Summer 8-11-2011

Robustness Of Two Formulas To Correct Pearson Correlation For Restriction Of Range

minh tran

ACCEPTANCE

This dissertation, ROBUSTNESS OF TWO FORMULAS TO CORRECT PEARSON CORRELATION FOR RESTRICTION OF RANGE, by DUNG MINH TRAN, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree Doctor of Philosophy in the College of Education, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chair, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty. The Dean of the College of Education concurs.

William Curlette, Ph.D.
Committee Chair

Susan Ogletree, Ph.D.
Committee Member

Chris Oshima, Ph.D.
Committee Member

Phillip Gagne, Ph.D.
Committee Member

Date

Sheryl Gowen, Ph.D.
Chair, Department of Educational Policy Studies

R. W. Kamphaus, Ph.D.
Dean and Distinguished Research Professor
College of Education

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education's director of graduate studies and research, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my written permission.

____

Dung Minh Tran

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University Library must be used only in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Dung Minh Tran
2621 Lavista Rd.
Decatur, GA 30033

The director of this dissertation is:

Dr. William Curlette
Georgia State University
Department of Educational Policy Studies
P.O. Box 3977
Atlanta, GA 30302-3977

CURRICULUM VITAE

Dung Minh Tran

ADDRESS Dung Minh Tran
2621 Lavista Rd.
Decatur, GA 30033

EMAIL minhttttran@yahoo.com

EDUCATION

Ph.D.	(2011)	Research, Measurement, and Statistics Georgia State University
M.S.	(1992)	Computer Science Georgia State University
M.B.A.	(1998)	Computer Information Systems Georgia State University

PROFESSIONAL EXPERIENCE

2002-present	Technical Application Developer III, Emory Healthcare (Emory University Hospitals)
2005-2008	GRA, Georgia State University
2000-2002	Web Implementation Manager, Emory Healthcare (Division of Marketing)
1998-2000	Senior Application Developer, Emory University (Division of Finance)
1992-1998	Senior Programmer Analyst, Georgia State University (Division of Information Systems and Technology)

ABSTRACT

ROBUSTNESS OF TWO FORMULAS TO CORRECT
PEARSON CORRELATION FOR
RESTRICTION OF RANGE
by
Dung Minh Tran

Many research studies involving Pearson correlations are conducted in settings where one of the two variables has a restricted range in the sample. For example, this situation occurs when tests are used for selecting candidates for employment or university admission. Often after selection, there is interest in correlating the selection variable, which has a restricted range, to a criterion variable. The focus of this research was to compare Alexander, Alliger, and Hanges's (1984) formula to Thorndike's (1947) formula and population values using Monte Carlo simulation when the assumption of normal distribution is violated in a particular way.

In both Thorndike's and Alexander et al.'s correction formulas, values for the variances in the restricted and the unrestricted situations are required. For both formulas, the variance in restricted situations was a sample estimate. In the Monte Carlo simulation, the difference between the two approaches was that in Thorndike's formula, the variance in the unrestricted situation was the population variance from the exogenous variable, whereas in Alexander et al.'s approach, the population variance was estimated based on the sample variance in the restricted situation. In the simulation, robustness situations were created from non-normal distributions for predicted group membership in a classification problem.

As expected, Thorndike's corrected correlation values were more accurate than Alexander et al.'s corrected correlation values, and Thorndike's formula had a smaller

standard error of estimates. Absolute values of the mean differences between the estimated and population correlations for Alexander et al.'s approach compared to Thorndike's approach in robustness situations ranged from 1.37 to 2.15 larger. Nevertheless, Alexander et al.'s approach, which is based only on estimated variances, appears to be a worthwhile correction in most of the simulated situations with a few notable exceptions for non-normal distributions.

ROBUSTNESS OF TWO FORMULAS TO CORRECT
FOR PEARSON CORRELATION FOR
RESTRICTION OF RANGE

by
Dung Minh Tran

A Dissertation

Presented in Partial Fulfillment of Requirements for the
Degree of
Doctor of Philosophy
in
Research, Measurement, and Statistics
in
the Department of Educational Policy Studies
in
the College of Education
Georgia State University

Atlanta, GA
2011

ACKNOWLEDGMENTS

I dedicate my work in the loving memory of my father, Dr. Tran Minh Man. He inspired me to be the man that he was in terms of his commitment, dedication, and care for others.

I'd like to thank to my dissertation's chair, Dr. William Curlette, for helping and guiding me. I'd like to thank to Dr. Chris Oshima, Dr. Phillip Gagne, and Dr. Susan Ogletree for helping and supporting me during this process. I also want to express my thanks to my mother Nguyen Thi Hanh and brother Tran Minh Tan for their continuous support.

