Cohen, 1983).

The effect of dichotomization in factor analysis is evident when a batch of scaled items is dichotomized. The resulting phi-coefficient and the factor loadings yielded are approximately two thirds as large as the product moment correlation coefficients in the original data with communalities less than half as large (Cohen, 1983). In attitudinal scales involving Likert type scales with 4 or 6-point agree-disagree scales, dichotomization is effected at the middle of the scale to simplify the analysis. This results in loss of up to two thirds of the variance accounted for on the original variable and a concomitant loss of power equivalent to discarding about to two thirds of the sample. Such losses cannot be aptly justified. Thus, the implications of using categorized or Likert data in correlation analysis and regression models need to be investigated and the loss of information resulting from the practice be addressed in the light of validity evidence and meaningful data interpretation.

Martin (1978) investigated the effect of Likert scaling on the correlation coefficient. The study investigated the consequences of varying the number of Likert scale intervals on a statistical model and the distortion in the Pearson correlation due to restricted number of scale points. A simulation data set was used to evaluate the effect of scaling on the correlation coefficient using a simplified standard form of the bivariate normal distribution. For this distribution, the means of X and Y were assumed to be zero, and variances were assumed to be one, resulting in the following joint distribution of x and y:

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right) \qquad (13)$$

where ρ represents the correlation between two variables in X and Y with means of zero and variances of one. To create interval scale points, relative frequency distributions on both X and Y were determined based on the joint probability distribution function:

$$P(x,y) = \int_a^b \int_c^d f(x,y)\,dy\,dx \qquad (14)$$

The frequency of a given response falling between two given values *(a, b)* in X, and simultaneously between two Y values *(c, d)* can be determined by using the above double integral function. This is the relative frequency of occurrence for a bivariate normal distribution given a certain level of correlation. The procedure was used to calculate the frequency for all X and Y combinations of scaling between the mean, and 3.2 standard deviations away from the mean with a sample size of 1000. The total distribution was aggregated to provide frequency distribution for different number of scaling units. This was done for equal width of observation resulting in equal interval. Pearson correlation was then computed for different scales of X and Y.

Pearson correlation for continuous X and Y were computed and set to 0.1, 0.2, 0.5, 0.7, 0.8, and 0.9. Estimates of these correlations using different combinations of scale points of 20, 16, 14, 10, 8, 6, 4, and 2 of X and Y were compared to the true correlations derived from the continuous data.

The largest information loss occurred within smaller number of scale points in both X and Y. For example, an $r = 0.800$ using a 4-point scale in both X and Y results in $r = 0.588$. The resulting r^2 then dropped from 0.640 to 0.346, close to 50% drop from the original. There was a substantial drop in the magnitude of the Pearson correlation as the scale points decreased. Amount of information lost by collapsing the scales was greater when the original correlation was high than when the correlation was low. Furthermore, the amount of information lost is determined by the smaller number of scale points when the number of scale points is unequal for the two variables. For example, r^2 for 20 scale points of X and 2 scale points of Y is approximately the same as the r^2 for 14 scale points of X and 2 scale points of Y. The determining factor in the two cases is the 2-point scale.

Comparison of the Pearson correlation results to those of biserial correlation and tetrachoric correlation for continuous and dichotomized scales showed that Pearson correlation yielded a larger loss of information. However, Sheppard's correction formula provided better estimates than both biserial and tetrachoric correlations provided the assumptions were met. The overall observation is that ordinal categorization resulted in substantial information loss.

A limitation of the study is that the simulation was based on normally distributed data for equal interval scale points, addressing only one aspect of the measurement problem associated with Pearson correlation besides the limited sample size. Unequal interval and non-normal distribution of Likert data were not addressed for a wide range of sample sizes from low to high. Further studies need to be considered regarding the effects of skewed distributions and unequal intervals of scales in Likert data. Evidently, Likert scaling of the variables had a meaningful significant effect on the magnitude, model fit and interpretation of data as model builders are limited by the loss of information.

Likert data and Structural Equation Modeling

Johnson and Screech (1983), using simulated data, investigated the effect of measuring continuous variables with multiple indicators collapsed into ordinal categories (Likert scale points). The study found that categorization errors occurred, which increased with small sample size and fewer number of categories used. They observed that when continuous variables are collapsed into ordinal categories, the measurement errors introduced were correlated. This condition violated the assumption of classical measurement theory of independence of errors. Thus, the impact of measurement errors introduced by ordinal categorization of variables needs to be investigated, as it is not known.

In the study, a structural equation model with three latent variables was simulated. Each of the latent variables had two categorized indicators. Random errors were introduced in the underlying categorized indicator variables. Using SPSS, a data set was generated using random variables that simulated a normal distribution with mean of zero and a variance of one for a population of 5000.The resulting z-scores were then used to create the latent variables and the indicators. Combinations for conditions of the variables was created by crossing the beta coefficients of 0.4 and 0.6, and the lambda coefficients of 0.6 and 0.9.These represented the conditions in which the relationship among the unmeasured variables are high or low, and the error of measurement were high (lambda coefficient of 0.60) or low (lambda coefficient of 0.9). Seven different models were generated for each condition with ordinal categories, 2, 3, 4, 5, 10, 20, and 36. The sample size for each category and the distribution (normal and uniform) were varied, resulting in 56 different conditions of errors and ordinal categorization. To test for linear relationships among the categories and the continuous variable, tracelines were plotted for the underlying variables and the categorized indicators.

Results showed that non-linearity was more pronounced when two ordinal categories were used than in the larger ordinal categories. Errors in the correlation were substantial for smaller ordinal categories, uniform distribution and high lambda (0.9). While the size of the beta coefficients did not have an effect on the residuals, the size of lambda systematically affected the residuals. High lambda values resulted in large residuals and low values resulted in smaller residuals. It was noted that the fewer the number of categories, the lower the value of lambda (correlation between latent variables and indicator variables), and the greater the variability in the estimates. Thus, ordinal categorization resulted in distortions in multiple indicator models. The distortion reflected the non-linear relationship of the ordinal categorical variables with continuous variables, producing correlated errors. From the results, it was concluded that small ordinal categories such as 2, 3, and 4 should be used with caution, especially in small sample sizes because of the low efficiency of estimates.

The study did not address the effect of ordinal categorization when the underlying variables had a non-normal distribution. It was recommended that the behavior of the parameter estimates and correlated residuals in the models for ordinal categorical variables be investigated in future studies. In the study, intervals in each ordinal categorical variable were assumed to be equal. Studies were yet to be conducted on the effects of ordinal categorization on the observed variable for unequal interval scale. Although the study looked at the effects of ordinal categorization on multiple indicators, no mention was made on the effects on multiple linear regression models, as the study was based on structural equation modeling, which examines model fit and the relationship among latent and observed variables as well as residuals. Moreover, the assumptions on the estimation procedures differ in structural equation modeling from those in multiple linear regression models. Thus, the study's contribution to the current investigation can be viewed from its methodological approach and important results obtained pertaining to the effects of variable categorization and the effects of varying the underlying distribution.

A major concern on the effect of ordinal categorization of continuous criterion variables which results in Likert scale points is the impact of Likert variables on the identification of the dimensionality of the latent variables represented by the observed categorized variables. Coupled with ordinal categorization, which results in Likert scales points, is the influence of the underlying distribution of the observed and latent variables. Due to these concerns, several studies have examined the impact of Likert variables on the identification of dimensionality under different ordinal categorization scheme and distribution.

Bernstein and Teng (1989) demonstrated that multi-categorical item responses resulted in spurious factors and hence wrong identification of the dimensions of the categorized data. Three procedures were examined namely: Principal Component Analysis, Exploratory Maximum Likelihood Factor Analysis, and Confirmatory Factor Analysis using LISREL for five levels of ordinal categorization of items. Data was simulated for 1000 observations based on responses to 20 items (X_{ij}, with i denoting observation, and j denoting the item) using the equation:

$$X_{ij} = aF_i + (1-a^2)^{1/2} e_{ij} \qquad (15)$$

The pattern parameter a, took values 0.5, 0.71, and 0.87, generating three desired values of the pattern p of inter-variable correlations (0.25, 0.5, and 0.75). F_i is the factor score for subject i, and e_{ij} is the random error for subject i on item j. Both F_i and e_{ij} are in standard score form. Six correlation matrices were generated from X_{ij} within each level of p, corresponding to the latent, binary 0.5, binary 0.16, binary high and low (H/L), multi-categorical-equal interval (MC-equal), MC-H/L high and low. All the 18 resulting matrices were yoked to a common random number seed.

Results were analyzed for the procedures stated, namely reliability and inter-variable correlation, Principal Component Analysis, Exploratory Maximum Likelihood Factor Analysis and the Unifactor Confirmatory Factor Analysis using LISREL. Item categorization led to incorrect inferences concerning the dimensionality of the data for all the stated procedures used. For principal component analysis, evidence for a spurious second factor appeared most strongly at lower levels of inter-variable correlation, low reliability and with dichotomous items. This was reflected by the size of eigen values across the levels of inter-variable correlation and categories from low to high. For multi-categorical H/L data the number of factors increased for high reliability and inter-variable correlation. In the case of exploratory and confirmatory factor analysis, spurious factors emerged at higher categories of the data and high inter-variable correlation and reliability.

Similar results were found for the unifactor confirmatory factor analysis (LISREL). At the three levels of inter-variable correlation and reliability, all the levels of categories resulted in spurious factors. It was all noted that spurious factors emerged, independent of the sample sizes used (N= 100, 200, 500, and 1000). Bernstein and Teng explained their results as being an artifact of the heterogeneity of the item distribution for the multi-categorical items. Their results indicated that spurious factors emerged more often in multi-categorical than in dichotomous data. This contradicted results from earlier studies.

A limitation of the Bernstein and Teng study is that the data were generated with a normally distributed underlying factor. It would have been informative to include uniform and non-normal distributions. For dichotomous items, both symmetric and non-symmetric distributions were represented, whereas for multi-categorical items, only non-symmetric distributions were examined. Analysis of both symmetric and non-symmetric distributions in the multi-categorical items would provide additional information on the effect of ordinal categorization on the dimensionality of the underlying data. Even then, the study significantly contributed to understanding of the effects of ordinal categorization on dimensionality of data, given that most social science data involve ordinal categorization of continuous criterion variables with underlying continuous latent normal distribution. The study is relevant to the present investigation of the effect of ordinal categorization on multiple regression models in terms of the methodological contribution. The effect of ordinal categorization of items resulting in Likert data and frequently used normal distribution was also investigated. A few questions need to be addressed: namely, the effect of non-normality on the Likert data, and whether all the commonly used methods of model fit (Maximum Likelihood, Generalized Least-square and Ordinary Least-square method) as well as the recently developed, Asymptotic Distribution Free (ADF) method, would yield the same result and accuracy.

Muthen and Kaplan (1985; 1992) addressed these questions by considering factor analysis with non-normal Likert variables with various models and sample sizes. In their initial study, Muthen and Kaplan (1985) compared two methods of model fit (Maximum Likelihood, ML, and Generalized Least-square, GLS method) for factor analysis of normal and non-normal Likert variables. The study found no significant differences between the methods. However, in their 1992 study, Muthen and Kaplan included additional methods of model fit together with an increase in the size of models. The new methods of model fit included were, the asymptotic distribution free estimator (ADF) developed by Browne (1982), which does not assume normality, and the Categorical Variable Method (CVM) estimator developed by Muthen (1984) that explicitly takes into account the categorical nature of the variables, while assuming that the underlying variable is continuous and normally distributed. All the model fits were compared to each other and to CVM.

The study tested the hypotheses that first, estimators based on normal distribution were independent of the number of variables and factors. Second, the number of degrees of freedom is a significant factor in the robustness of the chi-square test of model fit, and third, the underlying distribution of the variables was not a significant factor in the robustness of the chi-square test of model fit. Two sample sizes, 500 and 1000 were simulated for each method of model fit. Data was generated for continuous random normal variable with known factor structure. The resulting variables were then categorized in such as way as to yield six different levels of non-normal categorical variables Y, with factor analyses models with the same number of factors but different parameter values. Four models of increasing sizes were considered. For non-normality, a negative kurtosis and skewness were considered as they were hypothesized to result in an underestimation of chi-square values.

The first simulated model was a six-variable two-factor model (with 8 degrees of freedom). The second was a nine-variable three-factor model (with 24 degrees of freedom), the third was a twelve- variable three-factor model (with 51 degrees of freedom), and the fourth one was a fifteen-variable three-factor model. One thousand replications were conducted for 2 sample sizes (N= 500 and 1000), 6 types of non-normality, and 4 model sizes. Data was simulated for one Likert scale and one continuous scale. A large replication was done in order to adequately assess standard errors, and the behavior and rejection frequency (type I error rate) of the chi- square.

Results indicated that ADF performed better than GLS and ML. The previous study (Muthen & Kaplan, 1985) however, indicated that there were no significant differences between GLS and ML. Compared to ADF, the chi-square values in GLS were consistently underestimated for higher categories and non-normality conditions. For ADF, there was a downward bias in the estimation of the standard errors. In both ADF and GLS, bias was more pronounced in the low sample size of 500 than it was for the 1000. For continuous data, which served as a baseline for comparison with the other ordinal categorization, all model fit methods considered performed well. In the case of ordinal categorical data, GLS was sensitive to sample size, level of ordinal categorization and non-normality. This was evident from the large standard errors in the conditions stated. Bias was constant across sample sizes and model sizes.

Results indicated that ADF chi-square values were sensitive to model sizes and low sample size. It was also found that for moderate size models, and for those models with fewer variables, chi-square values and standard errors were not as robust to non-normality as previously presented. The study concluded that ADF does not appear to function sufficiently well as a means of compensating for the effect of non-normality as previously assumed, unless the model is small and the sample size is large. Findings were that GLS standard errors are mainly affected by degrees of non-normality, while GLS chi-square, ADF chi-square, and ADF standard errors, are affected by both non-normality and model size.

The study did not address the effects of ordinal categorization on the model fit methods for underlying uniform distribution of the variables. Only one categorization procedure was simulated and compared to continuous data. Different categories needed to be studied to adequately address the effect of ordinal categorization of model fit. However, the study highlighted the importance of non-normality as a significant factor in the appropriateness of model data fit. The importance of the assumption of normality in the case of linear regression analysis and model fit can be viewed from this perspective. Therefore, this study has a significant contribution and relevance in terms of the methodology used to investigate violation of the assumption of normality and how non- normality affects model fit.

Green, Akey, Fleming, Hershberger and Marquis (1997) investigated the effect item ordinal categorization with various Likert scale points on chi-square fit statistics for confirmatory factor analysis to assess whether increase in categories results in the identification of spurious factors as previously reported by Bernstein and Teng (1989). In fact, the study was a follow up to the Bernstein and Teng study with the same simulation procedures and a few additional variables, which distinguished the study from that of Bernstein and Teng. Four types of continuous single factor data were simulated for a 20-item test. The four distributions of the item responses were (1) uniform distribution for all items (2) symmetric and unimodal (3) negatively skewed for all items and (4) negatively skewed for ten items and positively skewed for the rest of the ten items. For each of the four types of distributions, item responses were divided to yield 2, 4, and 6 equidistant response categories (equal interval). Distributional properties of the responses were crossed with number of scale points. The crossed factorial design facilitated the conclusion about the effects of the number of Likert scale points on the chi-square indices, separate from the effects associated with the underlying distributions. Data generation method was as follows: items were generated following a normal distribution with one underlying dimension. Four different transformations were applied to the normally distributed data to yield four types of non-normally distributed scores that were continuous with the ranges 0 to 1. Non-normally distributed scores were transformed to produce Likert data with 2, 4, and 6 Likert scale points. Continuous and Likert data were factor analyzed to assess how well a single factor model fit different types of data.

Confirmatory maximum likelihood factor analyses were conducted to evaluate the existence of a single factor for each of the five distribution types in continuous and Likert variables. Because the null hypothesis was assumed to be correct, the mean of the chi-square statistics was expected to be equal to the degrees of freedom. Standard chi-square goodness of fit measure as well as Bentler's Comparative Fit index (CFI; Bentler, 1990) was computed. Also computed was the Satora-Bentler scaled chi-square as it was hypothesized to be robust to non-normality and therefore less prone to suggest spurious factors than the non-scaled ADF chi-square.

Results indicated that chi-square statistics for evaluating the one-factor model was more inflated at the 2-point Likert scale responses than at the 4, and 6-points Likert scales. Chi-square statistic became less inflated with the increase in the number of Likert scale points. The degrees of freedom were inflated for uniform distribution and for negatively skewed distribution. Chi-square values were more inflated for Likert data than for continuous data. As the Likert points increased, the mean chi-square values approached that of the continuous data. Across the distributions, the Satora-Bentler chi-square scaled statistic was close to the expected value, except for skewed distribution and for two category data. For different skewness, both scaled and unscaled chi-square indicated spurious factors, even for large number of Likert scale points. Increase in number of Likert scale points improved the chi-square statistic in terms of type I error rate. The CFI was sufficiently large (CFI>0.99) and close to one, for all model fits except in the case of 2-point Likert scale data. Results obtained are contrary to what Bernstein and Teng (1989) found, which was that the spurious factor increased with increase in number of Likert scale points. The study indicated that as the categories increased the spurious factors reduced. Therefore, the fewer the Likert scale points, the less accurate the results. It was found that categorization effects depended on the number of scale points, distribution of responses and the underlying factor.

A limitation of the study was that intervals in the Likert scale points were assumed to be equal with same thresholds. In reality, not all Likert scales have equal interval and incrementally equal thresholds, from the respondents' point of view, and response distribution patterns. Unequal interval scale data should have been generated to investigate this aspect of scale categorization, given that interval scales may be affected by wording and verbal anchor, and therefore different thresholds.

The strength of the study compared to that of Bernstein and Teng's study is that the results are supported by independent studies on correlation coefficients and ordinal categorization (Bollen & Barb, 1981) as well as studies on factor analysis and ordinal categorization (Babakus, Ferguson & Joreskog, 1987). The study has a significant methodological contribution as it addresses the effects of ordinal categorization with varying underlying distributions on the accurate identification of factors. Only two studies so far have addressed the two variables simultaneously, namely the Bernstein and Teng study, and the Green, Akey, Hershberger, and Marquis (1997) study.

Summary of Research Concerns

Research concerns that emerged from the studies reviewed on the effect of ordinal categorization and therefore Likert variables on correlation, factor analysis, and confirmatory factor analysis using structural equation modeling are that Pearson's r correlation was underestimated for low number of categories and uniform distribution. However, the estimation improved and was close to that of continuous variables for more Likert scale points than in fewer Likert scale points. Early studies such as that of Labovitz (1975) did not find any differences in the use of ordinal categories with those of continuous variables and concluded that ordinal categorical data could be analyzed as if continuous. This is can be explained by the fact that, besides the study's methodological limitations, ordinal categorical variables were hypothesized to be a monotonic transformation of the underlying continuous variable.

O'Brien (1979) found contradicting results to that of Labovitz and stated that the correlation between continuous variable and ordinal categorical versions of the same scale was not a monotonic function of the number of ordinal categories, as well as Likert scales, as had been previously proposed. O'Brien used only normal and uniformly distributed data. The effect of non-normal distribution of the data was not investigated and was recommended for future research. This has been investigated in the current study on linear regression models and Likert variables.

Bollen and Barb (1981) conducted a follow up study to that by O'Brien (1979) and found that the difference in the correlation of the continuous variable and the ordinal categorical variable decreased with the increase in the number of categories. The average standard deviation of the correlation in the categorized variable was greater than that of the correlation in the continuous variable. The standard deviation of the correlations decreased with the increase in number of categories. Thus, the number of categories has a meaningful impact on the correlation of the variables as demonstrated in the study by Bollen and Barb (1981). Moreover, the impact of ordinal categorization on Pearson correlation has important implications to psychometric properties inferred from instruments in terms of underestimated reliability coefficients. The present study differs from that of Bollen and Barb in that it is a case of ordinal categorization (resulting in Likert data) effects on multivariate regression models, compared to previous studies of the effect of ordinal categorization on Pearson's product moment correlation, which takes into account only one predictor and one criterion.

It was recommended in previous studies (Bollen & Barb, 1981) that studies be conducted on multiple correlations to address the effects of ordinal categorization on multiple correlation measures and precision. Results are expected to differ from the case of bivariate correlation measure (Pearson correlation), as the beta-weights for each predictor in the multiple correlation depend on other predictors in the regression equation. The modeled R^2 is the squared correlation between the criterion variable Y and optimal linear combination of the predictor variables X_i. It is for this reason that the multivariate case in Likert data was studied separately. Therefore, the present study addressed the effect of the number of Likert scale points in the model fit on linear regression models. However, with regard to distribution of the data, Bollen and Barb study also failed to address the issue of non-normality of the data. The present study extends beyond Bollen and Barb's study as both positive and negatively skewed distributions of responses were studied.

In studies involving structural equation modeling and confirmatory factor analysis, ordinal categorization of variables which resulted in Likert scales was observed to introduce errors, which in turn produce distortions in indicator models (Johnson & Screech, 1983). Errors, which are due to transformation, grouping and misclassification in categorization, attenuate the correlation of the variables involved. This results in distortion of the hypothesis and parameter estimates. The simulated model included three latent variables and six indicator variables, from which categorized variables were created. Conditions were created for low and high correlation of latent variables (beta coefficients) with low and high lambda coefficients. Measurement errors were also correlated. The test for linearity indicated that non-linearity was most pronounced for lower categories than for higher categories. When the correlation between latent variables and the indicator variables was high and uniformly distributed, the correlation among the residuals was also high, especially in the case of low number of categories. The fewer the categories, the lower the lambda, and the larger the variability in the estimates. Therefore, the number of categories was a significant determinant in the model fit, although the optimal number is not known. The study recommended that further investigation on the impact on the number of categories be conducted in future. Sample size was also noted to significantly affect parameter estimate, with large variability occurring at low sample sizes and fewer categories.