TABLE OF CONTENTS

LIST OF TABLES

ABBREVIATIONS

Thorndike	Thorndike's Formula
Alexander	Alexander et al.'s Formula
Thorn	Thorndike
T1	Thorndike Area1
T12	Thorndike Areas 1 and 2
A1	Alexander et al.'s Area 1
A12	Alexander et al.'s Areas 1 and 2
Alex	Alexander et al.
MSE	Mean Square Error

CHAPTER 1

VARIOUS CORRELATION FORMULAS TO CORRECT FOR RESTRICTION OF RANGE

In this dissertation, which follows the manuscript style format approved by the College of Education, the first chapter provides a literature review as background for the study. Following this chapter, the second chapter employs an extended manuscript format, supplemented by appendices, to describe the study. Many research studies involving Pearson correlations are conducted in settings where one of the two variables has a restricted range in a sample. For example, this occurs after tests are used for selecting people for employment or admission to school, resulting in a distribution which has been truncated. Often after selection, there is interest in correlating the selection variable, which has a restricted range, to a criterion variable. The focus of this review is on Thorndike's (1947) correction formula and Alexander et al.'s (1984) correction formula for use with Pearson's correlation formula for restriction of range, although I do mention other formulas.

Thorndike's (1947) correction formula uses known variance for the unrestricted situation to obtain an estimate of the corrected correlations, and Alexander et al.'s (1984) formula with unknown variance for the unrestricted sample uses Cohen's (1959) formula to obtain an estimate of the unrestricted variance. In other words, Cohen's approach, which is incorporated in Alexander et al.'s formula, uses the restricted variance and assumption of a normal distribution to estimate the unrestricted variance. The known variance for the unrestricted situation in Thorndike's formula is sometimes obtained from normative tables for an existing educational or psychological test.

In this review, there are two types of restriction of range to be considered, direct restriction of range and indirect restriction of range. Direct restriction of range occurs when there is a restriction of range on one of the two variables of interest. For example, if a researcher is interested in two variables, x and y, then the restriction of range occurs on variable x or variable y. Indirect restriction of range occurs when restriction of range occurs on a variable other than the two variables of interest. For instance, if a researcher is interested in two variables, x and y, then the restriction of range occurs on variable z. Restrictions of range and cumulative meta-analysis have a strong connection because one type of meta-analysis summarizes correlation coefficients from different studies. When the studies involve a correlation of a test with a criterion, this form of meta-analysis is known as validity generalization and is a potential application for Alexander et al.'s formula.

This literature review will present Pearson's correlation, Thorndike's (1947) formula, Cohen's (1959) ratio, Alexander et al.'s (1984) formula using Cohen's ratio, direct and indirect restriction of range, restriction of range in meta-analysis, restriction of range with attenuation, and contaminated normal as a background for the study reported in manuscript style in Chapter 2.

Pearson's Correlation

Pearson's correlation is a number between −1 and +1 that measures the relationship between the two variables. A positive number implies a positive association, whereas a negative number implies the inverse association. Pearson's correlation is a measure of the relationship between two variables x and y, and it could be defined in terms of the population correlation, $\rho_{x,y}$, where

$$\rho_{x,y} = COV(x,y) / \sigma_x\sigma_y \qquad 1$$

with the corresponding sample correlation $r_{x,y}$, given by

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \qquad 2$$

Here, COV(x,y) is the population correlation between x and y, σ_x is the population standard deviation of x, and σ_y is the population standard deviation of y. In the above formula, s_x, s_y are the sample standard deviations of x and y, respectively. The term $\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)}$ is the sample covariance. Additionally, Pearson's correlation could be expressed in terms of z-scores of x and y when the population means and population standard deviations of x and y are available. Goodwin and Leech (2006) stated that the Pearson's correlation could be defined in terms of z-score of x and y as follows:

$$\rho_{x,y} = \sum (z_x z_y) / N \qquad 3$$

where z_x is the z-score of the x variable, calculate using the population μ_x, and standard deviation σ_x,

z_y is likewise the z-score of the y variable, and

N is the number of pairs of scores.

The square of the correlation or the coefficient of determination (i.e., $\rho_{x,y}^2$) explains the portion of the shared variance, or the fraction of variance in one variable, x, that could be explained by the other variable, y. Furthermore, Rodgers and Nicewander (1988) suggested 13 different ways of interpreting a correlation value. Some of most common ways included the interpretation of a correlation value (1) as the standardized slope of the regression line in a z-score format, (2) as the proportion of variability in common between variables x and y, and (3) as a function of test statistics. Rovine and Von (1997)

added a 14th way to interpret the correlation value as the proportion of matches of the two variables of interest, x and y.