Important results in factor analysis studies involving Likert variables were the impact of ordinal categorization which led to identification of spurious factors (Bernstein & Teng, 1989; Green et.al 1997) and hence the wrong dimensionality. In Bernstein and Teng study, it was concluded that the increase in the number of Likert scale points resulted in more spurious factors, while in the Green et al study, it was concluded that the number of spurious factors reduced with the increase in number of Likert scale points. This led to conflicting findings that needed to be resolved. A research concern raised by the Green et. al, study was the common assumption of equal interval among Likert scales. While this is a common practice, the reality is that not all scales have uniform or equal interval between categories. Thus, the impact of unequal interval, which often occurs in practice, was addressed in the present study.

For structural equation modeling, the chi-square values were found to be inflated for fewer Likert scale points than for more scale points leading to poor model fit. In the study by Muthen and Kaplan (1992) in which models of various sizes were fit using confirmatory factor analysis, for non-normal distribution in Likert variables, it was found that the generalized least-square GLS method performed poorly compared to the other methods. The standard errors increased in lower sample sizes, and fewer Likert scale points for highly non-normal distributions. In the study by Curran, West and Finch (1996), factor loadings and factor correlations were found to be underestimated within fewer Likert scale points and skewed distribution than at more scale points. Evidently, the number of categories has a meaningful impact on the number of factors identified and correlation of the factors. In the studies reviewed, four variables were of major research concern and were examined in the current study. These were the effects of the number of categories in the variables, the underlying distribution, and the pattern of response for equal, and unequal interval.

An observed trend in the studies reviewed show that researchers moved from investigating the impact of Likert scales and categorical scales on Pearson correlation to factor analysis and structural equation modeling and were inclined to overlook multiple regression models. Researchers were fascinated by structural equation modeling and so structural equation modeling dominated the research topics in which polychoric and tetrachoric correlations were viewed as solutions to problems associated with correlations of discrete and categorical data in factor analysis. This instigated further work and discussion on the implication of Likert scales on factor analysis. Meanwhile, research consumers continued to use multiple linear regression models and hence the need to investigate the implication of using Likert data on linear regression models.

Although the studies reviewed were not conducted using regression models, they were essentially examined to facilitate the identification of appropriate variables that have been previously hypothesized to affect model fit in Likert data, and to provide direction for the methodological investigation of the current problem, as well as to answer the research questions posed. Pertinent research questions arose in the review of the effect of Likert data on Pearson correlation, especially with regard to the effect of response distribution. Therefore, due to these concerns, the present study addressed these questions and extended Bollen and Barb's (1981) study. The overall main concern was the impact of Likert data on the model fit in linear regression models in terms of bias in R^2 and relative importance of the predictors. Research questions were as follows:

Research Questions

1. What are the effects of Likert data on the estimation of R^2 across number of scale points, type of correlation matrices, response distributions and Likert conditions?
2. What are the effects of Likert data on variable ordering and relative importance as measured by relative Pratt index *d* across number of Likert scales, types of correlation matrix, response distributions and Likert conditions?
3. What are the effects of Likert data in the estimation of the Pearson Correlation coefficients across number of Likert scales, types of correlation matrix, response distributions and Likert conditions?

The last research question is an extension of Bollen and Barb's study. The following method was used to answer the research questions posed.

CHAPTER III

METHODOLOGY

Introduction

This was a simulation study in which the measurement process i.e. the process of responding to Likert scales was simulated. In essence, simulations were conducted to mimic the actual measurement process in responding to the Likert scales under controlled conditions enabling the researcher to determine how much information is lost. Thus, real data could not be used to realize this goal.

Little has been done in the area of Likert scales and regression models and so a large population of responses was simulated to adequately assess the influence of the independent variables stated in the research questions and to investigate the amount of bias resulting from the use of Likert data. The methodology has been adapted from similar studies on ordinal categorization of variables in correlation studies, factor analysis and structural equation modeling. Ordinal categorization of equal interval Likert scale points was adopted from the Bollen and Barb (1981) study. However the study goes beyond the Bollen and Barb study as it considers the effect of ordinal categorization resulting in Likert scaling on R^2 for different response distributions and, in particular, Likert scales in multiple regression models and not the bivariate cases as in the Bollen and Barb study. In addition to this, the present study examined the case of unequal interval Likert scale points for different response patterns and various combinations of Likert and continuous data across the predictors and criterion variable.

Three response patterns, three combinations of Likert conditions of both criterion and predictor variables that are either continuous or Likert, three types of correlation matrices depicting the relationship between the variables, and eight different scale points ranging from two to nine Likert scale points were studied using a fully crossed factorial design to assess the unique impact of each of the independent variables hypothesized.

Procedure

A program to generate a latent response set for 500,000 respondents was written to simulate the data. The resulting data set served as a population from which observed response conditions were generated. By using a large finite population the study sidesteps the matter of sampling variability and focused on the population-level results. For each combination of the conditions of responses, multiple ordinary least-square linear regression analysis was conducted for the predictors and criterion variable. Effect of Likert data on the regression model fit as well as information loss was analyzed through percent bias in the resulting R^2, relative Pratt index and Pearson correlation coefficients across the independent variables and conditions stated.

Study Design

The study design was as follows: Responses were generated using 3 correlation matrices, 3 response patterns, and 8 Likert scales, ranging from 2 to 9 scale-points. Ordinary least-square regression was conducted on 3 combinations of Likert data conditions and continuous data for predictors and criterion variable. This resulted in a 3x3x3x8 factorial design with 216 cells. Each cell consisted of the three dependent variables derived from the multiple linear regression analysis results.

Selection of Variables

To avoid problems associated with sampling, a population of 500,000 latent responses was simulated from which observed response patterns and conditions were manipulated. The independent variables were: patterns of responses, type of correlation matrix, combination of Likert conditions and continuous data, and Likert scale-points. The dependent variables were: percent bias in each of R^2, relative Pratt index, and Pearson correlation coefficients.

Response Patterns

Three response distributions, also referred to as response patterns, were simulated. These were (1) equal interval resulting in symmetric "normal looking" responses in the middle of the scale range ("Equal" or response pattern 1) as shown in figure 4, (2) unequal interval negatively skewed, resulting in responses bunched to the right of the scale ("Right" also referred as right bunching or response pattern 2), and (3) unequal interval positively skewed, resulting in responses bunched to the left of the scale ("Left" also referred to as left bunching or response pattern 3).

The first response pattern with equal interval scale points and symmetric distribution is similar to that used by Bollen and Barb (1981) in investigating the correlation coefficient. Thresholds in this pattern were simulated in a similar manner for all the conditions. Responses in this pattern were assumed to be normally distributed, and so across a standardized scale of $z = -3$ to $z = +3$, the scale points were divided equally for each ordinal item response process (see figure 4) in which the Likert scale points were simulated (see Table 5 in Appendix A). However, in the two cases of unequal interval, thresholds were set to generate responses that were positively and negatively skewed resulting in responses bunched to the left, and those bunched to the right.

Figure 4 depicts the thresholds for the symmetric response distribution with equal intervals. The top of Figure 4 lists a standard normal distribution and the bottom of Figure 4 lists three examples of the Likert scaling, namely 2-, 3-, and 4-point Likert scales. It is important to note that what is being divided in the response process is not the area under the Normal curve but rather the spatial distance along the continuum. What this represents is the item response model in which the response one provides depends on how much of the latent variable one possesses. For example, starting from the far left and using the 2-point scale, if one only has –1.5 standard units of the latent variable then they would respond "1" to the question. On the other hand, in the same context, if one has 0.5 standard units of the latent variable they would respond "2".

The thresholds in the pattern of responses that were negatively skewed (right bunching), were created using cut-off points on the continuum of a standard normal distribution for $z = -3$ to $z = +3$ as shown in Table 6 of Appendix B. For a two-point Likert scale, the threshold was set at the point, $z = -1.5$, below which the response value was 1 and above which the response value was 2. This was done for all the three predictors and criterion. Similarly, for a three-point Likert scale, two threshold points were created. The two scale points were $z = -1.5$ and $z = 0$. Thus, below $z = -1.5$, the response value was 1, between $z = 0$ and $z = -1.5$, the response values was 2, and above threshold $z = 0$, the response value was 3. In the case of positively skewed responses or bunched to the left pattern, the threshold points were created as shown in Table 7 in Appendix C.

As an illustration, using the previous example of the students' life satisfaction survey, the results of a pilot study with a simulated sample of 1000 subjects, yielded the histograms for selected response patterns based on the Likert scale points indicated. Figure 5 is the histogram for the Y criterion responses with a normal distribution and a continuous scale response. Figures 6 to 9 are histograms for 9 scale points, 8 scale points, 6 scale points and 4 scale points for equal intervals respectively. Figures 10 to 12 are histograms for negatively skewed response pattern bunched to the right for 9, 8 and 4 scale points Shown in Figure 13 is a histogram for responses in a criterion variable Y with eight scale-point responses bunched to the left or positively skewed.

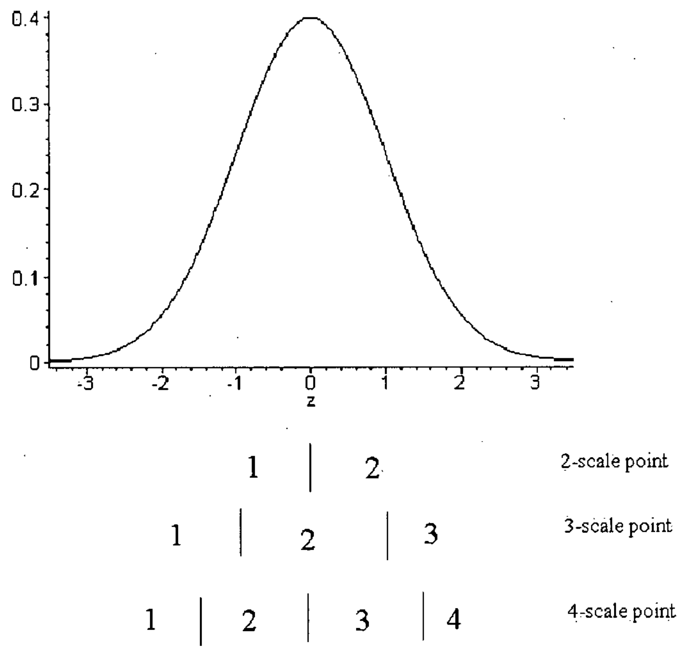

Figure 4

Likert responses from equal interval thresholds.

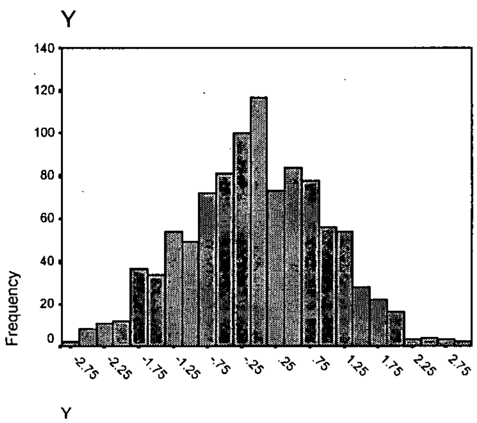

Figure 5

Histogram for a continuous Y scale normally distributed response pattern.

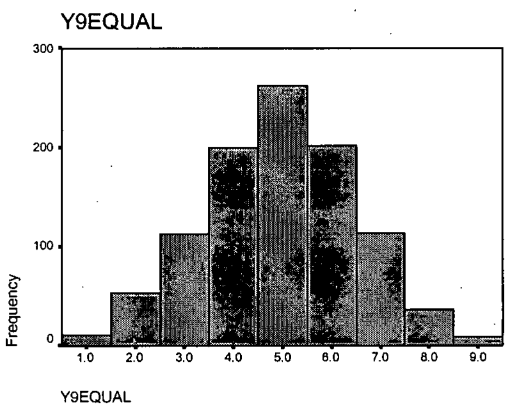

Figure 6

Histogram for 9 scale-point equal interval response pattern.

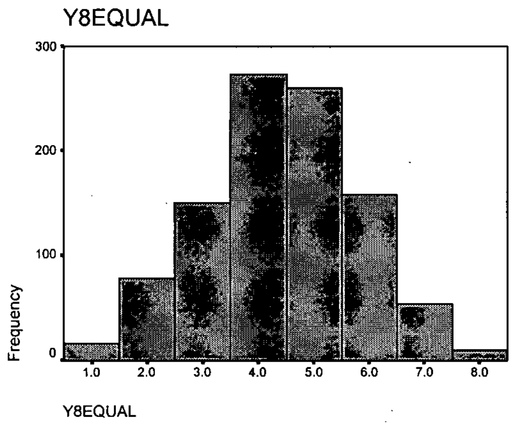

Figure 7

Histogram for 8 scale-point equal interval response pattern.

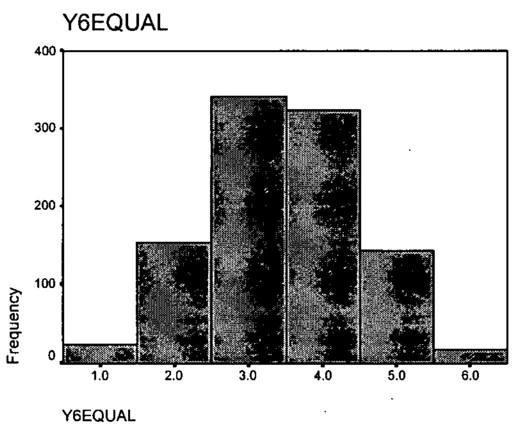

Figure 8

Histogram for 6 scale-point equal interval response pattern.

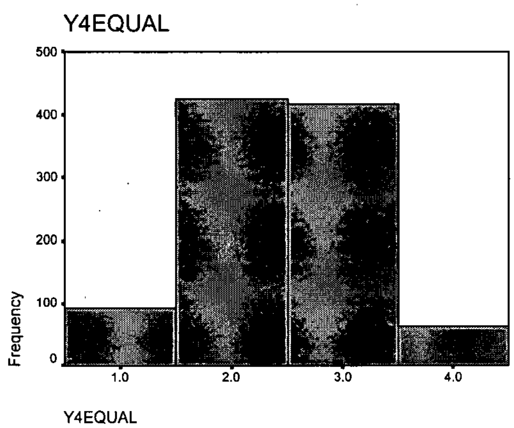

Figure 9

Histogram for 4 scale-point equal interval response pattern.

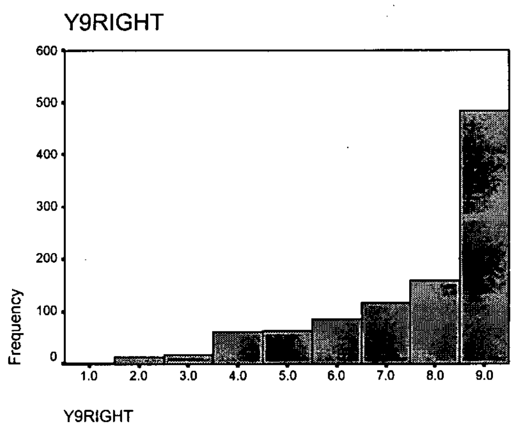

Figure 10

Histogram for 9 scale-point unequal interval response pattern bunched to the right.

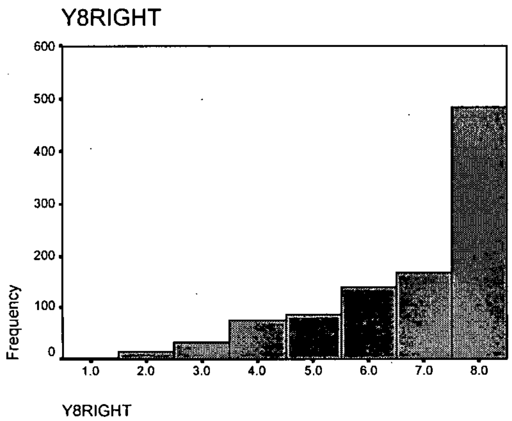

Figure 11

Histogram for 8 scale-point unequal interval response pattern bunched to the right.

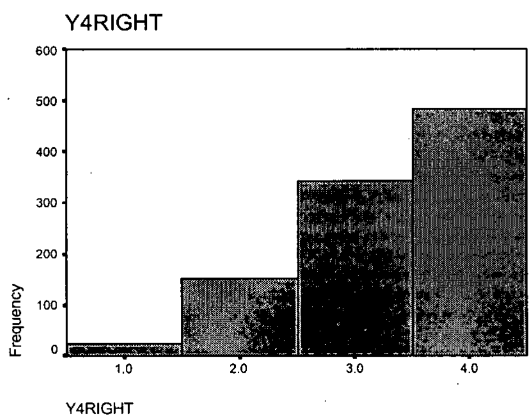

Figure 12

Histogram for 4 scale-point unequal interval response pattern bunched to the right.

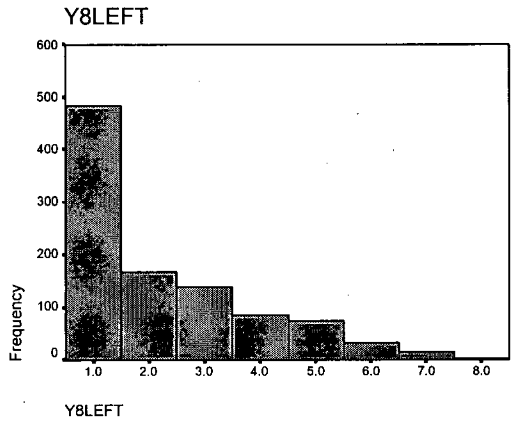

Figure 13

Histogram for 8 scale-point unequal interval response pattern bunched to the left.

Correlation Matrices

Correlation matrices selected for simulating the population of responses were based on the typical occurrence in social science research of the relationship among predictors and criterion variables, in which (1) predictors are moderately correlated with each other and low correlation with the criterion, (2) predictors have a low correlation with each other but a moderate correlation with the criterion, and (3) predictors have a moderate to high correlation with themselves and moderately high correlation with the criterion variable. The correlation matrices were based on examples of the relationship between predictors and criterion variables often found in educational settings and social science research (Stevens, 1986). These were used in the estimation of R^2, beta-weights, and Pearson correlation coefficients between each predictor and criterion. The resulting beta-weights and the Pearson correlation coefficients were used to compute the relative Pratt index for each predictor under each combination of conditions.

It was hypothesized that the type of correlation matrix together with Likert variables would impact the estimation of R^2, relative Pratt index, and Pearson correlation coefficient for each combination of the conditions of the independent variables. Shown in Table 1 is the correlation matrix for predictors with low inter-variable correlation with the criterion, but moderate inter-variable correlation among themselves. This matrix is referred to as the low inter-variable correlation matrix (Correlation matrix 1) in the study. The resulting R^2 and beta weights of the model fit of the matrix for continuous data are shown immediately below Table 1. Also shown are measures of variable importance namely, the relative Pratt indices for each predictor variable in the correlation matrix.

Table 1

Correlation matrix with low inter-variable correlation of predictors and the criterion, and moderate inter-variable correlation among the predictors.

	Y	X_1	X_2	X_3
Y	1			
X_1	.20	1		
X_2	.10	.50	1	
X_3	.30	.40	.60	1

R-squared = 0.118

Beta1 = 0.157, Pratt 1 = 0.269

Beta2 = -0.188, Pratt 2 = -0.158

Beta3 = 0.350, Pratt 3 = 0.890

Table 2 shows the correlation matrix for which the predictors have low inter-variable correlation among themselves but moderate inter-variable correlation with the criterion variable. This matrix is referred to as the moderate inter-variable correlation matrix (Correlation matrix 2) in the study. The resulting R^2, beta weights as well as relative Pratt indices derived from the model fit with continuous data in the pilot study are shown immediately below the matrix in Table 2.

Table 2

The correlation matrix with moderate inter-variable correlation of predictors and the criterion, and low inter-variable correlation among predictors.

	Y	X_1	X_2	X_3
Y	1			
X_1	.60	1		
X_2	.50	.20	1	
X_3	.70	.30	.20	1

R-squared = 0.753

Beta1 = 0.378, Pratt 1 = 0.3020

Beta2 = 0.321, Pratt 2 = 0.2136

Beta3 = 0.522, Pratt 3 = 0.4853

The third correlation matrix shown in Table 3 depicts a situation where the predictors have moderate to high inter-variable correlation among themselves and moderately high with the criterion variable. This is referred to as the high inter-variable correlation matrix (Correlation matrix 3) in the study. The resulting R^2, beta-weights, and relative Pratt indices of predictors in the model fit for the continuous data in the pilot study are shown immediately below Table 3.

Table 3

The correlation matrix with moderately high inter-variable correlation of predictors and the criterion, and moderate to high inter-variable correlation among the predictors

	Y	X_1	X_2	X_3
Y	1			
X_1	.60	1		
X_2	.70	.70	1	
X_3	.70	.60	.80	1

R-squared = 0.562

Beta1 = 0.185 Pratt 1 = 0.1972

Beta2 = 0.275 Pratt 2 = 0.3425

Beta3 = 0.370 Pratt 3 = 0.4069

Throughout the study the type of correlation matrix as a variable is also referred to as CORMTX with three levels namely low, moderate, and high inter-variable correlation matrices. In the study, low inter-variable correlation matrix is also referred to in the results and discussion as correlation matrix 1 (low correlation matrix), and moderate inter-variable correlation is referred to as moderate correlation matrix or correlation matrix 2, whereas high inter-variable correlation matrix is also referred to as high correlation matrix or correlation matrix 3.

Variable Combination of Likert conditions

Three variable combinations of Likert conditions of the predictors and criterion variables were studied. The case of continuous predictors and continuous criterion variables were examined for each population generated from the correlation matrices. The results of the regression analysis in the continuous data served as baseline against which the other conditions were compared in order to assess information loss across the stated conditions. The three variable combinations of Likert conditions (VARCOMB) referred to are variable combination 1, variable combination 2, and variable combination 3, and were defined as follows:

(1) When all predictors are continuous and the criterion is Likert (Variable combination 1).