Thorndike's Formulas for Restriction of Range

Thorndike (1947) presented formulas for correcting restriction of range for Pearson's (1904) formula. Thorndike developed three formulas to calculate the restriction of range under different circumstances.

Thorndike's Formula 1

The first formula is used to estimate the correlation between two variables of interest, x and y, when the range of restriction occurs on variable y; the observed correlation between the two variables of interest, x and y, is known; and the standard deviations of the unrestricted sample and the restricted sample of variable y are also known. Thorndike's Formula 1 is expressed as

$$r_{x,y\ \text{unrestricted}} = \sqrt{1 - \frac{SD^2_{y\ \text{restricted}}}{SD^2_{y\ \text{unrestricted}}}\ (1 - r^2_{x,y\ \text{restricted}})} \qquad 4$$

where $r_{x,y\ \text{unrestricted}}$ = correlation between variables x and y in an unrestricted sample,
$r_{x,y\ \text{restricted}}$= correlation between variables x and y in a restricted sample,
$SD_{y\ \text{unrestricted}}$= standard deviation of the variable y in an unrestricted sample, and
$SD_{y\ \text{restricted}}$= standard deviation of the variable y in a restricted sample.

It should be noted that the ratio of restricted standard deviation to unrestricted standard deviation is not on the variable for which the restriction occurred. Thorndike (1947) stated that "This situation was rarely encountered in practice" (p. 65).

Thorndike's Formula 2

Thorndike's second formula estimates the correlation between the two variables of interest, x and y, when the restriction of range occurs at variable x; the observed correlation value between the variables of interest, x and y, is available; and the standard deviations of the unrestricted and restricted distribution x are also known. Thorndike's Formula 2 is

$$r_{x,y\ \text{unrestricted}} = \frac{r_{x,y\ \text{restricted}}\left(\frac{SD_{x\ \text{unrestricted}}}{SD_{x\ \text{restricted}}}\right)}{\sqrt{1-r^2_{x,y\ \text{restricted}}+[\ r^2_{x,y\ \text{restricted}}\left(\frac{SD^2_{x\ \text{unrestricted}}}{SD^2_{x\ \text{restricted}}}\right)]}} \qquad 5$$

where $r_{x,y\ \text{unrestricted}}$ is the correlation of x and y in an unrestricted sample,

$r_{x,y\ \text{restricted}}$ is the correlation of x and y in a restricted sample,

$SD_{x\ \text{unrestricted}}$ is the standard deviation of variable x in an unrestricted sample, and

$SD_{x\ \text{restricted}}$ is the standard deviation of variable x in a restricted sample.

In contrast to Formula 1, Thorndike's Formula 2 is for situations where the required information about restricted standard deviation to unrestricted standard deviation is on the same variable that had restricted range. Formula 2 is commonly used in real-world situations. For example, Oleksandr and Deniz (1999) illustrated a case when Thorndike's (1947) Formula 2 was used in the Graduate Record Examination (GRE) validation studies with students who were already enrolled at the school. Because the selection was based on GRE scores, the range of scores of the students is restricted (i.e., most of the GRE scores in the sample are high). In general, no criterion information is available for low-scoring persons because these applicants are not admitted to the graduate program. Although the researchers could compute the correlation between GRE

score and the graduate performance criterion in the restricted sample of the students that are already enrolled at the school, they cannot compute the correlation for the total group of applicants who applied to the graduate school. Thus, the correlation for the total group of graduate applicants is not immediately available using Pearson's correlation formula alone. Thorndike's (1947) Formula 2 was used to estimate the corrected correlation from an unrestricted sample (i.e., the total group of applicants who apply to the graduate school) from the correlation of the restricted sample (i.e., the correlation of students who are already admitted to the graduate program).