(2) When all predictors are Likert and the criterion is continuous (Variable combination 2).

(3) When all predictors are Likert and the criterion is also Likert (Variable combination 3).

Regression analyses were limited to the following situation (which was seen in the example from the students' life satisfaction study at the beginning).

For the left bunching of the responses (positively skewed response distribution):

$$Ycont_{normal} = \beta_1 X_1 left + \beta_2 X_2 left + \beta_3 X_3\ left \qquad (16)$$

$$Yleft = \beta_1 X_1 cont_{normal} + \beta_2 X_2 cont_{normal} + \beta_3 X_3 cont_{normal} \qquad (17)$$

$$Yleft = \beta_1 X_1 left + \beta_2 X_2 left + \beta_3 X_3\ left \qquad (18)$$

Importantly, it should be noted from above that the patterns in the regressions are not mixed. Left and right bunching responses were ordinal categorical with Likert conditions, while the continuous predictors and criterion referred to were continuous and normally distributed. Similar regression models were fit for right bunching (negatively skewed response distribution) and for the equal interval symmetric distribution of responses.

Likert scale points

For all conditions, the Likert variables ranged from 2 to 9 scale points, which are commonly found in questionnaires and instruments in social science research. It was hypothesized as in the previous studies, that the number of scale points affects parameter estimates. The number of scale points was created through the simulation of responses with specified thresholds shown in

Appendix A, B, and C (See Table 5, 6 and 7 respectively). A schematic representation of the study design with all the independent variables is shown on Table 4.

Table 4

Schematic representation of the study design for the population generated from each of the three correlation matrices (Low, Moderate and High inter-variable correlation matrices) response patterns and Likert conditions

COMBINATIONS (VARCOMB)	Variable Combination 1 (Likert criterion Y and Continuous predictors X)			Variable Combination 2 (Continuous criterion Y and Likert predictors X)			Variable Combination 3 (Likert criterion Y and Likert predictors X)		
PATTERNS (Response patterns)	Equal	Right	Left	Equal	Right	Left	Equal	Right	Left
CATEGORIES (Likert Scale points)	2-9	2-9	2-9	2-9	2-9	2-9	2-9	2-9	2-9

Levels of categories (number of Likert scale points), other than the continuous one, were selected because low number of categories (coarse categorization) were previously hypothesized in correlation studies, to result in large standard errors of Pearson's correlation (Bollen & Barb, 1981) and spurious factors in both factor analysis and confirmatory factor analysis (Green et.al, 1997; Johnson & Screech, 1983), while high number of categories (refined categorization of scale) resulted in low standard errors in the correlation estimates and fewer spurious factors. These findings were disputed by Bernstein and Teng (1989) and resulted in a follow-up study by Green et al (1997).

Because of the conflicting results in the correlation studies regarding the effect of number of Likert points, it was prudent to examine a wide range of Likert scale points from low to high. Based on the literature, it was therefore hypothesized that R^2 would be impacted by fewer number of scale points than more scale points.

Both equal and unequal interval Likert scales were addressed in the present study, as suggested in previous studies by Bernstein and Teng (1989) as well as Green et al (1997). This is because all previous studies assumed equal interval and normal distribution between categorical scales in responses. In reality not all ordinal categorical response scales are of equal interval and normal distribution. This is a limitation of the commonly used Likert scales, which assumes equal interval in categories. Each of the dependent variables was analyzed separately to assess the specific effects of ordinal categorization (Likert scaling) on each measure and an overall implication to the linear regression model and Pearson correlation.

Relative Pratt index and Variable importance

In performing regression analysis researchers and stakeholders often ask which predictors are important, and which ones contribute most to the estimation of the criterion variable. Several indices have been proposed in determining relative importance of predictors in regression models among which are the relative Pratt index (Thomas, Hughes and Zumbo, 1998), Beta weights, t-values, commonality components, semi-partial correlation sr_j and partial correlations pr_j. However, Pedhazur (1982) and Darlington (1990) have suggested that beta weights and partial correlations should not be used on their own as measures of variable importance as the two statistics lack additive and proportionality properties within the criterion variance. Therefore, the relative Pratt index with its additive property and which sums up to one for all the predictors in the equation emerges as the most suitable measure of relative importance. Relative importance of variable X_j in a regression model is determined by the proportion of the variance in the criterion variable accounted for by X_j (Bring, 1994, 1996; Kruskall, 1987; Pratt, 1987; Thomas, Hughes, & Zumbo, 1998).

Because of its additive and proportionality property as well as simplicity in interpretation and computation, the relative Pratt index lends itself as a promising and prudent choice of an index to use in determining relative importance and ordering of variables (Thomas, Hughes, & Zumbo, 1998; Thomas & Zumbo, 1996). The relative Pratt index is based on the axiomatic derivation of the product of the simple correlation and beta coefficient of a variable X_j from the regression equation of the form:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_j X_j. \qquad (19)$$

The relative Pratt index d_j associated with predictor X_j is then computed as follows:

$$d_j = \frac{\beta_j r_j}{R^2} \qquad (20)$$

The resulting quotient is the proportion of model R^2 accounted for by the predictor variable X_j while β_j is the beta weight associated with the predictor X_j, and r_j is the Pearson correlation between X_j and the criterion variable Y. Thus, the relative Pratt index is used in variance partitioning of R^2 for each predictor.

Values of relative Pratt index d_j that meet the condition;

$$d_j > \frac{1}{2p}, \qquad (21)$$

where p is the number of predictors in the equation, are the only ones that are usually interpreted (Thomas, et al. 1998). While relative Pratt index has been endorsed as an appropriate measure of relative importance of predictor variables in linear regression models (Thomas, et al. 1998), little is known of the effect of Likert data on its accuracy, consistency and performance under different conditions of Likert scale and response distribution. The present study addressed this concern through graph plots and response modeling of the percent bias in relative Pratt indices for each combination of the independent variables. A similar procedure of data analysis was adopted for percent bias in R^2 and Pearson correlation coefficients for each condition. What follows are the results of the data analysis.

CHAPTER IV

RESULTS

Introduction

To answer the research questions, analysis was conducted separately for R^2, relative Pratt index, and Pearson correlation coefficient to determine the effect of Likert data on these dependent variables and the overall effect on the linear regression models. For each of these dependent variables a statistical model was fit using the independent variables; namely, number of Likert points, type of correlation matrix, variable combination, and distribution of response patterns.

For each dependent variable, the percent bias (relative to the continuous case) was used as the object of analysis. This contrasting to the continuous case reflects the simulation methodology's focus on mimicking measurement. Zumbo and Zimmerman (1993) introduced a simulation strategy whereby one mimics the measurement process and therefore is able to experimentally study the effects of measurement on data analysis. This dissertation follows the Zumbo and Zimmerman strategy (also see Nanna & Sawilosky, 1998). Whereas most simulation studies have tended to present only graphical plots and tabular displays, statistical modeling of the simulation output was conducted in the present study to provide (1) statistical evidence of the effects, and (2) further insight into the potential interaction effects of the independent variables (Zumbo & Harwell, 1999). That is, given that there are potentially several rather complex (i.e., of high order) interaction effects among the independent variables, the statistical modeling allows one to select a most parsimonious description of the results. Note that all of the independent variables were treated as categorical explanatory variables in the modeling except for the number of Likert scale points, which was treated as a quantitative variable in the modeling.

As Zumbo and Harwell state, the fundamental premise of using the statistical modeling is that simulation plays the same role in methodological research as experimentation plays in other sciences (see also, Hoaglin & Andrews, 1975). Note, however, that for the present simulation the statistical modeling serves a descriptive purpose. That is, the output from the simulation is a population result (based on a population size of 500,000) and hence will have little variation upon replication. The statistical model fit, hence, aids in describing the potentially complex response surface (i.e. hyperplane) of the simulation results (Box & Draper, 1987; Khuri & Cornell, 1987; Snee & Hare, 1992).

In what follows each dependent variable will be discussed in order of R^2, relative Pratt index, and the Pearson correlation.

Effect of Likert data on R-squared

Although it was used for all of the dependent variables, the percent bias will be described in detail only for R^2. To evaluate the effect of Likert data on R^2 resulting from the linear regression model fit, differences between R^2 derived from the regression model based on the experimental conditions and those derived from the baseline model with continuous predictors and continuous criterion variables were computed. Bias was defined by the expression:

$$Bias\Delta = R^2_{mod} - R^2_{cont} \qquad (22)$$

where R^2_{mod} is the value of R^2 from the model with Likert condition and R^2_{cont} is the value of R^2 of the model with continuous data. Given that various values of R^2_{cont} were used in the model, it was considered prudent to transform differences in R^2 (bias) as a percentage of the R^2 from the continuous baseline regression model with continuous variables. The resulting quotient was referred to as percent bias and is computed as follows:

$$Percent\ bias\Delta R^2 = \{R^2_{mod} - R^2_{cont}\} / R^2_{cont} \times 100. \qquad (23)$$

Statistical Model fit

A statistical model was fit for percent bias in R^2 as the dependent variable. As evident in Table 8, there were main effects for number of Likert points (CATEGORY), response pattern (PATTERN), and variable combination (VARCOMB). The main effect of correlation matrices was, strictly speaking, not statistically significant. However, given that the p-value was close to the cut-off of 0.05, this effect was considered to be marginally significant and worthy of consideration. Effect sizes of the variables are as follows: for the correlation matrices, eta squared are 0.035 (small), for variable combination, eta squared is 0.062 (moderate), while for response pattern distribution eta squared is 0.046 (small). In the case of number of Likert scale points (CATEGORY), eta squared is 0.520 (large). The effect size of Likert points (CATEGORY) is much larger than those of the other independent variables. Thus, number of Likert points seems to have a larger effect than the rest of the variables. The following criteria, based on Cohen (1992) and Kirk (1996), was used to interpret effect sizes: 0.01 to 0.058 small, 0.059 to 0.137 moderate, and greater than 0.138 as large. All two, three and four-way interactions of the explanatory variables in the statistical model fit were not significant. To assess the effects of the explanatory variable further, post hoc tests were conducted for each of the significant main effects.

Table 8

Statistical model fit for Percent bias in R-square for type of matrix, variable combination, response pattern, and Likert scale points

Source	df	Mean Square	F	Sig.
CORMTX	2	421.666	2.974	.054
VARCOMB	2	761.444	5.371	.006
PATTERN	2	556.993	3.929	.022
CATEGORY	1	24892.208	175.586	.000
CORMTX * VARCOMB * PATTERN * CATEGORY	8	4.080	.029	1.000
CORMTX * VARCOMB * PATTERN	16	29.297	.207	1.000
CORMTX * VARCOMB * CATEGORY	4	11.936	.084	.987
CORMTX * PATTERN * CATEGORY	4	15.859	.112	.978
VARCOMB * PATTERN * CATEGORY	4	13.001	.092	.985
VARCOMB * PATTERN	4	46.642	.329	.858
VARCOMB * CATEGORY	2	312.417	2.204	.114
PATTERN * CATEGORY	2	.385	.003	.997
Error	162	141.767		
Total	216			

For the independent variable correlation matrices, pairwise comparisons were conducted between the means of percent bias of R^2 at low, moderate and high correlation matrices. Figure14 displays the mean percent bias in R^2 for the three levels of the correlation matrix CORMTX. There were statistically significant differences in the means of percent bias between low inter-variable correlation matrix and moderate correlation, and low inter-variable correlation matrix and high inter-variable correlation matrix ($p<0.05$). However, there were no significant differences in the mean percent bias of R^2 between moderate and high correlation matrices.

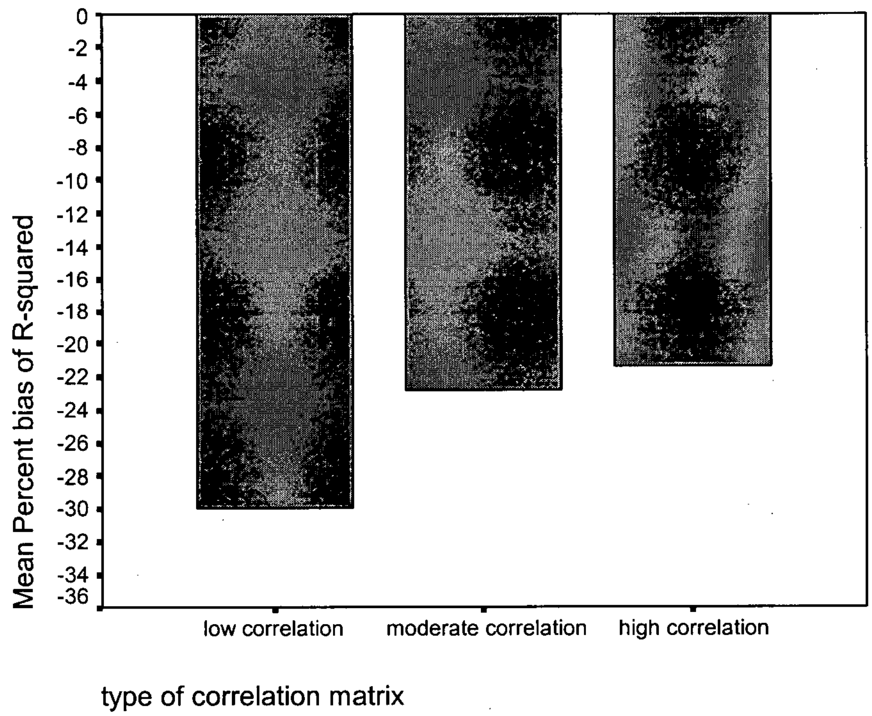

Figure 14

Bar Graph of the main effect of type of correlation matrix for percent bias in R-squared.

Post hoc tests were done on the mean percent bias for the main effect of variable combinations. Figure 15 displays the mean percent bias for the various variable combinations. For the pairwise comparisons of the mean percent bias, there were statistically significant differences between means in variable combination 2 (Continuous Y and Likert X) and variable combination 3 (Likert Y and Likert X), using a Bonferroni correction at a nominal alpha of 0.05. There were no significant differences in the means of percent bias for the other pairwise comparisons.

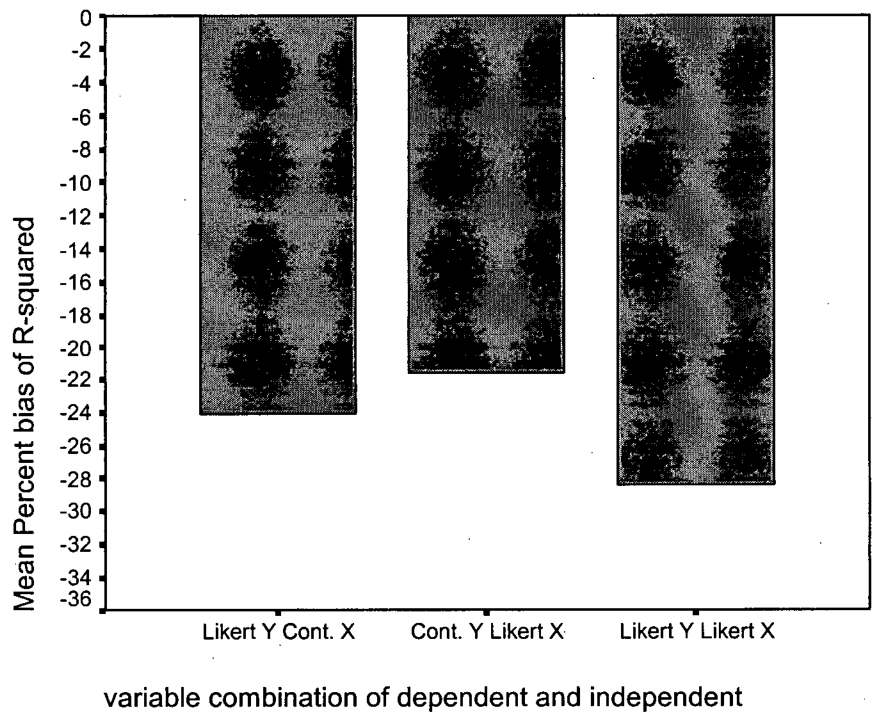

Figure 15

Bar Graph of the main effect of type of variable combinations for percent bias in R-squared.

For the main effects on response patterns, post hoc tests were again done through pairwise comparison of the mean percent bias of R^2 among the three response patterns at a nominal alpha of 0.05 using a Bonferroni correction. Figure 16 displays the mean percent bias in R^2 for the three response patterns. There were statistically significant differences between response pattern 1(equal interval) and 2 (right bunching), and response pattern 1 and 3 (left bunching). There were no statistically significant differences between response pattern 2 and 3.

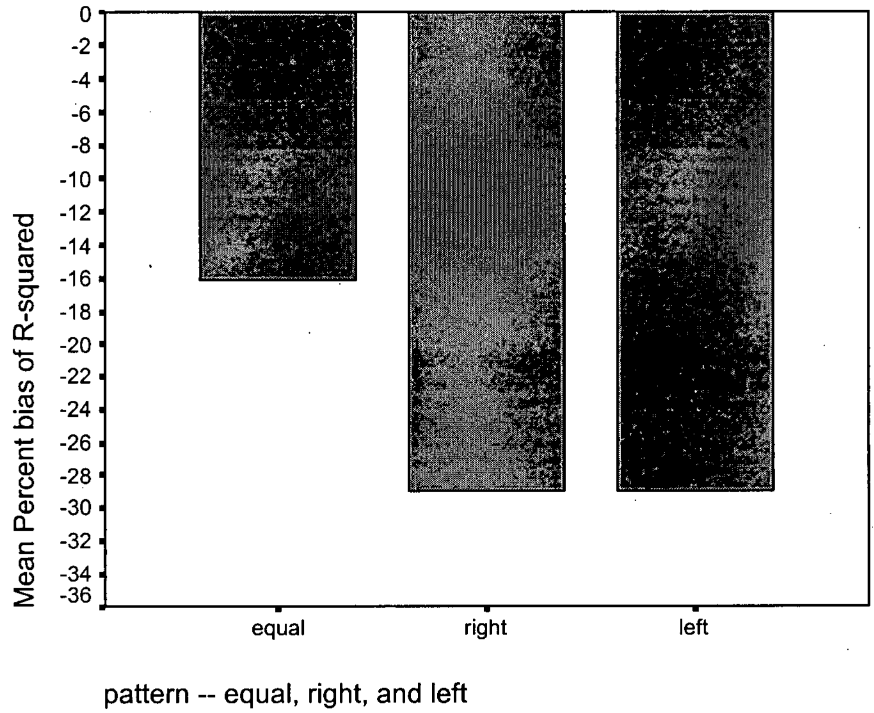

Figure16

Bar Graph of the main effect of response patterns for percent bias in R-squared.

Post hoc tests for the main effects of number of Likert points on percent bias of R^2 were conducted through pairwise comparison of the scales at 95% confidence interval. The results are summarized in Table 9. There were statistically significant differences between the 2-point Likert scale and the rest of the scale points. Similarly, there were statistically significant differences between the 3-point Likert scale and Likert scale points of 6, 7, 8 and 9. However, there were no statistically significant differences between the 3-point Likert scale, and the 4-point Likert scale as well as 3-point Likert scale and 5-point Likert scale. Comparison of the 4-point Likert scale with the rest of the scales from 5-point scale to 9-point scale showed no significant differences. Table 9 provides a summary of the mean and confidence interval bounds at 95%. Note that an asterisk (*) denotes a statistically significant comparison. Figure 17 depicts the relationship between number of Likert scale points and the percent bias in R^2.

Incorporating the results from Table 9, it is noted that the bias flattens at the four-point Likert scale and beyond.

Figure 17 provides a graphical display of how the mean percent bias in R-squared reduces as the number of scale points increase and flattens at four Likert scale points.

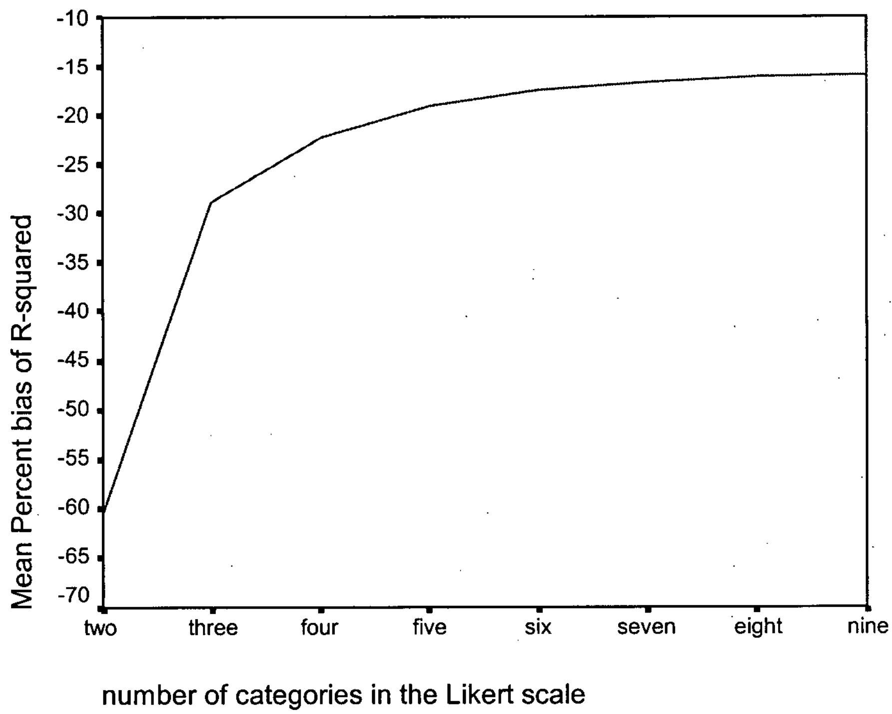

Figure 17

Main effect of number of Likert scale points on percent bias of R-Squared.