Thorndike's Formula 3

Alexander et al. (1990) note that there are two types of restrictions of range: (1) the direct restriction of range and (2) the indirect restriction of range. The direct restriction of range occurs when there is a restriction of range at one of the two variables of interest; for example, if there are two variables of interest, x and y, the range restriction occurs on variable x. The indirect restriction of range occurs at some third variable, z, other than the two variables of interest, x and y. Thorndike's indirect restriction of range Formula 3 is

$$r_{x,y\ un} = \frac{r_{x,y\ re} + r_{x,z\ re} r_{y,z\ re}\ [(SD^2_{z\ un}/SD^2_{z\ re}) - 1]}{\sqrt{[1 + r^2_{x,z\ re}((SD^2_{z\ un}/SD^2_{z\ re}) - 1)][1 + r^2_{y,z\ re}((SD^2_{z\ un}/SD^2_{z\ re}) - 1)]}} \qquad 6$$

where $r_{i,j\ un}$= correlation of x and y in an unrestricted sample,

$r_{i,j\ re}$= correlation of x and y in a restricted sample,

$SD_{i\ unrestricted}$= standard deviation of the variable i in an unrestricted sample, and

$SD_{i\ restricted}$= standard deviation of the variable i in a restricted sample.

In some situations, the observed correlation value between variables x and z is not known, and Formula 3 is expressed in terms of the correlation between variables x and z. Thus, Thorndike's (1947) indirect restriction of range formula could be rewritten as

$$r_{x,y\ un} = \frac{r_{x,y\ rest}\sqrt{1+r^2_{x,z\ un}\left(\frac{SD^2_{z\ rest}}{SD^2_{z\ un}}-1\right)}+r_{x,z\ un}r_{y,z\ rest}\left(\frac{SD_{z\ un}}{SD_{z\ rest}}-\frac{SD_{z\ rest}}{SD_{z\ un}}\right)}{\sqrt{1+r^2_{y,z\ rest}\left(\frac{SD^2_{z\ un}}{SD^2_{z\ rest}}-1\right)}} \qquad 7$$

where $r_{i,j\ un}$ = correlation between variables i and j in an unrestricted sample,
$r_{i,j\ rest}$ = correlation between variable i and j in a restricted sample,
$SD_{i\ un}$ = standard deviation of the variable i in an unrestricted sample, and
$SD_{i\ rest}$ = standard deviation of the variable i in a restricted sample.

Thorndike's (1947) third formula (Formula 3), an indirect restriction of range formula, addresses correcting the correlation between the two variables of interest, x and y, when the restriction of range occurs at a third variable, z. Additionally, Thorndike's Formula 3, the indirect restriction of range, estimates the corrected population correlation between the variables of interest, x and y, when the following are known: (1) the observed correlation values between the variables x and z and between y and z, (2) the standard deviation of z in the unrestricted sample, and (3) the standard deviation of z in the restricted sample.

The Cohen Ratio

Cohen (1959) proposed a ratio of sample variance in the restricted sample over the square of difference between the sample mean in the restricted sample and the point of truncation in the restricted sample. The formula for the Cohen ratio is:

$$\text{Cohen ratio} = \frac{SD^2_{\text{restricted}}}{(x'_{\text{restricited}} - x_{\text{c restricted}})^2} \qquad 8$$

where $SD^2_{\text{restricted}}$ is the sample variance in a restricted sample,

$x'_{\text{restricited}}$ is the mean in a restricted sample, and

$x_{\text{c restricted}}$ is the highest sorted x value in a restricted sample.

To apply this formula, sort in ascending order the array of elements of the restricted sample from the lowest value to the highest value. The $x_{\text{c restricted}}$ value is the highest sorted value in the restricted sample.

The Cohen ratio is used to find the restricted standard normal standard deviation and the z-score from Cohen's table (see Appendix B). Cohen's table is based on the fact that his formula results in a unique value for truncation point $x_{\text{c restricted}}$. More specifically, the Cohen table has three columns. The first column is the Cohen ratio; the second column represents the table restricted normal standard deviation (i.e., SD_{tab}), which is the "standardized value of the standard deviation after truncation (with a non-truncated value of 1.0). That table value also represents the proportion reduction in standard deviation due to range restriction" (Alexander et al., 1984, p. 432). The third column in the Cohen table represents the z-score. From the table, the restricted normal standard deviation value and its z-score are found by doing a table look-up based on Cohen's ratio.

Alexander et al.'s Formula

Since the unrestricted standard deviation is not directly known in Alexander, Alexander et al.'s (1984) formula uses the Cohen ratio to obtain the SD_{tab} value from Cohen's table (see Appendix B). As Alexander et al. (1984) points out, "Cohen's ratio has the advantage of ease of calculation from sample data" (p. 432). Once the SD_{tab} value

is obtained, Alexander et al.'s formula estimates the unrestricted standard deviation based on the values of the observed standard deviation in the restricted sample and the table restricted normal standard deviation in the restricted sample. Alexander et al.'s formula estimates the unrestricted standard deviation as follows:

$$SD' = SD_{obs} / SD_{tab} \qquad 9$$

where SD' is the estimate of the unrestricted standard deviation,

SD_{obs} is the observed standard deviation in the restricted sample, and

SD_{tab} stands for avalue for the restricted normal standard deviation.