Table 9

Confidence intervals of the mean differences of percent bias in R^2 for all Likert scale points

COMPARISON	MEAN1 - MEAN2	95% CI OF DIFFERENCE
1: 2v3 *	-31.62559	-41.91889 to -21.33229
2: 2v4 *	-38.28406	-48.57736 to -27.99076
3: 2v5 *	-41.49404	-51.78734 to -31.20074
4: 2v6 *	-42.97422	-53.26752 to -32.68092
5: 2v7 *	-43.81377	-54.10707 to -33.52047
6: 2v8 *	-44.29653	-54.58983 to -34.00323
7: 2v9 *	-44.58969	-54.88299 to -34.29639
8: 3v4	-6.65777	-16.95107 to + 3.63553
9: 3v5	-9.86845	-20.16175 to + 0.42485
10: 3v6 *	-11.34863	-21.64193 to -1.05533
11: 3v7 *	-12.18818	-22.48148 to -1.89488
12: 3v8 *	-12.67094	-22.96424 to -2.37764
13: 3v9 *	-12.96410	-23.25740 to -2.67080
14: 4v5	-3.21074	-13.50404 to + 7.08256
15: 4v6	-4.69092	-14.98422 to + 5.60238
16: 4v7	-5.53047	-15.82377 to + 4.76283
17: 4v8	-6.01323	-16.30653 to + 4.28007
18: 4v9	-6.30639	-16.59969 to + 3.98691
19: 5v6	-1.48018	-11.77348 to + 8.81312
20: 5v7	-2.31973	-12.61303 to + 7.97357
21: 5v8	-2.80249	-13.09579 to + 7.49081
22: 5v9	-3.09565	-13.38895 to + 7.19765
23: 6v7	-0.83955	-11.13285 to + 9.45375
24: 6v8	-1.32231	-11.61561 to + 8.97099
25: 6v9	-1.61547	-11.90877 to + 8.67783
26: 7v8	-0.48276	-10.77606 to + 9.81054
27: 7v9	-0.77592	-11.06922 to + 9.51738
28: 8v9	-0.29316	-10.58646 to + 10.00014

* significant at $p < 0.05$

Effect of Likert data on the relative Pratt index

The purpose of computing the relative Pratt index is to determine variable importance among the set of predictors in a regression model. This serves as a criterion of ordering the variables or examining relative importance in the regression model. To evaluate the effect of Likert data on the measures of variable importance in the linear regression models, the relative Pratt index was computed from the Pearson correlation coefficients and beta weights for each predictor variable and for each regression model fit. The difference between the models' relative Pratt indices based on Likert data, and the corresponding baseline relative Pratt indices for continuous data, was then transformed into percent bias of the relative Pratt index for each predictor variable under the three treatment conditions and Likert scales. The relative Pratt index associated with variable X_1 was referred to as Pratt 1 index, that associated with variable X_2 was referred to as Pratt 2 index, and that associated with X_3 was referred to as Pratt 3 index. For each predictor variable, the computed percent bias of the relative Pratt indices were analyzed separately to assess the effect of Likert data on the three indices under the conditions of the stated independent variables. It was hypothesized that percent bias in the relative Pratt index would affect variable importance and therefore variable ordering of the predictors under the stated conditions of type of correlation matrix, response distribution, variable combination and number of Likert scale points.

Statistical Model fit

Separate statistical model fit were conducted for percent bias in Pratt 1, Pratt 2, and Pratt 3, respectively.

Pratt Index One

The results of the model fit for the percent bias in Pratt 1 is shown in Table 10. First, it is important to note that the main effect of response distribution PATTERN was not statistically significant, and nor were any interactions involving this independent variable. There was a statistically significant three-way interaction of CORMTX by VARCOMB by CATEGORY. It will be evident in the results listed below that due to the type of interaction, the main effects will not be interpreted.

Given the three-way interaction, a simple effects analysis was conducted to decompose the interaction. That is, at each level of VARCOMB (variable combination) a two-way model was fit for CORMTX by CATEGORY – indicating that the three-way interaction implies that the two-way interactions depend on the values of the third independent variable, VARCOMB. Table 11 lists the three two-way interactions. At variable combination 1 (Likert Y and Continuous X), interaction effects of CORMTX by CATEGORY were not significant. However, at Variable combination 2 and 3, the interaction effects of CORMTX by CATEGORY were statistically significant. Figures 18 to 20 show the interaction plots for the three two-way interactions listed in Table 11. From the Figures it is evident that for:

(a) Variable combination 1 (Likert Y and Continuous X) there was no bias (Figure 18).
(b) Variable combination 2 (Continuous Y and Likert X) the high correlation case was substantially more biased than the low and moderate correlation cases across the Likert scale points. Furthermore, the percent bias at two Likert scale points was substantially more for the high correlation case than for the low and moderate correlation cases (Figure 19).
(c) Variable combination 3 (Likert Y and Likert X) there was little to no bias for the low and moderate correlation cases irrespective of the number of scale points. However,

for the high correlation case the percent bias peaks at three Likert scale points and then tends toward zero as the number of scale points increases (Figure 20).

Table 10

Statistical model fit for the percent bias in Pratt1 for type of correlation matrix, variable combination, response pattern and number of Likert scale points

Source	Sum of Squares	df	Mean Square	F	Sig.
CORMTX	998.963	2	499.482	105.531	.000
VARCOMB	892.175	2	446.087	94.250	.000
PATTERN	4.014	2	2.007	.424	.655
CATEGORY	238.577	1	238.577	50.407	.000
CORMTX * VARCOMB	698.046	4	174.512	36.871	.000
CORMTX * PATTERN	4.258	4	1.064	.225	.924
CORMTX * CATEGORY	254.934	2	127.467	26.931	.000
VARCOMB * PATTERN	46.461	4	11.615	2.454	.048
VARCOMB * CATEGORY	243.531	2	121.765	25.727	.000
PATTERN * CATEGORY	9.002	2	4.501	.951	.389
CORMTX * VARCOMB * PATTERN	4.872	8	.609	.129	.998
CORMTX * VARCOMB * CATEGORY	181.353	4	45.338	9.579	.000
CORMTX * PATTERN * CATEGORY	14.854	4	3.714	.785	.537
VARCOMB * PATTERN * CATEGORY	4.710	4	1.178	.249	.910
CORMTX * VARCOMB * PATTERN * CATEGORY	9.249	8	1.156	.244	.982
Error	766.750	162	4.733		
Total	8327.964	216			

Table 11

Statistical model fit at variable combination levels for Correlation Matrix by Category for Pratt 1

Variable combination of dependent and independent	Source	Sum of Squares	df	Mean Square	F	Sig.
Likert Y Cont. X	CORMTX * CATEGORY	.709	2	.355	1.368	.262
Cont. Y Likert X	CORMTX * CATEGORY	334.405	2	167.202	10.903	.000
Likert Y Likert X	CORMTX * CATEGORY	101.173	2	50.587	13.455	.000

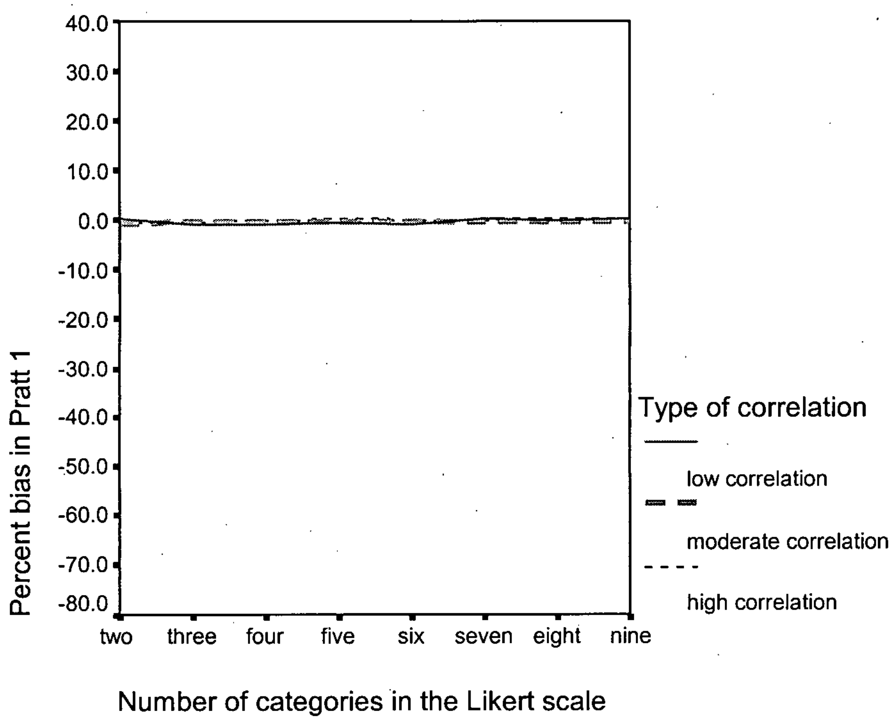

Figure 18

Percent bias in Pratt 1 for variable combination 1 (Likert Y and Continuous X) within type of correlation matrix and number of Likert scales.

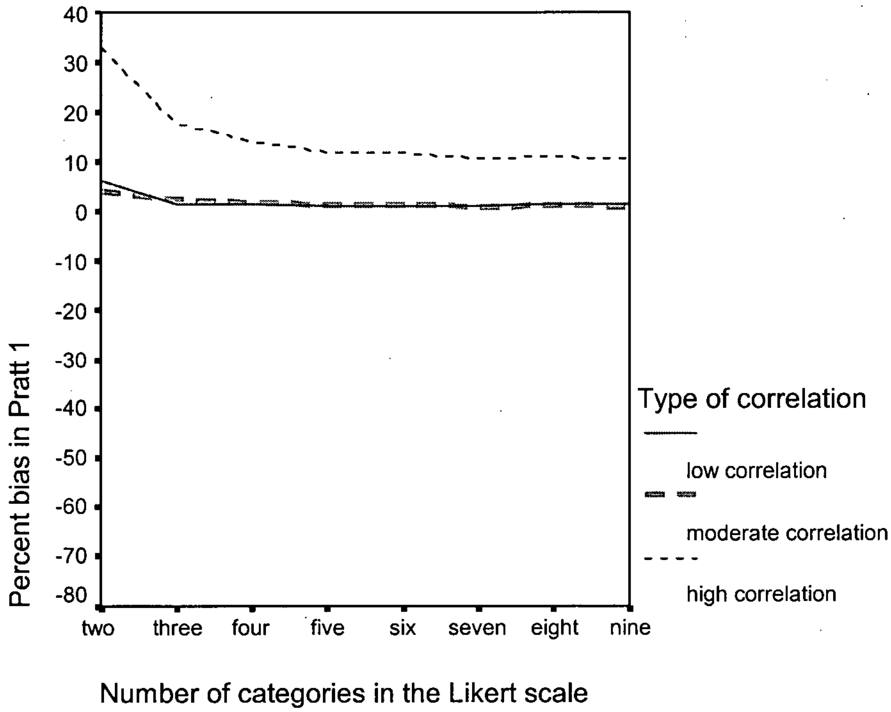

Figure 19

Percent bias in Pratt 1 for variable combination 2 (Continuous Y and Likert X) within type of correlation matrix and number of Likert scales.

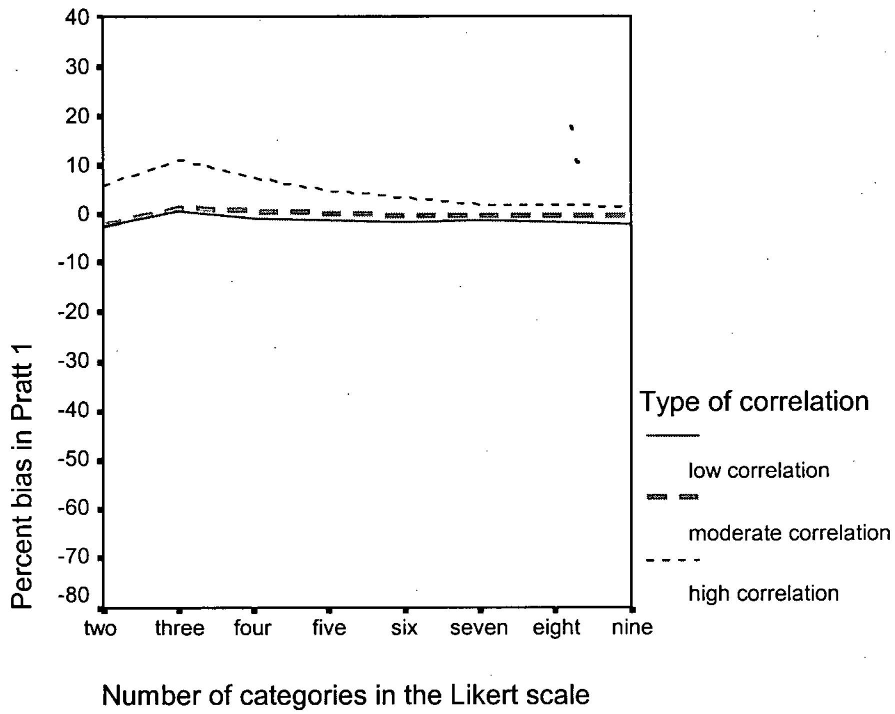

Figure 20

Percent bias in Pratt 1 for variable combination 3 (Likert Y and Likert X) within correlation matrix and number of Likert scales.

Pratt Index Two

The same analyses as Pratt 1 were conducted for Pratt 2. Note, however, as reported in the methodology Chapter, Pratt 2 is negative in the low correlation matrix case. The magnitude of the percent bias for low correlation is larger than in Pratt 1 for the same correlation matrix. The statistical model fit (shown in Table 12) results in a statistically significant 3-way interaction of CORMTX by VARCOMB by CATEGORY. Therefore, the simple effects two-way models were fit for each level of VARCOMB. The results of the 2-way interaction effects are listed in Table 13 and the interactions are plotted in Figures 21 to 23. From the Tables and Figures it is evident that for:

(a) Variable combination 1 (Likert Y and Continuous X) the Table and Figures give contradictory interpretations. That is, although the test of the two-way interaction shows a statistically significant interaction, the plot in Figure 21 indicates that there is no interaction – in fact, the Figure indicates that there was no bias across all of the number of scale points (Figure 21). This apparent contradiction can be accounted for by noting that the sum of squares of the interaction term was, relative to the other two interactions in Table 13, quite small. This information indicate that the magnitude of the interaction effect was moderate (an eta-squared of 0.128).

(b) Variable combination 2 (Continuous Y and Likert X) the high and moderate correlation cases showed very little percentage bias at two Likert scale points where the moderate case was slightly more biased. For the low correlation case, however, there is 80 percent negative bias at two Likert scale points and nearly 20 percent negative bias at nine scale points (Figure 22).

(c) Variable combination 3 (Likert Y and Likert X) was similar to Variable combination 2 except that the difference in percent bias was larger than in Variable combination 2 between moderate and high correlation cases for at two scale points (Figure 23).

Table 12

Statistical model fit for percent bias in Pratt 2 for type of correlation matrix, variable combination, response pattern and number of Likert scale points

Source	Sum of Squares	Df	Mean Square	F	Sig.
CORMTX	16744.062	2	8372.031	217.769	.000
VARCOMB	3717.458	2	1858.729	48.348	.000
PATTERN	19.930	2	9.965	.259	.772
CATEGORY	2492.421	1	2492.421	64.832	.000
CORMTX * VARCOMB	8903.093	4	2225.773	57.896	.000
CORMTX * PATTERN	11.252	4	2.813	.073	.990
CORMTX * CATEGORY	5568.378	2	2784.189	72.421	.000
VARCOMB * PATTERN	13.505	4	3.376	.088	.986
VARCOMB * CATEGORY	1182.580	2	591.290	15.380	.000
PATTERN * CATEGORY	14.154	2	7.077	.184	.832
CORMTX * VARCOMB * PATTERN	61.384	8	7.673	.200	.991
CORMTX * VARCOMB * CATEGORY	2976.225	4	744.056	19.354	.000
CORMTX * PATTERN * CATEGORY	46.397	4	11.599	.302	.876
VARCOMB * PATTERN * CATEGORY	5.876	4	1.469	.038	.997
CORMTX * VARCOMB * PATTERN * CATEGORY	25.461	8	3.183	.083	1.000
Error	6228.015	162	38.445		
Total	74987.212	216			

Table 13

Statistical model fit for variable combination levels for Correlation matrix by Category for Pratt 2

Variable combination of dependent and independent	Source	Sum of Squares	df	Mean Square	F	Sig.
Likert Y Cont. X	CORMTX * CATEGORY	2.197	2	1.099	4.831	.011
Cont. Y Likert X	CORMTX * CATEGORY	4858.342	2	2429.171	40.382	.000
Likert Y Likert X	CORMTX * CATEGORY	3684.064	2	1842.032	32.791	.000

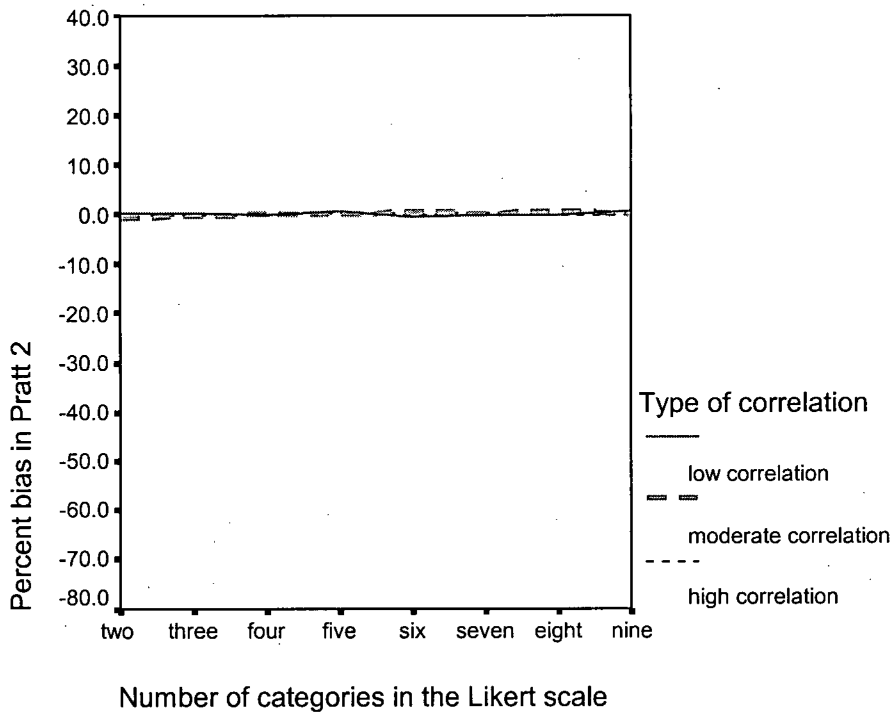

Figure 21

Percent bias in Pratt 2 for variable combination 1 (Likert Y and Continuous X) within correlation matrices and number of Likert scales.

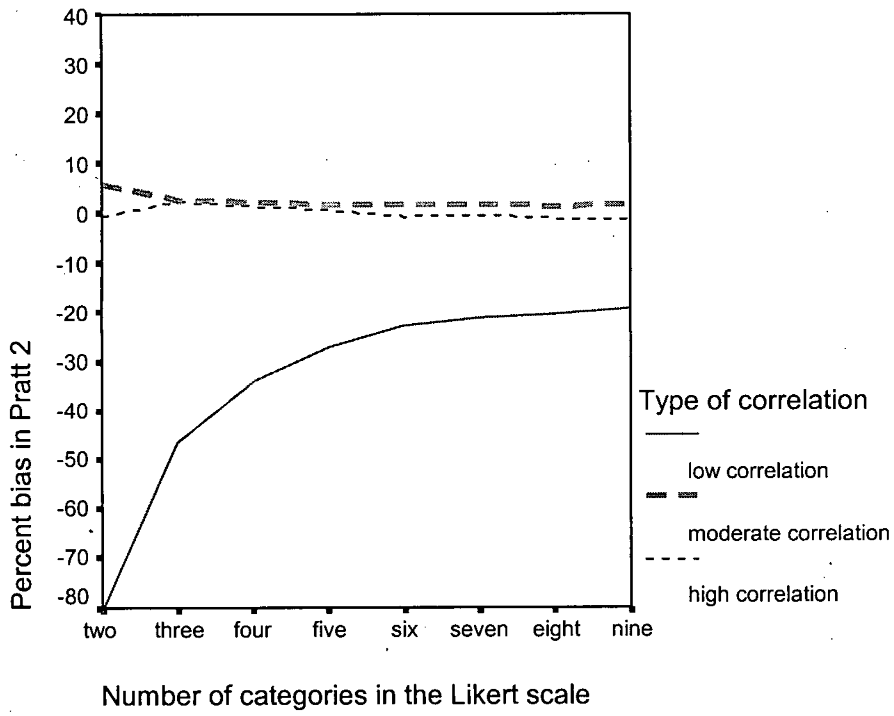

Figure 22

Percent bias in Pratt 2 for variable combination 2 (Continuous Y and Likert X) within correlation matrices and number of Likert scales.

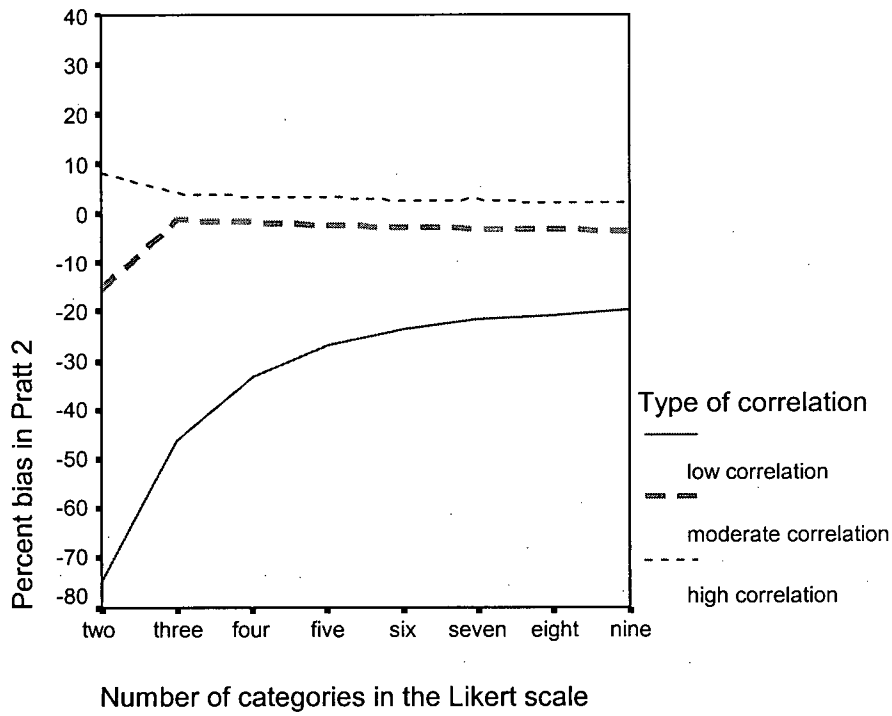

Figure 23

Percent bias in Pratt 2 for variable combination 3 (Likert Y and Likert X) within type of correlation matrices and number of Likert scales.

Pratt Index Three

Table 14 contains the overall model fit for percent bias in Pratt 3. Again, the three-way interaction of VARCOMB by CORMTX by CATEGORY is statistically significant. The simple effects two-way interactions of CORMTX by CATEGORY for the various levels of VARCOMB are listed in Table 15. Furthermore, the plots of the interactions are provided in Figures 24 to 26. For variable combination 1 there was no interaction. However, for variable combinations 2 and 3 the moderate correlation situation was distinct from the low and high correlation cases; and in addition, the effect was largest at 2 scale points.

Table 14

Statistical model fit for percent bias in Pratt 3 for type of correlation matrix, variable combination, response pattern and number of Likert scale points

Source	Sum of Squares	df	Mean Square	F	Sig.
CORMTX	483.290	2	241.645	84.864	.000
VARCOMB	617.007	2	308.504	108.344	.000
PATTERN	2.309	2	1.155	.406	.667
CATEGORY	302.153	1	302.153	106.113	.000
CORMTX * VARCOMB	256.073	4	64.018	22.483	.000
CORMTX * PATTERN	7.144	4	1.786	.627	.644
CORMTX * CATEGORY	165.571	2	82.786	29.074	.000
VARCOMB * PATTERN	17.408	4	4.352	1.528	.196
VARCOMB * CATEGORY	203.407	2	101.703	35.717	.000
PATTERN * CATEGORY	3.942	2	1.971	.692	.502
CORMTX * VARCOMB * PATTERN	7.847	8	.981	.344	.947
CORMTX * VARCOMB * CATEGORY	87.708	4	21.927	7.701	.000
CORMTX * PATTERN * CATEGORY	.603	4	.151	.053	.995
VARCOMB * PATTERN * CATEGORY	5.116	4	1.279	.449	.773
CORMTX * VARCOMB * PATTERN * CATEGORY	1.608	8	.201	.071	1.000
Error	461.288	162	2.847		
Total	4830.270	216			

Table 15

Statistical model fit for levels of variable combinations for correlation matrix by category for Pratt3

Variable combination of dependent and independent	Source	Sum of Squares	df	Mean Square	F	Sig.
Likert Y Cont. X	CORMTX * CATEGORY	.227	2	.114	1.320	.274
Cont. Y Likert X	CORMTX * CATEGORY	75.415	2	37.708	9.942	.000
Likert Y Likert X	CORMTX * CATEGORY	177.637	2	88.819	18.203	.000

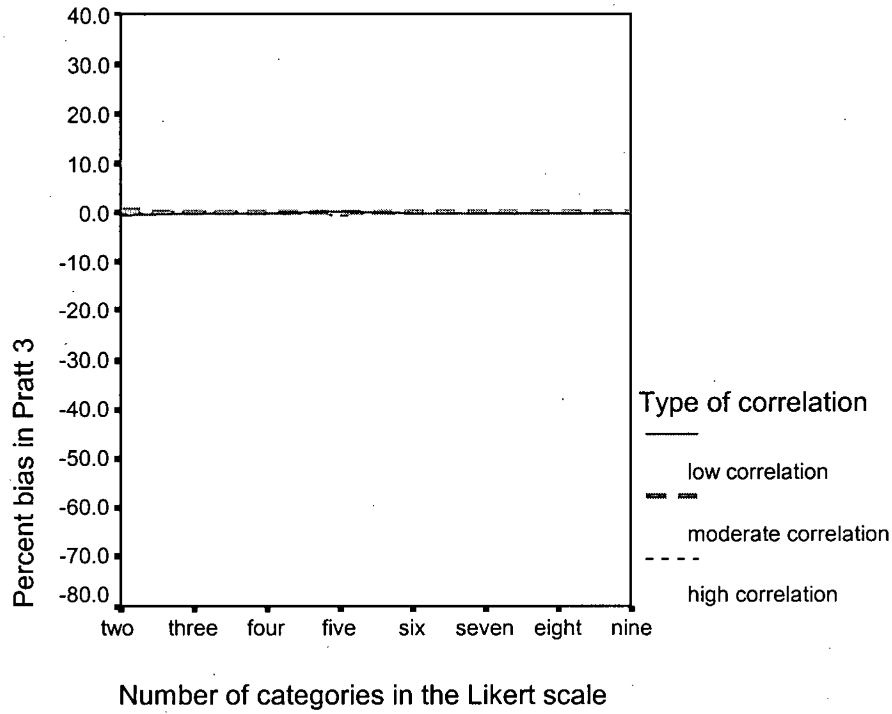

Figure 24

Percent bias in Pratt 3 for variable combination 1 (Likert Y and Continuous X) within type of correlation matrices and number of Likert scales.

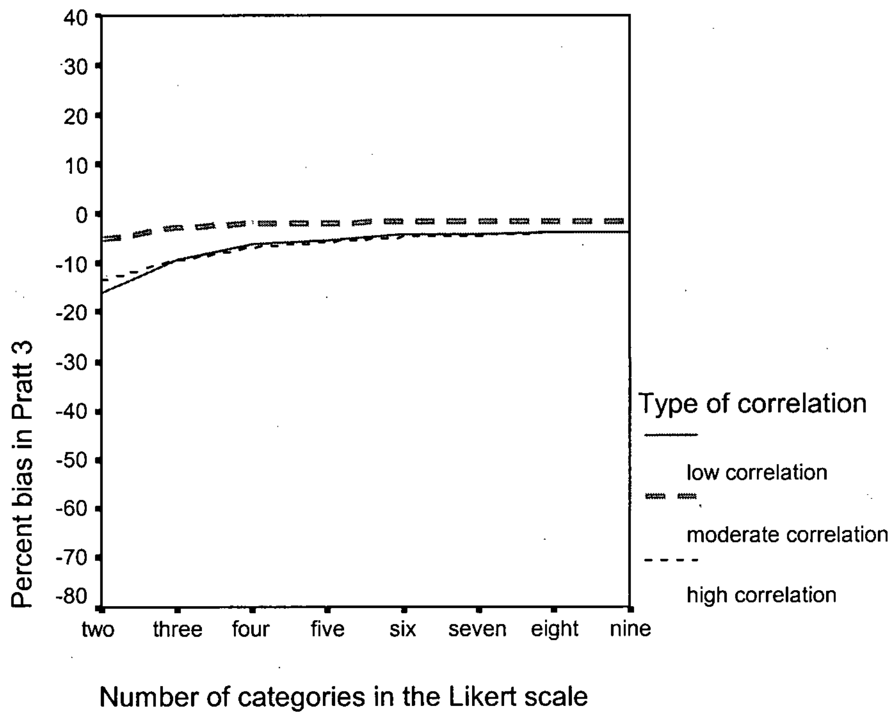

Figure 25

Percent bias in Pratt 3 for variable combination 2 (Continuous Y and Likert X) within type of correlation matrices and number of Likert scales.

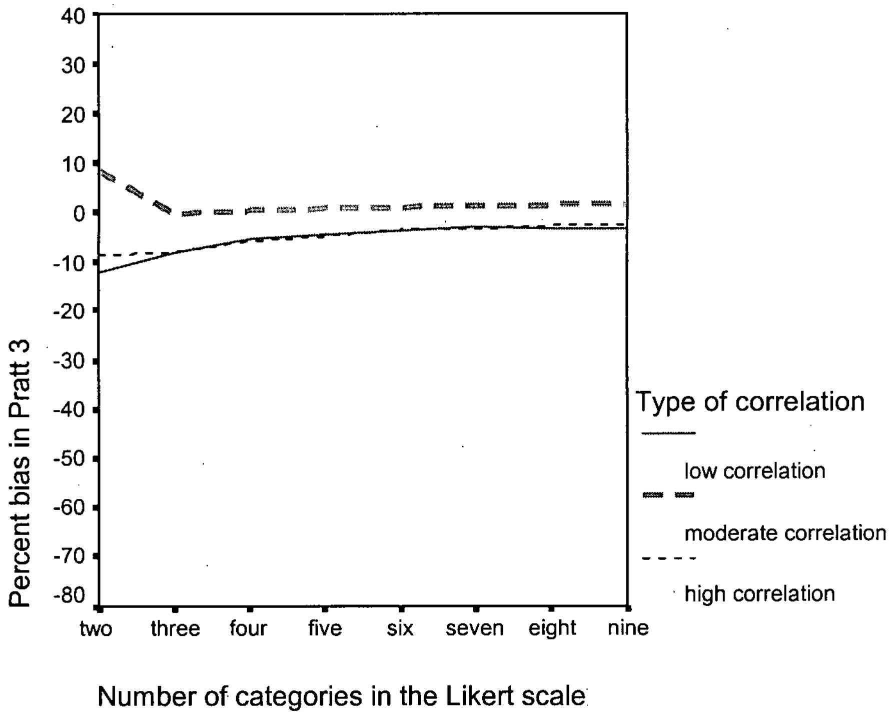

Figure 26

Percent bias in Pratt 3 for variable combination 3 (Likert Y and Likert X) within type of correlation matrices and number of Likert scales.

In summary, for the three Pratt indices taken together, there was no effect of response pattern. However, the remaining three independent variables show complex and differing interaction patterns for each of the Pratt Indices and combination of independent variables -- with variable combination appearing to consistently have the largest effect on the interactions. Percent bias was substantially large in Pratt 2. Given that Pratt 2 value was negative and the associated beta weight was also negative, this could indicate that the predictor was a suppressor variable, hence the large bias in the resulting relative Pratt index. Also, given that (a) the percent bias in the Pratt indices ranged from -80% to 40% and that this variation was explained by various patterns of interactions among the independent variables, and (b) that the primary use of the Pratt index is to order the predictor variables in terms of importance, the consistency in variable ordering across the variables' experimental conditions was investigated.

Likert Data and Variable Importance

To investigate the effect of Likert data on variable ordering, it was first noted that the values of Pratt 3 were greater than Pratt 1 and Pratt 2 in the baseline condition. Two indicator variables were created for continuous and Likert data i.e. for baseline and treatment conditions, showing that Pratt 3 was greater than Pratt 1 and 2, and for the relationship between Pratt 1 and 2. A cross tabulation was done to compare variable ordering in the continuous (baseline) data and in Likert data. The result indicated that variable order remained consistent i.e. invariant across the treatment conditions in spite of the presence of bias in the Pratt indices.

Thus, irrespective of scaling, the order of importance remained invariant and variable X_3 with the largest Pratt index remained the most important and furthermore the relationship between variables X_1 and X_2 was also invariant of Likert scaling. Therefore, the order of importance was not affected by the Likert data across the conditions. Consistency of the relative Pratt index was maintained under different conditions of correlation matrix, variable combination response pattern and number of Likert points. The relative proportion of variance attributed to each variable did not change under the studied conditions.

Effect of Likert Data on Pearson Correlation

To assess the effect of Likert data on the Pearson correlation for each predictor X_i (i.e. X_1, X_2 and X_3) and the criterion Y, the correlations r_{yx1}, r_{yx2}, and r_{yx3} were computed for each variable combination, response pattern, correlation matrix, and Likert scale points as well as for the continuous distribution of X and Y. For the continuous case, the resulting Pearson's correlation was labeled as $r_{yxi\ cont}$. Thus, for the first predictor X_1 and Y, the correlation was $r_{yx1\ cont}$ while for the second X_2 and third predictor X_3 the resulting correlations with Y were $r_{yx2\ cont}$ and $r_{yx3\ cont}$ respectively. Pearson correlation was computed for data in each of the three correlation matrices. The value of $r_{yxi\ cont}$ in all the continuous conditions ranged from 0.202 to 0.600. These correlation values represent common and typical range of values that are frequently encountered in social science research. Similarly, Pearson correlation was computed in the data for the eight Likert scales in each of the three variable combinations and response pattern. The resulting correlations were referred to as $r_{yxi\ model}$ for each condition and Likert data.

To obtain the amount of bias in the estimation of Pearson correlation $r_{yxi\ model}$, the difference between $r_{yxi\ cont}$ and $r_{yxi\ model}$ were computed as follows:

$$Bias\ \Delta r_{yx} = r_{yxi\ model} - r_{yxi\ cont}. \qquad (25)$$

The resulting bias was then expressed as a percentage of the Pearson correlation $r_{yxi\ cont}$ as follows:

$$Percent\ bias = [Bias\ \Delta r_{yx} / r_{yxi\ cont}] \times 100 \qquad (26)$$

The percent bias computed at each combination of the condition of the independent variables was used to analyze the effect of Likert data on the Pearson correlation under the stated conditions. The type of correlation matrix was not expected to impact on the estimation of Pearson correlation between the predictors and the criterion, and so analysis was conducted for each predictor and criterion at each level of variable combinations, type of response pattern and number of Likert scales. The number of Likert scale points, variable combination and type of response pattern were hypothesized to impact the estimation of Pearson correlation of each predictor with the criterion. Percent bias in the Pearson correlations r_{yx1}, r_{yx2} and r_{yx3} for the correlation of X_1 and Y, X_2 and Y, as well as X_3 and Y were denoted by X1PBIAS-r, X2PBIAS-r and X3PBIAS-r respectively. Percent bias in each of the three correlations was analyzed separately.

Statistical Model fit

To provide insight on the effect of Likert scales on the estimation of Pearson correlations, statistical models were fit for each combination of the conditions namely, type of correlation matrix, response pattern, variable combination, and number of Likert scale points using percent bias as the object of analysis.

Statistical model fit for X1PBIAS-r

The results of statistical model fit for the percent bias in X1PBIAS-r are shown in Table 16. There were significant main effects for variable combination (VARCOMB) response pattern (PATTERN) and number of Likert scale points CATEGORY. However, the type of correlation matrix (CORMTX) was not statistically significant. There were statistically significant three-way interaction effects for VARCOMB by PATTERN by CATEGORY. A simple effects analysis was conducted to decompose the three-way interaction effects. Thus, at each level of VARCOMB (variable combination) a two-way model was fit for PATTERN by

CATEGORY. This indicated that the two-way interaction effects depended on the third variable namely VARCOMB.

Table 16

Statistical model fit for Percent bias X1PBIAS-r in Pearson correlation for type of correlation matrix, variable combination, response patterns, and number of Likert scale points

Source	df	Mean Square	F	Sig.
VARCOMB	2	3605.473	42.414	.000
PATTERN	2	31556.566	371.221	.000
CORMTX	2	15.889	.187	.830
CATEGORY	1	4266.517	50.190	.000
VARCOMB * PATTERN * CORMTX * CATEGORY	8	0.07143	.001	1.000
VARCOMB * PATTERN * CORMTX	8	1.315	.015	1.000
VARCOMB * PATTERN * CATEGORY	4	511.149	6.013	.000
VARCOMB * CORMTX * CATEGORY	4	.946	.011	1.000
PATTERN * CORMTX * CATEGORY	4	.241	.003	1.000
VARCOMB * PATTERN	4	7126.916	83.839	.000
VARCOMB * CORMTX	4	15.074	.177	.950
VARCOMB * CATEGORY	2	1435.174	16.883	.000
PATTERN * CORMTX	4	1.211	.014	1.000
PATTERN * CATEGORY	2	1993.454	23.450	.000
CORMTX * CATEGORY	2	1.423	.017	.983
Error	162	85.007		
Total	216			

The significant main effects were not analyzed. Table 17 lists the two-way interactions at the three levels of variable combination. At variable combination 1 (Likert Y and Continuous X) the two-way interaction effect of PATTERN by CATEGORY was significant. This implied that percent bias in Pearson correlation X1PBIAS-r varied across similar categories for different response patterns. Similar results are noted for variable combination 2 (Continuous Y and Likert X). At variable combination 3 (Likert Y and Likert X) the interaction effect of PATTERN by CATEGORY was not significant. Figures 27 to 29 show the interaction plots of the two-way interaction effects listed in Table 17.

Table 17

Statistical model fit for Percent bias X1PBIAS-r in Pearson correlation for levels of variable combination in response pattern distribution and number of Likert scale points

Variable combination of dependent and independent VARCOMB	Source	Sum of Squares	df	Mean Square	F	Sig.
Likert Y Cont. X	PATTERN * CATEGORY	3047.914	2	1523.957	22.782	.000
Cont. Y Likert X	PATTERN * CATEGORY	2982.329	2	1491.164	21.952	.000
Likert Y Likert X	PATTERN * CATEGORY	1.261	2	.631	.008	.992

From the Figures it is evident that for:

(a) Variable combination 1 (Likert Y and Continuous X) percent bias was largest for the positively skewed response patterns. The least bias was in the case of equal interval and symmetric response distribution in which the percent bias tended to zero as the number of scale points increased (see Figure 27). The negatively skewed and equal conditions showed similar bias patterns except for the two Likert scale point case.

(b) Variable combination 2 (Continuous Y and Likert X) yielded the same percent bias as in variable combination 1 (Likert Y and Continuous X) as expected given that the correlation is symmetric with regard to X and Y (see Figure 28).

(c) Variable combination 3 (Likert Y and Likert X) right and left bunching of response patterns performed the same. Percent bias was largest at the two-point Likert scale for all the response patterns but systematically reduced beyond the three-point Likert scale. For the equal interval response distribution percent bias was relatively larger in variable combination 3 compared to variable combination 1 and 2 across the number of Likert scales. This implied that percent bias varied across variable combinations for a given response pattern (see Figure 29).

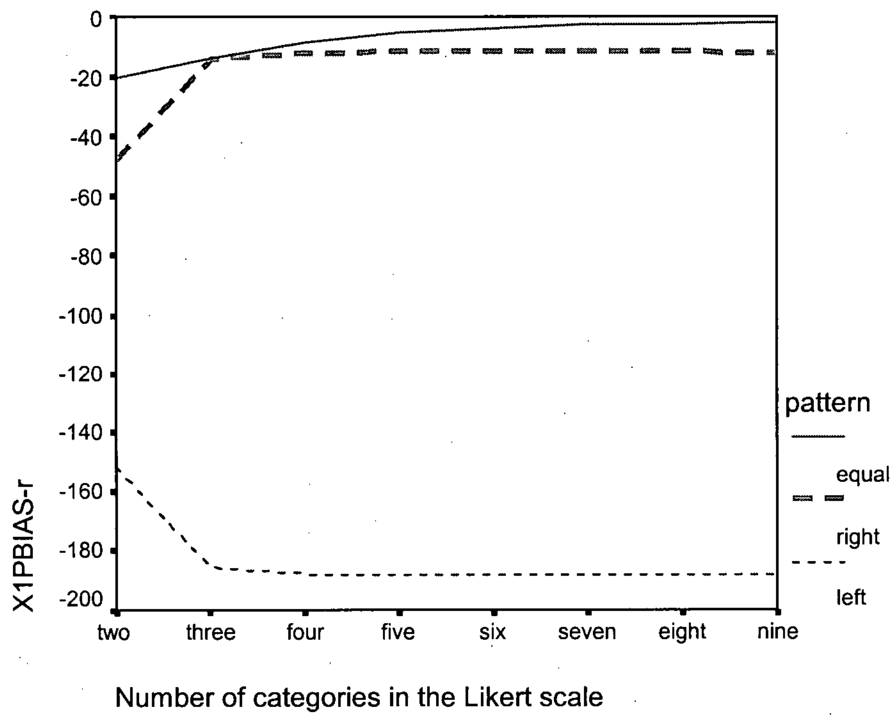

Figure 27

Percent bias X1PBIAS-r in Pearson correlation for variable combination 1 (Likert Y and Continuous X) within response patterns and number of Likert scales.

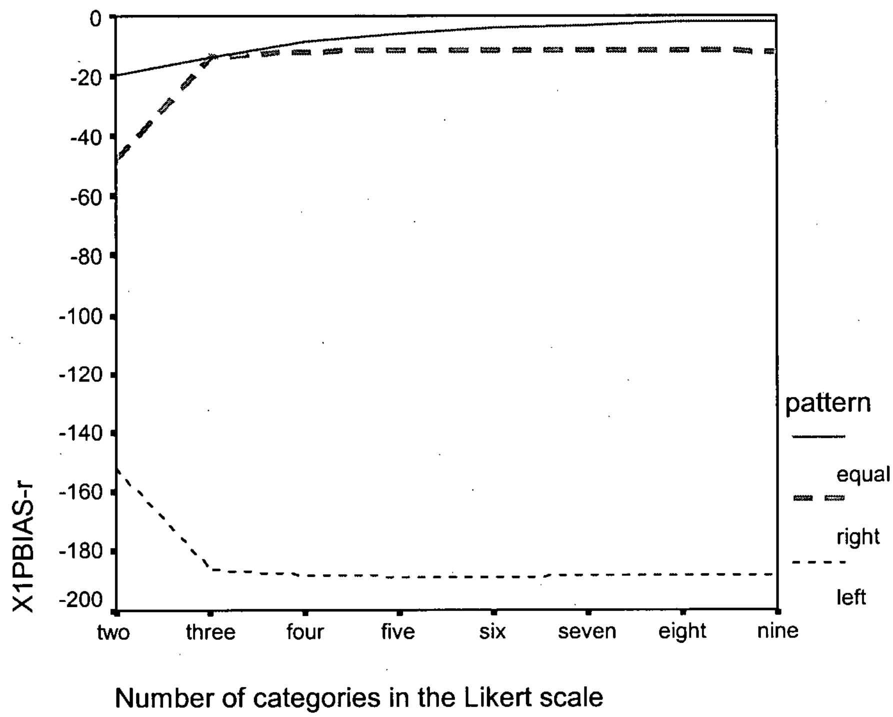

Figure 28

Percent bias X1PBIAS-r in Pearson correlation in variable combination 2 (Continuous Y and Likert X) within response patterns and number of Likert scales.

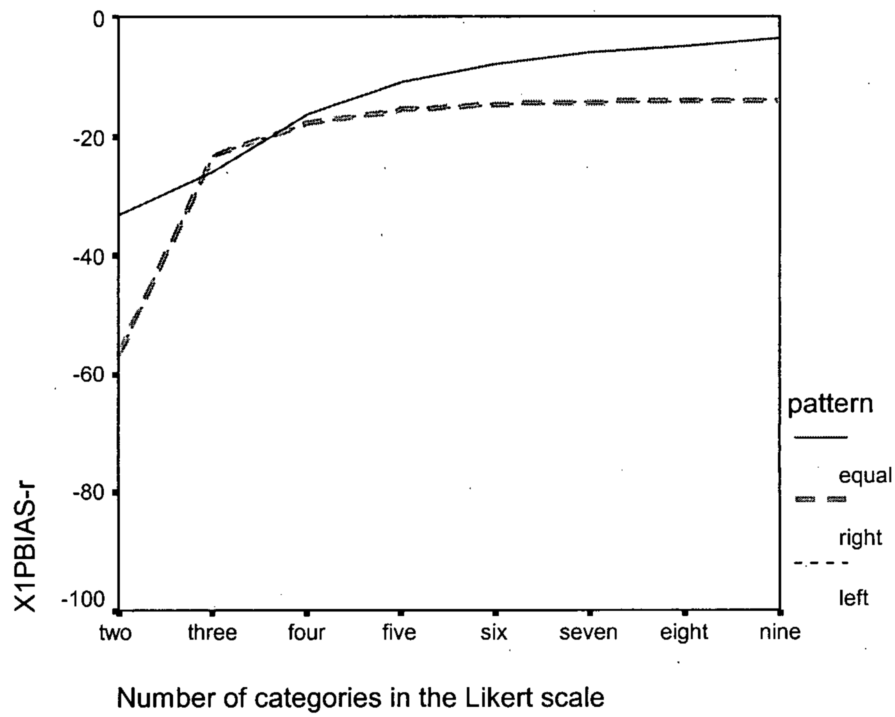

Figure 29

Percent bias XIPBIAS-r in Pearson correlation in variable combination 3 within response patterns and number of Likert scales.

Statistical model fit for X2PBIAS-r

A statistical model was fit for percent bias X2PBIAS-r in the Pearson correlation for the second predictor X_2 and Y, for all possible combination of main and interaction effects of the independent variables (see Table 18). Results of the model fit in X2PBIAS-r were very similar to those of X1PBIAS-r. The main effect for CORMTX was not statistically significant, although those of PATTERN, VARCOMB and CATEGORY were statistically significant. There was a statistically significant three-way interaction effect of VARCOMB by PATTERN by CATEGORY.

Table 18

Statistical model fit for percent bias X2PBIAS-r in Pearson correlation for type of correlation matrix, variable combination, Likert scale points, and response patterns

Source	df	Mean Square	F	Sig.
VARCOMB	2	3527.842	39.866	.000
PATTERN	2	31870.728	360.155	.000
CORMTX	2	31.115	.352	.704
CATEGORY	1	4221.468	47.705	.000
VARCOMB * PATTERN * CORMTX * CATEGORY	8	.947	.011	1.000
VARCOMB * PATTERN * CORMTX	8	3.270	.037	1.000
VARCOMB * PATTERN * CATEGORY	4	503.196	5.686	.000
VARCOMB * CORMTX * CATEGORY	4	2.119	.024	.999
PATTERN * CORMTX * CATEGORY	4	.755	.009	1.000
VARCOMB * PATTERN	4	7116.063	80.415	.000
VARCOMB * CORMTX	4	32.101	.363	.835
VARCOMB * CATEGORY	2	1418.853	16.034	.000
PATTERN * CORMTX	4	3.835	.043	.996
PATTERN * CATEGORY	2	1955.075	22.093	.000
CORMTX * CATEGORY	2	2.438	.028	.973
Error	162	88.492		
Total	216			

Given the significant three-way interaction like in the case of X1PBIAS-r, a simple effects analysis was conducted to decompose the interactions. At each level of VARCOMB two-way model fits were conducted for PATTERN by CATEGORY as well as that of VARCOMB by PATTERN. A significant two-way interaction of PATTERN by CATEGORY implied that percent bias of Pearson correlation varied across response patterns for a given set of category of Likert scale points. Table 19 lists the two-way interactions at each level of variable combination.

At variable combination 1 (Likert Y and Continuous X) there is a significant PATTERN by CATEGORY interaction. At variable combination 2 (Continuous Y, Likert X) there were significant main effects for PATTERN and significant two-way interaction effects between PATTERN and CATEGORY, while at variable combination 3 (Likert Y and Likert X) the two-way interaction effect of PATTERN by CATEGORY was not significant. This result in variable combination 3 differs from those in variable combination 1 and 2 indicating the influence of number of Likert scales when both criterion and predictor are Likert. Figures 30 to 32 show the interaction plots for the two-way interaction listed in Table 19.

Table 19

Statistical model fit for Percent bias X2PBIAS-r in Pearson correlation for levels of variable combination in response pattern distribution and number of Likert scale points

Variable combination of dependent and independent	Source	Sum of Squares	df	Mean Square	F	Sig.
Likert Y Cont. X	PATTERN * CATEGORY	2896.048	2	1448.024	21.877	.000
Cont. Y Likert X	PATTERN * CATEGORY	3025.342	2	1512.671	22.062	.000
Likert Y Likert X	PATTERN * CATEGORY	1.546	2	.773	.008	.992

From the Figures it is evident that for:

(a) Variable combination 1 (Likert Y and Continuous X) to variable combination 3 (Likert Y and Likert X) the left bunching cases showed substantially more negative bias than the right or equal conditions. However, for the left bunching case, bias reduced at the two scale points whereas the bias increased for right bunching. The least bias was in the case of equal interval and symmetric response distribution in which the percent bias tended to zero as the number of scale points increased (see Figure 30).

(b) Again, as expected variable combination 2 (Continuous Y and Likert X) yielded the same percent bias as in variable combination 1 (Likert Y and Continuous X) as shown in Figure 31.

(c) Variable combination 3 (Likert Y and Likert X) percent bias was largest at the two-point Likert scale for all the response patterns but systematically reduced beyond the three-point Likert scale. For the equal interval and symmetric response distribution percent bias was relatively larger in variable combination 3 compared to variable combination 1 and 2 across the number of Likert scales. This implied that percent bias varied across variable combinations for a given response pattern. Percent bias in left bunching response pattern (positively skewed) was the same as in right bunching response pattern (negatively skewed). This is evident in Figure 32.

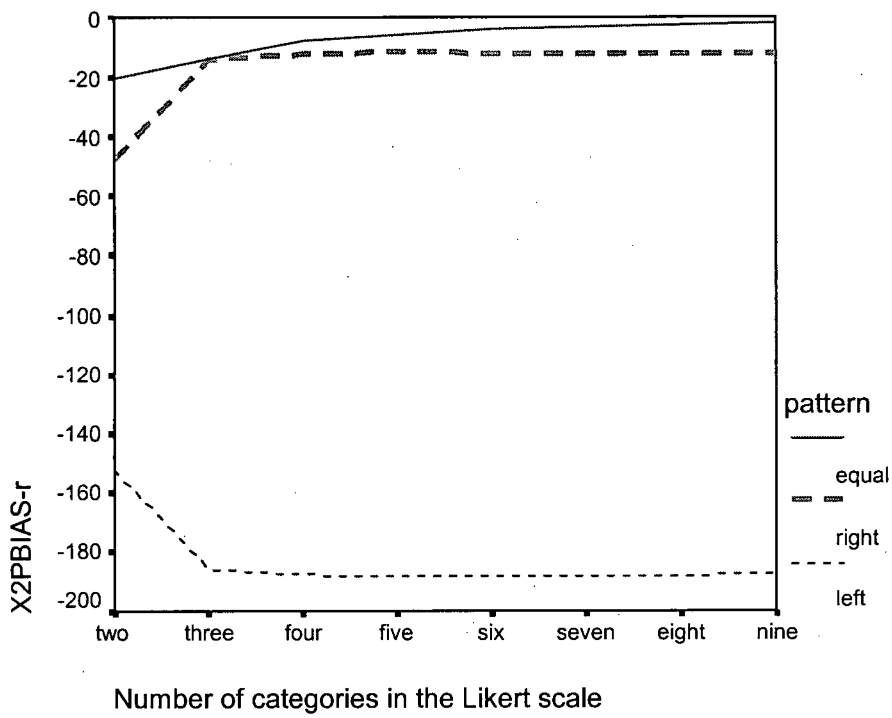

Figure 30

Percent bias X2PBAS-r in Pearson correlation for variable combination 1 (Likert Y and Continuous X) within response patterns and number of Likert scales.

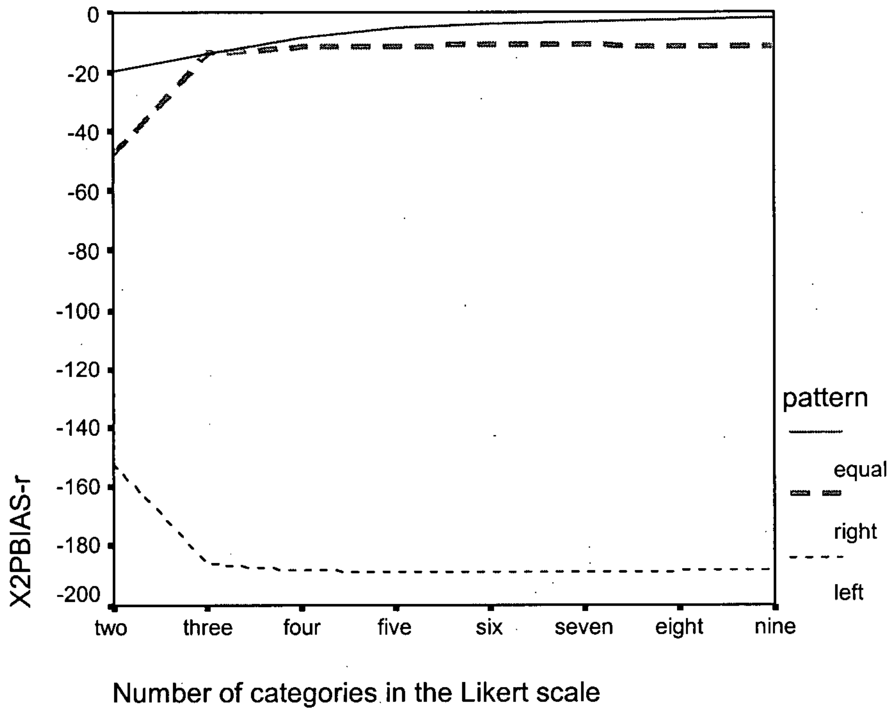

Figure 31

Percent bias X2PBIAS-r in Pearson correlation in variable combination 2 (Continuous Y and Likert X) within response pattern and number of Likert scales.

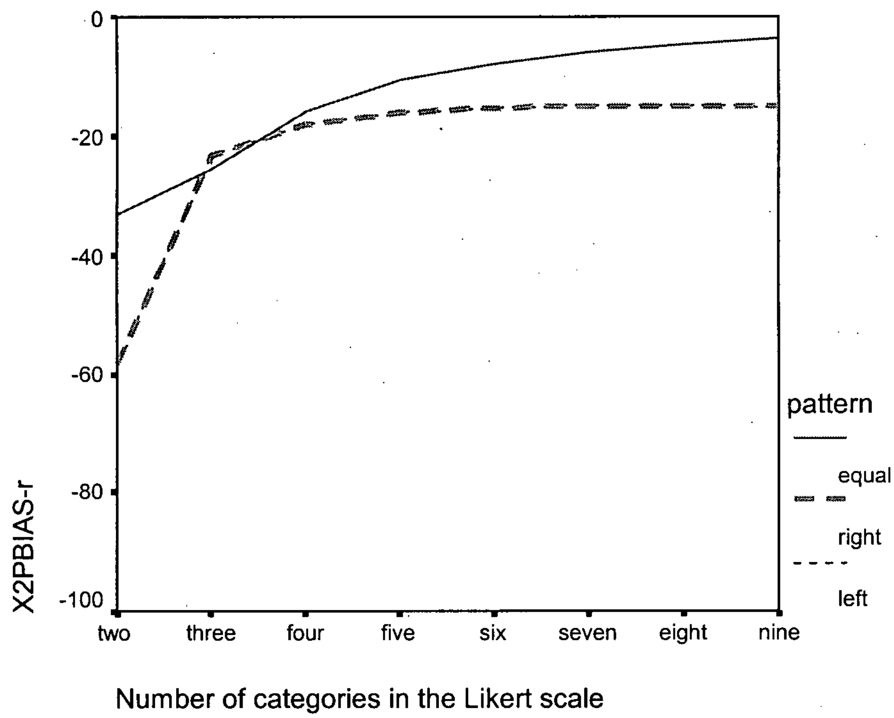

Figure 32

Percent bias X2PBIAS-r in Pearson correlation for variable combination 3 (Likert Y and Likert X) within response pattern and number of Likert scales.

Statistical model fit for X3PBIAS-r

Statistical model fit was conducted for percent bias X3PBIAS-r in Pearson correlation for the predictor X_3 and Y, with the independent variables CORMTX, VARCOMB, PATTERN and CATEGORY. All possible combinations of the interaction between the independent variables were included in the model as shown in Table 20.

Table 20

Statistical model fit for Percent bias X3PBIAS-r in Pearson correlation for type of correlation matrix, variable combination, response pattern and number of Likert scale points

Source	df	Mean Square	F	Sig.
Corrected Model	53	20448.516	254.530	.000
Intercept	1	123663.558	1539.283	.000
CORMTX	2	25.777	.321	.726
PATTERN	2	31411.382	390.988	.000
VARCOMB	2	4129.448	51.401	.000
CATEGORY	1	4010.012	49.914	.000
CORMTX * PATTERN * VARCOMB * CATEGORY	8	7.094E-02	.001	1.000
CORMTX * PATTERN * VARCOMB	8	1.511	.019	1.000
CORMTX * PATTERN * CATEGORY	4	5.727E-02	.001	1.000
CORMTX * VARCOMB * CATEGORY	4	3.159	.039	.997
PATTERN * VARCOMB * CATEGORY	4	513.228	6.388	.000
CORMTX * PATTERN	4	1.411	.018	.999
CORMTX * VARCOMB	4	25.203	.314	.869
CORMTX * CATEGORY	2	2.919	.036	.964
PATTERN * VARCOMB	4	7204.292	89.674	.000
PATTERN * CATEGORY	2	1976.278	24.599	.000
VARCOMB * CATEGORY	2	1313.654	16.351	.000
Error	162	80.338		
Total	216			

The resulting statistical model fit based on the univariate analysis of variance was similar to those of X1PBIAS and X2PBIAS-r. There were three significant main effects, of VARCOMB, PATTERN and CATEGORY. However, the type of correlation matrix CORMTX was not significant. Like in the Pearson correlation of the other two predictors and criterion, there was a significant three-way interaction effect among the three variables, VARCOMB by

PATTERN by CATEGORY. This was followed by a further analysis of simple effects to decompose the interaction by fitting two-way statistical models at each level of variable combination VARCOMB. Shown in Table 21 are the two-way model fits of interactions at each level of variable combination and Figures 33 to 35 display the interaction plots.

Table 21

Statistical model fit for Percent bias X3PBIAS-r in Pearson correlation for levels of variable combination in response pattern distribution and number of Likert scale points

Variable combination of dependent and independent	Source	Sum of Squares	df	Mean Square	F	Sig.
Likert Y Cont. X	PATTERN * CATEGORY	2999.652	2	1499.826	22.119	.000
Cont. Y Likert X	PATTERN * CATEGORY	3002.932	2	1501.466	22.157	.000
Likert Y Likert X	PATTERN * CATEGORY	2.887	2	1.443	.020	.980

It is noted that the results of percent bias in Pearson correlation for the first, second, and third predictors with the criterion are similar and are therefore replicated across the correlations irrespective of the magnitude of the correlations. Thus, percent bias was influenced only by the independent variables in the study except type of correlation matrix, and not the magnitude of the Pearson correlation. This implies that the findings can be generalized to all the range of Pearson correlation values studied.

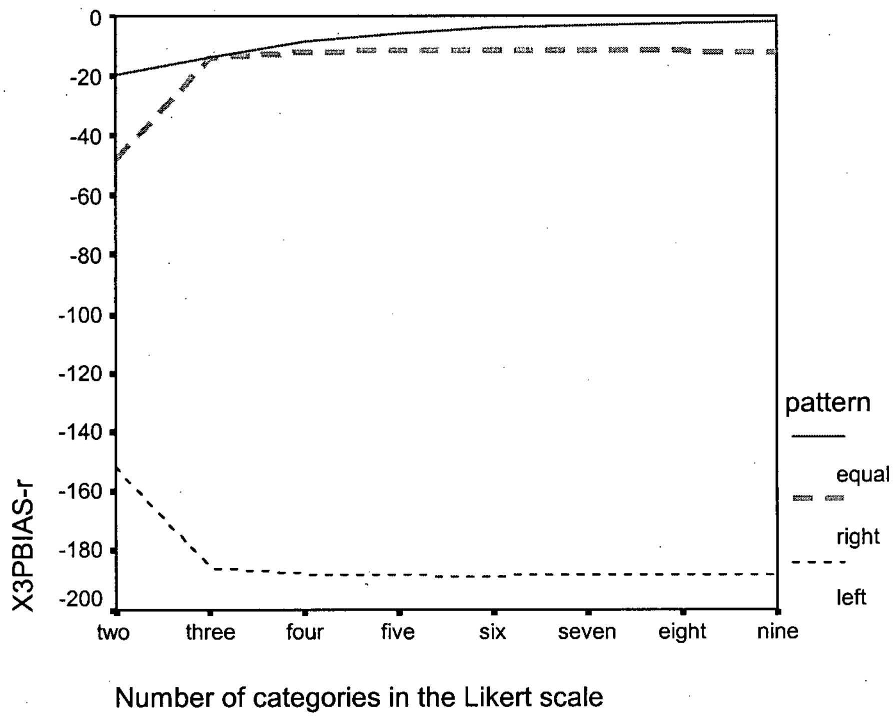

Figure 33

Percent bias in Pearson correlation X3PBIAS-r in variable combination 1 (Likert Y and Continuous X) within response pattern and number of Likert scales.

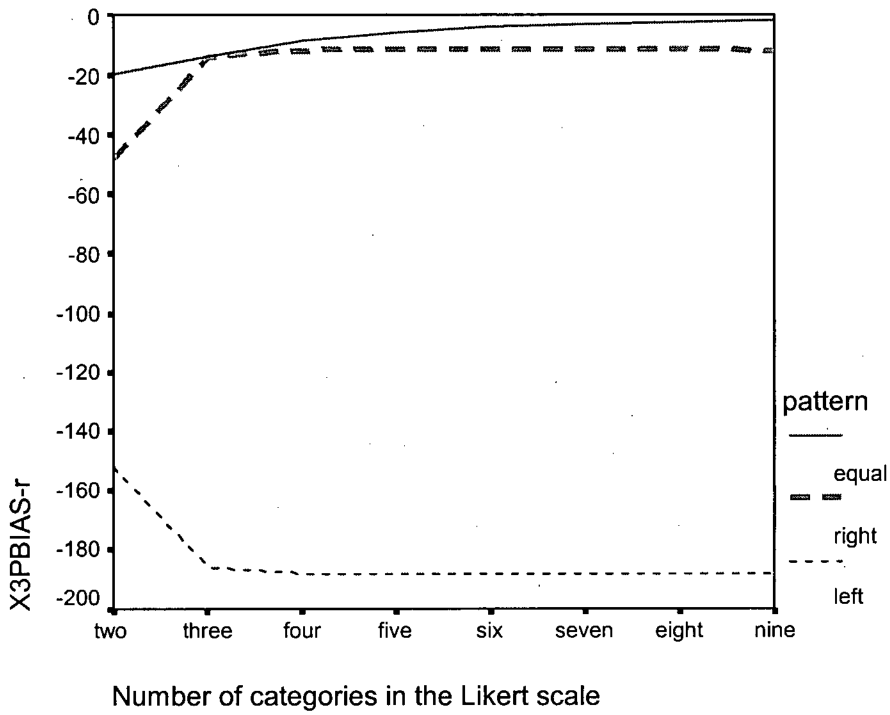

Figure 34

Percent bias in Pearson correlation X3PBIAS-r for variable combination 2 (Continuous Y and Likert X) within response patterns and number of Likert scales.

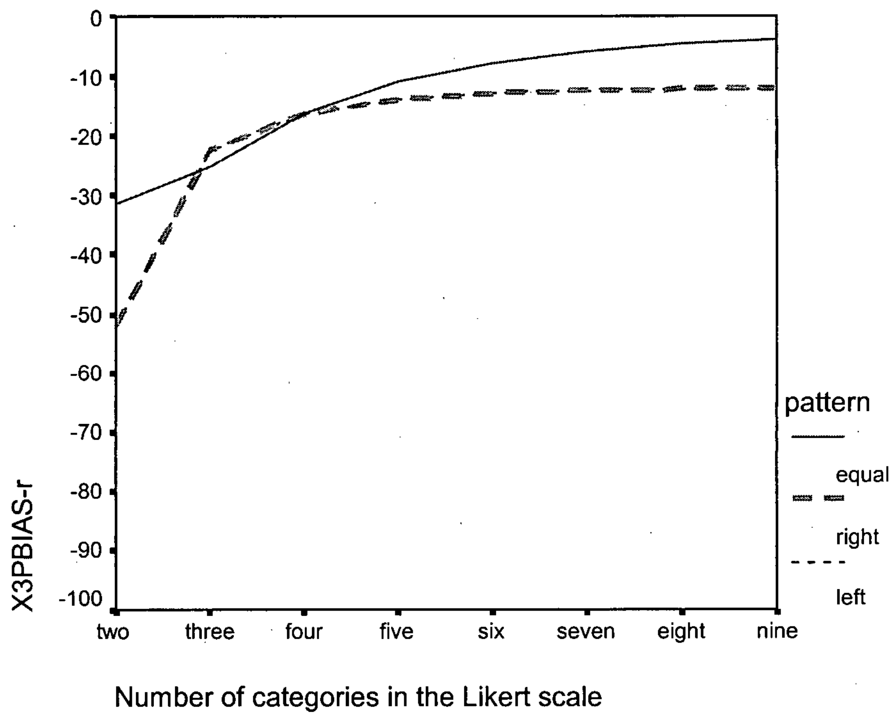

Figure 35

Percent bias in Pearson correlation X3PBIAS-r for variable combination 3 (Likert Y and Likert X) within response patterns and number of Likert scales.

CHAPTER V

DISCUSSION

Introduction

The study examined the impact of using Likert data in linear regression models through bias in the model fit. When ordinary least square linear regression models were fit for the Likert data and compared to the same models with continuous data, bias occurred in the estimation of R^2, relative Pratt index, and Pearson correlation. Amount of bias increased with the decrease in the number of Likert scale points under the stated conditions of the independent variables. Bias in the estimation of the dependent variable was addressed in terms of percent bias at each combination of conditions namely, type of correlation matrix, distribution of response pattern, and variable combination of Likert conditions in both predictors and criterion. The four-point Likert scales were found to be optimal in that there was no substantial gain of information in using more than four Likert scale points.

Type of correlation matrix did not affect the estimation of Pearson correlation. However, for Pratt indices, percent bias was substantially large at Pratt 2 for low correlation matrix but negligible at Pratt 1 and Pratt 3 across all the correlation matrices. This could be explained by the negative value of Pratt 2 in the low correlation matrix. The associated beta weights were also negative, implying that the predictor was a suppressor variable. Thus, the effect of Likert data on Pratt indices with suppressors will need to be investigated. Relative Pratt index was least affected by response distribution and remained robust across the range of response distributions. The response distribution, and violation of normality assumption affected the estimation of R^2 resulting in bias in R^2 and Pearson correlation coefficient.

Equality of intervals of scale had an effect on the estimation of the three dependent variables in that bias was larger for unequal interval Likert data than for the equal interval case. While bias was observed in the relative Pratt index across the independent variables, it was not significant, except for the number of Likert scale points. Thus, the relative Pratt index is robust to response distribution, type of correlation matrix, and type of Likert combination condition. This has implications to research practice in that the robust relative Pratt index is a prudent choice for determining variable importance among other indices given the limitations of beta weights, communality measures, and partial correlations (Courville & Thompson, 2001; Darlington, 1990; Pedhazur, 1982).

The study is unique in terms of methodology in that it is the first study in which statistical model fits have been used to investigate the effects of Likert data on linear regression models and Pearson correlation coefficient. Previous studies also used simulation methods and presented tabular and graphical comparison of results but did not provide statistical evidence and insight of their findings other than the displays of R^2 and Pearson correlation coefficients.

In previous studies on Pearson correlation and Likert data, normality assumption was made for the response data. This study examined violation of the assumption using skewed distribution in the Likert data and its implications. Previous studies also assumed the equality of intervals in Likert scales. The present study investigated both equal and unequal intervals of Likert data and its impact on the model fit in linear regression models. Similar to previous studies, several Likert scale points were generated together with the combination of the independent variables for each model fit. Bias in the statistics and model fit resulted in information loss, which has validity implications in terms of interpretation of data, construction of tests and survey instruments where Likert scales are frequently use.

The discussion presented addresses four aspects of the study, (1) the bias in R^2, Pearson correlation coefficient, and relative Pratt index in linear regression models when using Likert data, (2) the relationship of the present findings to previous studies on the statistics, (3) theoretical, research and practical implications of the present findings, (4) future research and directions. Finally, recommendations are suggested for research practitioners and consumers.

Bias in the estimation of R-squared

Results of R^2 estimation and the percent bias were presented at each independent variable condition and combination. For types of correlation matrix, percent bias in R^2 increased with the reduction in Likert points. Percent bias was largest at the low inter-variable correlation matrix and lowest at the high inter-variable correlation matrix. Across the three types of correlation matrices, percent bias was largest at the 2-point Likert scale and least at the 9-point Likert scale. This confirmed the hypothesis on the effect of type of correlation matrix on the estimation of R^2 in Likert data. Furthermore, increase in percent bias associated with the decrease in Likert point, confirmed the hypothesis that the amount of bias increases with the overall decrease in Likert scale points.

Bias in the estimation of R^2 in Likert data was affected by the distribution of response patterns for different skewed conditions, non-normal conditions and for equal as well as unequal intervals. Three conditions of non-normal (skewed) for equal and unequal interval conditions of response patterns were evaluated. There was a systematic reduction in the large percent bias from the 2-point Likert scale to the 9-point Likert scale as the number of Likert points increased. This trend was observed across the three distributions of response patterns.

At the two skewed distributions of left and right bunching (positive and negatively skewed distribution) with unequal interval, percent bias was largest at the 2-point Likert scale but reduced rapidly at the 3-point Likert scale, leveling off at the 4-point Likert scale. The overall biases in the two skewed distributions were larger than for the symmetric and equal interval distribution pattern. Thus, the hypothesis on the effect of type of response distribution on the estimation of R^2 for Likert scale was confirmed. It can be concluded that bias in R^2 increased with skewness in the response distributions.

For equal and unequal interval percent bias was largest for the unequal interval skewed distribution of responses at the 2-point Likert scale in both left and right bunching (positive and negatively skewed distribution). Percent bias varied in the skewed distribution across the correlation matrices and variable combinations. This was more pronounced, and larger in variable combination 3, when both X and Y are Likert, and for low inter variable correlation matrix than in the rest of the Likert conditions. The effect of coarse categorization on both predictors and criterion results in larger biases than when only the predictors are categorized or when only the criterion is categorized. Skewed distributions further accentuated the amount of bias in the estimation of R^2.

Percent bias of the estimation of R^2 across variable combination of Likert conditions increased with the decrease in number of Likert scales. The amount of bias was the same in the two skewed distributions but larger than for the symmetric distribution of response patterns. Percent bias was largest at the 2-point Likert scale in all the three variable combinations of Likert conditions i.e. when Y is Likert and Continuous X, when Y is continuous and X is Likert, and when both Y and X are Likert. Percent bias rapidly reduced from the 2-point Likert scale to 3-point Likert scale and leveled off after the 4-point Likert scale in variable combination 1 and 2, when Likert Y and Continuous X, and when Continuous Y and Likert X. Percent bias persisted after the 4-point Likert scale although this varied across correlation matrices and patterns of response. This confirmed the hypothesis on the effect of Likert data on both predictors and criterion on the estimation of R^2.

Statistical modeling of the percent bias within the four independent variables of number of Likert points, types of inter variable correlation matrix, distribution of response patterns, and variable combination of Likert condition showed only three main effects. Three of the independent variables affected the estimation of R^2 in Likert data in terms of bias. These were, number of Likert scale points, pattern of responses and variable combination of Likert conditions. Type of correlation matrix was not significant. Thus, the type of correlation matrix did not affect percent bias in Likert data. This was a surprise finding as it was hypothesized that type of correlation matrix would influence estimates of R^2 in Likert data. There were no statistically significant interaction effects among the independent variables.

Post hoc tests on the significant main effects of number of Likert points, response patterns and variable combinations of Likert condition revealed that as the number of Likert points reduced, percent bias increase, and that percent bias was higher for skewed response distribution with unequal interval than in symmetric distribution with equal intervals. Percent bias was also highest for the variable combination in which both predictors and criterion were Likert, confirming the research hypothesis on the effect of categorization. These findings confirm what Cohen (1983) proposed with regard to bias and the high cost of dichotomization of both criterion and predictor in a simple bivariate case. This can be extended to the case of multiple regression models and coarse categorization of both predictors and criterion.

Relationship of Finding to Previous Research

Findings on the model fit indicate that Likert data impacted on the fit of linear regression models by yielding biases in the estimation of R^2, which resulted in consistent underestimation of R^2. Bias increased with the decrease in Likert scale points. Hardly any research had been done on the impact of Likert data on linear regression model fit, except in the studies by Russell, Pinto, and Bobko (1991), and Russell and Bobko (1992) in which the effect of Likert data on the effect sizes of moderated linear regression models in terms of changes in R^2 were examined. Even then, amount of bias and information loss in the model fit was not addressed. The impact of Likert data on the estimation of linear regression models was not addressed in the two studies.

Other studies addressed the impact of Likert data on multivariate analysis such as factor analysis and in structural equation modeling, but not linear regression models. However, it was evident that Likert data resulted in poor model fit as supported by studies on the impact of Likert data on factor analysis and structural equation modeling (Bernstein & Teng, 1989;

Bollen, 1989; Curran, West & Finch, 1996; Green et. al, 1997; Muthen & Kaplan, 1985, 1992). These studies did not address the impact of Likert data in linear regression model fit. Therefore, findings of the current study provide interesting and unique insight to the problem of using Likert data in linear regression models in terms of bias in the model fit and relative loss of information as well as precision. This has validity implications in the interpretation of data from theoretical, practical and research methodology point of view.

Implications of Findings

Theoretical implications of the findings on bias in the estimation of R^2 and the information loss resulting in the fit of linear regression model is that ordinal and logistic regression models should be used where the criterion and predictors are Likert or categorized. As shown in the study by Cohen (1983), the cost of dichotomization in terms of information loss and precision is substantial, which results in distortions of interpretation of the data. This can be extended and generalized to Likert scales in which the number of Likert points was noted to impact on the fit of linear regression models in terms of bias in the estimation of R^2.

Krieg (1999) provided an analytic approach to resolve the issue of bias in Pearson correlation and compared the resulting Pearson correlations with Sheppard's correction formula. Results of the proposed analytic approach to addressing bias in Pearson correlation is not applicable to multiple regression model due to the multivariate nature of such models and the underlying assumptions, and so this simulation study provided a clear insight of the extent of the bias in linear regression models, and the need to be cognizant of the information loss resulting from using Likert data in such regression models. Interpretation of linear regression models with Likert data should be done with caution given the amount of bias associated with the use of such data.

Research implication of the findings on the bias in R^2 is that ordinary least square regression method should not be used with Likert data as this has a negative impact on the accuracy of the interpretation of the data. A common practice of collapsing data into ordinal categories to simplify analysis should also be avoided, as this results in information loss. Where possible, continuous data it its original form should be used. However, there are occasions when the researcher is compelled by circumstances in the study, or the construct of interest, to use Likert scales with few scale points. For researchers in cognitive and developmental psychology as well as related disciplines, the optimal choice of number of Likert scale points may depend on the context and situation. For example, children and certain subpopulation of adults may find Likert scales with more than two or three Likert scale points cognitively demanding or confusing. In such situations researchers may use dichotomized scales or three-point Likert scales. On the other hand, when fine-grained distinction among respondents is demanded by specific research questions, then the researcher may decide to adopt four or more Likert scale points.

As a general recommendation to avoid the bias resulting from using ordinary least squares regression to analyze Likert data, ordinal logistic regression models should be used (see Ananth & Kleinbaum, 1997). One of the reasons for the occurrence of bias in ordinary least square regression is that it is assumed in the model, that the criterion is continuous. Thus, an ordinal categorical or Likert criterion is bound to result in bias in the estimation of R^2 whenever ordinary least square linear regression is used. Although Krieg (1999) recommends the use of graphical scales in place of Likert scales as a means of overcoming bias, researchers may find the former method more cumbersome than the latter. Note that a graphical scale is one in which the respondent is provided a continuous line anchored at the two ends. The respondent is instructed to indicate their response by marking a point on the line. The researcher then needs to convert this marking to some sort of numerical code on a common metric – e.g., measuring the distance from the far left of the line. Whatever the researcher chooses to do, two assumptions must hold: (a) there is a common metric across individuals, and (b) the metric (i.e., the units) are meaningful to the respondent. Therefore, for example, if one measures the response in centimeters then a one-centimeter difference is meaningful to the respondent – i.e., the resolution of measurement needs to be meaningful to the respondent. Because Likert scales are often used by researchers, ordinal logistic regression models are recommended in place of ordinary least square linear regression model for analyzing Likert data.

Beyond the four-point Likert scales, there was negligible improvement in the information gain in regression as bias remained almost constant and low. For normally distributed Likert response data, bias asymptotically reduced with the increase in Likert points. Researchers need to consider amount of bias associated with using Likert scales with few scale points and skewed distribution as demonstrated in this study. When reading research in which Likert scales with fewer than four scale points have been used, the results will have considerably more bias in the estimation of R^2 than when four or more scale points have been used.

Practical implication of the findings on the impact of Likert data on the estimation of R^2 is that the researcher needs to be cognizant of the presence of bias that is inherent in Likert scales with respect to ordinary least square linear regression model fit. However, beyond the four-point Likert scale point, reduction in percent bias was negligible and so using more than four scale points may have little gains in terms of information and precision.

Unequal interval and varying threshold difference for skewed distribution resulted in larger percent bias in estimating R^2 than the case of equal intervals, although in practice the assumption of equal interval and normally distributed response pattern is hardly ever met. The number of scale points and the distribution of response patterns need to be considered.

In order to assess the impact of Likert data on the interpretation of regression models in terms of variable importance and ordering, beta weights, zero order correlation of the each predictor and the criterion, and the resulting R^2 were computed for each combination of the independent variables. Due to lack of additive and ratio properties of beta weights in relation to the criterion, variable importance was assessed using relative Pratt index, which is a functional transformation of beta weights, zero order correlation of the respective predictors and overall R^2.

Bias in Relative Pratt Index

Bias in relative Pratt index for each predictor was computed and transformed into percent bias. Plots of the percent bias for each combination of conditions of the independent variables against number of Likert scale points showed that percent bias was largest for 2-point Likert scale, but reduced drastically after the 3-point Likert scale. However, at and beyond the 4-point Likert scale, change in percent bias was negligible for the symmetric distribution with equal interval and not significant in the skewed distribution of response patterns. Percent bias in the estimation of relative Pratt index was relatively larger in the case of variable combination where both criterion and predictors were Likert than when only one of the variables were Likert. The order of variable importance was invariant across variable combinations of Likert conditions of the independent variables stated. Thus, relative Pratt index remained consistent and robust in identifying variable order across types of correlation matrix, response pattern distribution, number of Likert scale points and variable combination of Likert conditions.

Statistical modeling of the percent bias in relative Pratt index based on the four independent variables revealed a replicate of pattern of results for percent bias in the three predictors, X_1, X_2, and X_3 for each model fit. There was a significant main effect of number of Likert scale points for variable combination of Likert condition in which Y is continuous and X is Likert for all the correlation matrices and distribution of response patterns. Thus, the number of Likert scale points influenced the amount of bias in the estimation of relative Pratt index. The fewer the number of Likert scale points, the larger the percent bias in the estimation of relative

Pratt index. Percent bias was more pronounced in variable combination of Likert condition in which Y is continuous and X is Likert, and where Y is Likert and X is Likert.

The type of correlation matrix and the distribution of response patterns did not affect percent bias in relative Pratt index. Therefore, order and relative importance of the predictors were maintained across the stated conditions. The hypothesis that type of correlation matrix, and response distribution would affect the relative Pratt index was not confirmed. Relative Pratt index was found to be robust to skewness of the distribution and type of correlation matrix of the response data. However, the hypothesis on the effect of number of Likert scales and variable combination of Likert condition on the model fit was confirmed.

Relationship of Findings to Previous Research

No previous studies have been conducted on the stability of relative Pratt index in Likert data either for Likert criterion and continuous predictors or for Likert predictors and continuous criterion. The research hypothesis on relative Pratt index was postulated on the premises that, through the zero order correlation and beta weights associated with each predictor, relative Pratt index would vary within each combination of the independent variables for Likert data. Bias in relative Pratt index was not affected by the distribution of response data but by number of Likert points in the scales. The fewer the Likert points the larger the bias.

The largest percent bias occurred at the two-point Likert scale. However, relative importance of the predictors among themselves was not affected by the relative change in bias. Relative Pratt index was therefore robust to distribution of response, type of correlation matrix and Likert condition. The most important and significant effect was the number of Likert scale points with regard to amount of bias, the fewer the number of Likert points the larger the amount of bias.

Implications of Findings

Theoretical implications of the findings on the effect of Likert data on relative Pratt index is that given the additive property of the relative Pratt index and the fact that it is not affected by the distribution of the data, the index can therefore be used to compute relative importance of variable in both Likert and continuous data with considerable accuracy, computational ease and interpretive clarity compared to other measures of variable importance.

Research implication of the findings on the effect of Likert data on relative Pratt index is that given the robustness of the index across Likert data distribution and types of correlation matrices, researchers may use relative Pratt index in place of other indicators of relative importance such as beta weights, semi-partial correlation coefficients, and communality measures associated with specific predictors. The latter indicators lack additive and ratio properties (Darlington, 1990; Pedhazur, 1982), and are therefore implicitly prone to biases and inaccuracy under different distributions as well as when assumptions made in the model are violated.

Practical implications of the findings on relative Pratt index is similar to the research implication in that practitioners can now use an index that is additive (sum of the relative index measure add up to one for all the predictors in the regression equation) and robust to response distribution data and Likert condition. While relatively large percent bias was observed to persist at fewer Likert scale points, relative order of importance of the predictors remained invariant among the predictors across the range of Likert scales.

Bias in Pearson Correlation Coefficient

To assess the effect of Likert data on the Pearson correlation coefficient of each predictor and the criterion, percent bias of the estimation of the Pearson correlations were plotted and a statistical model fitted for each predictor and criterion. Type of correlation matrix was not expected to affect the estimation of Pearson correlation. However, the distribution of responses, number of Likert points, and variable combination of Likert conditions were hypothesized to affect the estimation of Pearson correlation.

For Likert Y and Continuous X, percent bias in Pearson correlation varied significantly across response distribution. Percent bias was larger in the skewed response distributions with unequal interval than for symmetric distribution with equal interval. Similarly, for Continuous Y and Likert X, percent bias in Pearson correlation was larger for skewed response distributions with unequal interval than for symmetric response distribution with equal interval. At variable combination of Likert condition with both Likert Y and Likert X, the type of response distribution had no effect on percent bias in Pearson correlation coefficients, unlike in the other two conditions of variable combination of Likert condition. Instead, percent bias varied with number of Likert scale points. This implied that bias in Pearson correlation was not significantly affected by type of response distribution when both variables were categorized, but rather by the number of Likert scale points. However, the type of response distribution only affected percent bias in the estimation of Pearson correlation when only one of the variables was categorized. The fewer the Likert points, the higher the percent bias observed. Similar results and conclusion were replicated for bias in Pearson correlation of the criterion and each predictor across the combination of independent variables. These findings can be generalized for low and high values of Pearson correlation coefficients.

Relationship of Findings to Previous Research

While research abounds on the impact of Likert data on Pearson correlation (Bollen & Barb, 1981; Krieg, 1999), findings of those studies were noted to be varied and contradictory, and because of the extra variables as predictors in multiple regression analysis compared to bivariate measures in Pearson correlation, the findings cannot be generalized to multivariate regression analysis. Instead, extensions of the studies need to be conducted to address concerns raised in the studies such as the effect of the type of underlying distribution of the response data and the amount of bias in Pearson correlation resulting from Likert scaling.

This study extended the work of Bollen and Barb (1981) on the effect of Likert data on Pearson correlation coefficient by taking into account non-normal skewed distribution, and unequal interval with varying increase in thresholds. While Bollen and Barb studied the effect of Likert data using Likert Y and Likert X with normal distribution, the present study addressed the effect of Likert Y and Continuous X, as well as Continuous X and Likert Y in addition to both Likert Y and X. This was done for both normal and skewed distributions with equal and unequal interval. Furthermore, Bollen and Barb's study was based on samples generated while results of the present study are based on populations, and so bias in Pearson correlation under the Likert conditions are adequately addressed without confounds from sample to sample variations.

It was found that skewed distribution and unequal interval contributed further to the bias in the estimation of Pearson correlation in addition to the number of Likert scale points. This confirmed the hypothesis on the effects of these variables and further supported findings by Wiley (1976) on the effect of the skewed distribution with equal intervals to the estimation of Pearson correlation in Likert data. Findings of the present study are unique as they are based on both skewed distribution and unequal interval Likert data with varying differences in the threshold values. Combination of these conditions had not been addressed in previous studies on Likert data. Bias resulting in using Likert data in estimating Pearson correlation indicate a substantial loss of information and so appropriate correlation coefficients such as the phi correlation coefficient, biserial correlation coefficient, and polyserial correlation coefficient should be used as the bias in these correlations is expected to be less than in Pearson correlation (Krieg, 1999).

Implications of Findings

Findings on the bias of estimation of Pearson correlation coefficient have theoretical, research, and practical implications which need to be addressed. The theoretical implication of findings is that Pearson correlation has been shown not to be as robust under non-normal distribution as previously thought. The effect of non-normal skewed distribution with unequal interval produced large biases especially in Likert scales with few scale points. This implies that assumptions made regarding the use of Likert data with continuous underlying normal distribution do not hold. Violation of these assumptions leads to underestimation of the Pearson correlation coefficient across different scale points. The use of Sheppard's correction formulae to reduce or eliminate bias was shown by Krieg (1999) not to be adequate in estimating Pearson correlation when the results were compared to simulation studies by Bollen and Barb (1981), and those by Krieg (1999). Therefore, polyserial or polychoric correlation should be used in place of Pearson correlation for Likert data, given that bias results in such uses (Olsson, Drasgow & Dorans, 1982).

Research implication of bias found in estimating Pearson correlation with Likert data are that loss of information resulting from the bias leads to distortions of conclusions made on findings based on such correlation studies. Pearson correlation has also been used in computation of reliability coefficients for Likert scales. Bias resulting from the computation of reliability coefficients based on Pearson correlation leads to underestimation of reliability measures of the instruments and distortion on the psychometric properties of the instrument (Chang, 1994; Cicchetti et.al. 1985; Matell & Jacoby, 1971). Studies on the effect of Likert scale on the estimation of reliability coefficient showed that the fewer the number of Likert scale points the lower the reliability coefficient estimate.

This study did not specifically investigate the effect of Likert data on reliability estimates. However, the impact of Likert data on Pearson correlation, which has been used to estimate reliability coefficient, has psychometric implications to inferences made from such computation, in that the reliability coefficient computed is consistently underestimated. While studies on the effect of Likert data on reliability have not been conclusive on the impact of Likert data together with other variables such as the distribution of responses, discrete and continuous data, inferences can be made from the present findings on the effect on Pearson correlation, as the latter is often used in the computation of reliability coefficients.

Practical implications of the findings of the bias in estimating Pearson correlation is that practitioners may adopt different correlation coefficients for discrete and Likert data in place of Pearson correlation as the study has shown substantial bias associated with the estimation of the Pearson correlation coefficient when using Likert data. This was more pronounced in skewed distribution, in fewer Likert scale points, and when both variables were Likert. Bias and loss of information was largest when both variables were dichotomized. These findings support Cohen (1983) findings on the high cost of dichotomization of variable. Tetrachoric correlation is recommended in such circumstances, as an efficient substitute of Pearson correlation as it has lower amount of bias although large sample sizes are required to realize this (Martin, 1978).

Significant Contribution of the Study

The present study advances the current literature on the impact of Likert data on linear regression models. Most research has concentrated on the impact of Likert data on factor analysis, and Pearson correlation at the expense of multiple linear regression models. The present study therefore fills the gap and expands on the literature on Likert data and multiple regression models using ordinary least-square estimation which frequently occurs in social science research. Furthermore, it was considered in the study that both criterion Y, and the predictors X in the regression model were measured random variables. This is a common practice in social science research context. The study also expanded on the early work of Bollen and Barb (1981) of the impact of Likert scales on Pearson correlation and highlighted the properties and advantages of using the relative Pratt index in determining relative variable importance in multiple linear regression models with Likert data.

The practical contribution and significance of the present study is that it provides information on the extent to which information is lost when researchers choose to treat and analyze Likert variables as if continuous in linear regression models. Combinations of continuous and categorical variables among predictors and criterion variables were investigated to provide an in-depth understanding of the extent of information loss that is essential in the valid and accurate interpretation of linear regression models that are frequently applied in education, and social science research.

The study examined to what extent accuracy is compromised in using Likert variables and the validity implications on the interpretation and utilization of findings derived from linear regression models. Based on the findings it is recommended that treating Likert variables as if continuous in linear regression models will need to be revisited and appropriate methods adopted for Likert data, for better precision and accurate interpretation of results, given the impact of the number of Likert scale points and the occurrence of non-normal data in practice. For Pearson correlation, the largest bias occurred within the dichotomized scales and lower Likert scale points, and when the response distribution was non-normal (skewed) implying that phi-correlation, tetrachoric and polyserial correlation should be used in place of the Pearson correlation for polytomous items or Likert variables for either predictors or criterion. With regard to relative importance and variable ordering, the relative Pratt index was found to be robust to response distribution and number of Likert scale points as well as types of inter-variable correlation matrices. Note that variable condition 2 (i.e., continuous Y and Likert X) can be found in practice when one uses the total scale score (even when each item of the scale are Likert-type) as the criterion and individual Likert items as predictors.

Limitations of the Study

The study is based on simulated data and not real responses of a Likert scale with known psychometric properties. This could be a criticism of the method. However, simulations provide an opportunity for manipulation of variables, which is a positive research characteristic of an experimental design. The number of Likert scales ranged from 2 to 9 scale points, which were selected as they typically occur in social science literature. Due to logistic reasons, it was not possible to study more scale points. Even then, the range of scale points considered represents what is typically encountered in surveys and social science research. Other studies have addressed Likert scale points ranging from 2 to 10 (Krieg, 1999) 2 to 12 (Wiley, 1976), 2 to 20 (Martin, 1978). However, as this study shows there is no significant information gain in using Likert scales with more than four scale points. This is particularly evident in skewed Likert data that is frequently encountered by researchers.

Future Research

Only three predictors and one criterion were included in this study. However, in future studies, more predictors should be studied to examine whether the number of predictors, and implicitly, the size of the correlation matrix has an effect on the estimation of the dependent variables in the study. No interaction terms were included in the linear regression model in this study. Future studies should assess the impact of Likert data on linear regression models with interaction terms and the estimation of R^2.

In this study the number of Likert scale points were set to be equal for both criterion and predictors. However, different combination of number of Likert scale points in the criterion and predictors should also have been included to further enhance generalizability of the findings to all Likert conditions. The effect of Likert data on relative Pratt index with negative values and the possible presences of suppressor variables were not separately addressed, as special cases of the impact of using Likert data on linear regression models. The stability of relative Pratt index in linear regression models with interaction terms was also not addressed, as only non-interaction terms were included in the model.

Future research should address limitations, together with research questions arising from the study as well as the literature on the impact of Likert data on linear regression models. The impact of Likert data on the estimation of three dependent variables studied namely R^2, Pearson correlation coefficient, and relative Pratt index was done using a population of responses. To further examine the stability of the resulting indices under different sample sizes and sample-to-sample variations, future studies will need to replicate the current study but using several sample sizes.

In the present study three types of inter-variable correlation matrices generated from three predictors were used. However, future studies will need to include more than three predictors with larger inter-variable correlation matrices to determine if the size and magnitude of the inter-variable correlation matrices have an impact on the model fit in linear regression models using Likert data within the independent variables.

The Likert condition of the criterion and the predictors were set to be the same for each combination of condition of the independent variable i.e., whenever the criterion had a four-point scale so were the predictors. However, future studies will need to consider the impact of Likert data on the three dependent variables namely R^2, Pearson correlation, and relative Pratt index in terms of bias in the estimation, and the stability of the indices for different Likert conditions and scale points in the criterion and the predictors i.e. a mixed ratio of Likert scale points for criterion and predictors. Furthermore, a mixed data format with some predictors

Likert and some continuous with a Likert criterion should also be explored to determine the effect of the Likert data on bias and stability of the three dependent variables under the conditions of the independent variables already stated. The effect of Likert data on linear regression models was studied for models without interaction terms. Future studies will need to address the effect of Likert data in linear regression models with interaction terms in the model. In the case of relative Pratt index, further studies need to address the stability and bias of the relative Pratt index in Likert data as well as the relationship between negative relative Pratt index and moderators and suppressor variables in the linear regression model.

The underlying assumption in fitting a linear regression model is that the response data is normally distributed. This study has shown that under non-normal distribution and Likert data condition both R^2 and Pearson correlation were underestimated. However, only two levels of skewness (positive, and negative) were addressed. The effect of non-normal response distribution with different kurtosis and more than three levels of skewness will need to be examined in order to further determine the impact of skewness and kurtosis on the estimation of R^2 and Pearson correlation coefficient in Likert data. In addition to this, non-normal latent distribution will also need to be investigated as the present study and earlier studies focused on normally distributed latent variables.

The assumption of equal interval in the Likert scales was also studied by considering equal and unequal interval scales. Unequal intervals were created using the method proposed by Bollen and Barb (1981) based on the partitioning of the normal distribution of the response data to reflect differences in threshold. Future studies will need to consider other methods of creating unequal intervals with different threshold to determine if the impact of unequal interval was invariant across method. Given the degree of bias in estimating R^2, future research should investigate how probit and logit in ordinal regression models compare with ordinary least-square regression model fit.

For relative Pratt index, it was observed that skewness, and by implication, violation of normality assumption of the response distribution was found not to impact variable ordering and importance. Future studies will need to address the impact of Likert data on skewness and negative relative Pratt index and its meaningful interpretation, as questions have been raised concerning negative values of the index and occurrence of multicollinearity in linear regression models (Bring, 1996; Pratt, 1987; Thomas, Hughes & Zumbo, 1998). Furthermore, the stability of relative Pratt index in linear regression models with moderators and suppressor variables using Likert data will also need to be addressed.

A population of responses was simulated to avoid sample-to-sample variation of the dependent variables and to adequately address the issue of bias in the estimation of R^2, Pearson correlation coefficient, and the relative Pratt index across the independent variables. However, to study the stability of the estimates of the dependent variables, the study should have addressed the effect of various sample sizes across combinations of the independent variables and Likert data on the linear regression model. This will be addressed in future studies.

Recommendations

It is recommended that researchers and practitioners recognize the substantial bias and loss of information that occurs in using Likert data in linear regression models. It is suggested from the results of the study that researchers use other methods of regression model fit such as ordinal regression, and logistic regression models when using ordinal categorical or Likert data to overcome bias and loss of information (for a review of the extensive literature, see Ananth & Kleinbaum, 1997; McCullagh, 1980). Ordinal regression models take into account the type of response distribution, and ordinal and categorical nature of the responses through appropriate probability transformation functions.

Ordinal regression as a recommended approach for analyzing ordinal or Likert data is now commonly used in biomedical research work e.g. epidemiology. Computer programs for ordinal regression model fit are now readily available in computer software such as SPSS and SAS. Given that educational researchers often encounter ordinal data in the form of Likert scales, and fit ordinary least-square linear regression models, it is recommended that they move in the same direction as their biomedical research counterparts and adopt ordinal regression model fit to overcome information loss as manifested in the ordinary least-square linear regression model fit.

Violation of the normality assumption on the distribution of the data also resulted in bias. Selection of appropriate models and application of suitable data transformation is recommended in determining and minimizing bias in both R^2 and Pearson correlation in Likert data. Readers and consumers of research need to treat with caution results and inferences derived from linear regression models using Likert data with non normal response distribution because of the large bias encountered in the model fit.

Few Likert scale points resulted in large biases. It is recommended that four Likert scale points or more be used. However, little or no substantial gain in information results in using more than four Likert points as evidenced by the study. Dichotomization of data results in substantial loss of information. Therefore, researchers should not dichotomize data unless it is inevitable for example, in variables such as gender. Appropriate correlation measures for bivariate dichotomized data such as phi-coefficient, and biserial correlation may be used instead of Pearson correlation. For Likert scales, polyserial and polychoric correlation may be used. However, caution should be taken because polychoric correlation matrices maybe non-Gramian. This leads to difficulties in obtaining inverses of the matrices with no definite solutions in computing beta-weights.

For variable ordering and relative importance, it is recommended that the relative Pratt index be used in place of other indices as it is robust to type of response distribution, type of correlation matrix and Likert scaling except for the number of Likert points. However, the order of variables remains unchanged irrespective of the number of Likert scale points despite the bias in the magnitude of the relative Pratt index.

References

Ananth, C. V. & Kleinbaum, D. G. (1997). Regression models for ordinal responses: A review of methods and applications. International Journal of Epidemiology, 26, 6, 1323-1333.

Babakus, E., Ferguson, C.E, & Joreskog, K.G.(1987). The sensitivity of confirmatory maximum likelihood factor analysis to violation of measurement scale and distributional assumptions. Journal of Markerting Research, 24, 222-228.

Bendig, A. W. (1953). The reliability of self-ratings as a function of the amount verbal anchoring and the number of categories of scales. Journal of Applied Psychology, 37, 38-41.

Bentler, P.M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238-246.

Bernstein, I.H. & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. Psychological Bulletin, 105, 3, 464-477.

Bollen, K. A. (1989). Structural equations with latent variables. New York: John Wiley.

Bollen, K.A., & Barb, K. (1981). Pearson's r and coarsely categorized measures. American Sociological Review, 46, 232-239.

Box, G. E. P., & Draper, N. R.(1987). Empirical model building and response surfaces. New York: John Wiley & Sons, Inc.

Bring, J. (1994). How to standardize regression coefficients. The American Statistician, 48, 209-213.

Bring (1996). A geometric approach to compare variables in a regression model. American Statistician, 50, 57-62.

Champney, H., & Marshall, H. (1939). Optimal refinement of the rating scale. Journal of Applied Psychology, 23, 323-331.

Cicchetti, D. V., Showalter, D., & Tyler, P. J. (1985). The effect of number of rating scale categories on levels of inter rater reliability: A Monte Carlo investigation. Applied Psychological Measurement, 9, 1, 31-36.

Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert type scales in relation to reliability and validity. Applied Psychological Measurement, 18,3, 205-215.

Cliff, N. (1994). Predicting ordinal relations. British Journal of Mathematical and Statistical Psychology, 47, 127-150.

Cohen, J. (1983). The cost of dichotomization. Applied Psychological Measurement, 7, 3, 249-253.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles : β is not enough. Educational and Psychological Measurement, 61, 2, 229-248.

Cox, N, R. (1974). Estimation of the correlation between continuous and discrete variable. Biometrics, 30, 171-178.

Curran, P.J., West, G., & Finch, J.F.(1995). Structural equation models with non-normal variables: Problems and remedies. In Hoyle, R.H. (ed). Structural equation modeling: Concepts, issues and applications. Newbury park CA: sage.

Curran, P.J., West, G., & Finch, J.F. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. Psychological Methods, 1, 16-29.

Darlington, R. B. (1990). Regression and linear models. New York: McGraw-Hill.

Dawes, R.M (1972). Fundamentals of Attitude Measurement. New York: John Wiley & Sons Inc.

Eom, H. J. (1993). The interactive effects of data categorization and non-circularity on the sampling distribution of generalizability coefficients in Analysis of Variance models: An empirical investigation. Unpublished Dissertation, University of British Columbia.

Green, S.B., Akey, T.M., Fleming, K.K., Hershberger, S.L. & Marquis, J.G. (1997). Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis. Structural Equation Modeling, 4, 2, 108-120.

Grether, D.M (1976). On the use of ordinal data in correlation analysis. American Sociological Review, 41, 908-912.

Henry, F. (1982). Multivariate analysis and ordinal data. American Sociological Review, 47. 299-307.

Hoaglin, D. C., & Andrews, D. F. (1975). The reporting of computation-based results in statistics. The American Statistician, 29, 122-126.

Jenkins, G. D., & Taber, T.D. (1977). A Monte Carlo study of factors affecting three indices of composite scale reliability. Journal of Applied Psychology, 62,392-398.

Johnson, D.R., & Creech, J.C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. American Sociological Review, 48, 598-407.

Khuri, A. I., Cornell, J. A (1987). Response Surfaces: Designs and Analyses. New York: Marcel Dekker, Inc.

Kim, J. (1978). Multivariate analysis of ordinal variables revisited. American Journal of Sociology, 84, 448-456.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. Educational and Psychological Measurement, 56, 746-759.

Komorita, S. S., & Graham, W. K. (1965). Number of scale points and reliability of scales. Educational and Psychological Measurement, 25, 987-995.

Krieg, E. F. (1999). Bias induced by coarse measurement scales. Educational and Psychological Measurement, 59, 5, 749-766.

Kruskall, W.(1987). Relative importance by averaging over oderings. The American Statistician, 41, 6-10.

Labovitz, S. (1975). Comment on the Henkel's paper: the interplay between measurement and statistics. Pacific Sociological Review, 18, 27-35.

Lissitz, R. W. & Green, S. B. (1975). Effect of number of scale points on reliability: A Monte Carlo approach. Journal of Applied Psychology, 60, 10-13.

Long, J.D. (1999). A confidence interval for ordinal multiple regression weights. Psychological Methods, 4,3, 315-330.

Martin, W. S. (1978). Effect of scaling on the correlation coefficient : Additional considerations. Journal of Marketing Research, 15, 304-308.

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. Educational and Psychological Measurement, 31, 657-674.

McCullagh, P. (1980). Regression models for ordinal data. Journal of Royal Statistical Society Series B 42, 2,109-142.

Muthen, B. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. British Journal of Mathematical and Statistical Psychology, 38, 171-189.

Muthen, B. & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. British Journal of Mathematical and Statistical Psychology, 45, 19-30.

Nanna, M. J., & Sawilowsky, S.S. (1998). Analysis of Likert scale data in disability and medical rehabilitation research. Psychological Methods, 3, 55-67.

Nunnally, J., & Bernstein, I.H.(1994). Psychometric theory (3rd ed.) New York: McGraw-Hill.

O'Brien, R.M. (1979). The use of Pearson's r with ordinal data. American Sociological Review, 44, 851-857.

O'Brien, R.M. (1981). Using rank category variables to represent continuous variables. Social Forces, 59,1149-1162.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. Psychometrika, 44, 443-460.

Olsson, U., Drasgow, F., & Dorans, N. (1982). The polyserial correlation coefficient. Psychometrika, 47, 3, 337-347.

Pearson, K. (1913). On the measurement of the influence of broad categories on correlation. Biometrika, 9, 116-139.

Pedhazur, E. J. (1982). Multiple Regression in Behavioral Research. New York: Holt, Rhinehart, and Winston.

Peters, C. C. & Van Voorhis, W. R. (1940). Statistical procedures and their mathematical bases. New York: McGraw-Hill.

Pratt, J. W. (1987). Dividing the indivisible: using simple symmetry to partition variance explained, In Pukila, T. and Puntanen, S. (Eds), Proceedings of the Second International Conference in Statistics, (pp 245-260), Tampere, Finland: University of Tampere.

Russell, C. J., & Bobko, P. (1992). Moderated regression analysis and Likert scales: too coarse for comfort. Journal of Applied Psychology, 77, 336-342.

Russell, C. J., Pinto, J. K., & Bobko, P. (1991). Appropriate moderated regression and inappropriate research strategy: A demonstration of information loss to scale coarseness. Applied Psychological Measurement, 15, 3, 257-266.

Stevens, J. (1986). Applied Multivariate Statistics for the Social Sciences. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Symonds, P. M. (1924). On the loss of reliability in ratings due to coarseness of the scale. Journal of Experimental Psychology, 7, 456-461.

Thomas, D. R., Hughes, E., & Zumbo, B.D. (1998). Variable importance in regression and related analyses. Social Indicators Research: International Journal for Quality-of-Life Measurement, 45, 253-275.

Thomas, D. R., & Zumbo, B. D. (1996). Using a measure of variable importance to investigate the standardization of discriminant coefficients. Journal of Educational and Behavioral Statistics, 21, 110-130.

Wiley, P. B. (1976). Effect of coarse grouping and skewed marginal distributions on Pearson correlation coefficient. Educational and Psychological Measurement, 36, 1-7.

Winship, C. & Mares, R. (1984). Regression models with ordinal variables. American Sociological Review, 49, 512-525.

Zumbo, B. D., & Harwell, M. R. (1999). The methodology of methodological research: Analyzing the results of simulation experiments. Paper number ESQB-99-2. Prince George, B.C: University of Northern British Columbia. Edgeworth Laboratory for Quantitative Behavioral Science.

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

APPENDIX A

SCALE INTERVALS THRESHOLDS

Table 5

Threshold values for the scale intervals in symmetric response distribution with equal interval

EQUAL

1	2	3	4	5	6	7	8	9
0.0000								
-1.0000	1.0000							
-1.5000	0.0000	1.5000						
-1.8000	-0.6000	0.6000	1.8000					
-2.0000	-1.0000	0.0000	1.0000	2.0000				
-2.1429	-1.2857	-0.4286	0.4286	1.2857	2.1429			
-2.2500	-1.5000	-0.7500	0.0000	0.7500	1.5000	2.2500		
-2.3333	-1.6667	-1.0000	-0.3333	0.3333	1.0000	1.6667	2.3333	

APPENDIX B

Table 6

Threshold values for the scale intervals in negatively skewed response distribution (right bunching) with unequal interval

BUNCHED TO the RIGHT

1	2	3	4	5	6	7	8	9
-1.5000								
-1.5000	0.0000							
-2.0000	-1.0000	0.0000						
-2.2500	-1.5000	-0.7500	0.0000					
-2.4000	-1.8000	-1.2000	-0.6000	0.0000				
-2.5000	-2.0000	-1.5000	-1.0000	-0.5000	0.0000			
-2.5714	-2.1429	-1.7143	-1.2857	-0.8571	-0.4286	0.0000		
-2.6250	-2.2500	-1.8750	-1.5000	-1.1250	-0.7500	-0.3750	0.0000	

APPENDIX C

Table 7

Threshold values for the scale intervals in positively skewed response distribution (left bunching) with unequal interval

BUNCHED TO the LEFT

1	2	3	4	5	6	7	8	9
1.5000								
0.0000	1.5000							
0.0000	1.0000	2.0000						
0.0000	0.7500	1.5000	2.2500					
0.0000	0.6000	1.2000	1.8000	2.4000				
0.0000	0.5000	1.0000	1.5000	2.0000	2.5000			
0.0000	0.4286	0.8571	1.2857	1.7143	2.1429	2.5714		
0.0000	0.3750	0.7500	1.1250	1.5000	1.8750	2.2500	2.6250	