The SD' and SD_{obs}values will then be used to calculate the U ratio, which in turn will be used to estimate the corrected correlation using Thorndike's (1947) Formula 2. Using the U ratio and Thorndike's (1947) Formula 2, Alexander et al. (1984) developed the following formula:

$$r' = (r * \frac{SD'}{SD_{obs}}) / SQRT\ (1 - r^2 + (r^2 * \left(\frac{SD'}{SD_{obs}}\right)^2)) \qquad 10$$

where r' is the corrected correlation value,
r is the observed correlation value in the restricted sample,
SD' is the estimate of the unrestricted standard deviation, and
SD_{obs} is the observed standard deviation in the restricted sample.

Alexander et al. (1984) employed a Monte Carlo computer simulation program to compare the means of the observed correlation values with the corresponding means of the corrected estimated correlations using his formula. In his computer simulation program, Alexander et al. used one sample size (i.e., N = 60), different ρ values (i.e., 0.2,

0.4, 0.6, 0.8), and various truncation values (i.e., −2.0, −1.5, −1.0, −0.5, 0, 0.5, 1.0). The computer simulation produced the table of comparison between the two means (i.e., observed correlation values and the corrected correlation values), which is presented in Table 1. Alexander et al. did not investigate non-normality nor did he compare his mean correlation estimates with estimates from Thorndike's formula.

Even though the unrestricted variance was estimated through Alexander et al.'s (1984) formula, in every case except for the truncation value at −2.0, the results of the mean estimated corrected correlation values were closer to the true ρ correlation values than the mean observed correlation values, as can be seen in Table 1. Similarly, in every case except for the severe cut (i.e., +1.0), the mean estimated corrected correlation value was an over estimate for truncation values (i.e., −2.0, −1.5, −1.0, −0.5, 0, +0.5). The overestimations were either 0.01 or 0.02. For the severe cut, the mean estimated corrected correlation value was underestimated by 0.01.

Table 1

Computer Simulation Results from Alexander et al.: Comparison of Mean Observed and Corrected Estimated Correlations for Different ρ Values and Truncation Points

	ρ			
Truncation	.20	.40	.60	.80
–2.0	0.19 (0.21)	0.38 (0.42)	0.58 (0.62)	0.78 (0.81)
–1.5	0.18 (0.21)	0.36 (0.41)	0.55 (0.62)	0.76 (0.81)
–1.0	0.16 (0.21)	0.33 (0.41)	0.51 (0.61)	0.72 (0.81)
–0.5	0.14 (0.21)	0.29 (0.41)	0.46 (0.61)	0.68 (0.81)
μ	0.12 (0.21)	0.25 (0.42)	0.41 (0.61)	0.62 (0.81)
+0.5	0.10 (0.21)	0.22 (0.41)	0.36 (0.61)	0.56 (0.80)
+1.0	0.09 (0.19)	0.19 (0.39)	0.31 (0.59)	0.50 (0.79)

Note. Corrected correlation values using Alexander et al.'s formula are in parentheses. Data presented are mean correlations for samples with N = 60 for 5,000 replications. This table was taken from Alexander et al. (1984). Permission is found in Appendix D.

Alexander et al. (1984) also used the estimate unrestricted standard deviation, z-score, and the highest sorted value of x from the previous equation to derive a formula to correct mean. Alexander et al.'s corrected mean formula is as follows:

$$\overline{x_c} = x_c - zSD' \qquad 11$$

where $\overline{x_c}$ is the corrected mean,

X_c is the highest sorted x value in in a restricted sample,

z is the z-score from the Cohen's table, and

SD' is the estimated unrestricted standard deviation.

Thus, Alexander et al.'s (1984) formula is a special case of Thorndike's (1947) Formula 2. In Thorndike's formula, the restricted variance and unrestricted variance are

presented as population values, but in Alexander et al.'s version, the restricted variance is estimated from the sample and the unrestricted variance needs to be estimated from the sample variance. Consequently, Alexander et al. employed Cohen's (1959) ratio formula to obtain the estimate of the unrestricted variance. From the estimated unrestricted variance, Alexander et al. computed the U ratio, which was then entered into the Thorndike's (1947) Formula 2 to estimate the corrected correlation value.

Attenuation in Restriction of Range

Sackett and Yang (2000) stated that there are two common methods for correcting correlation values in a direct restriction of range and an indirect restriction of range. Hunter and Schmidt (2004) mentioned that majority of correlation values are not true correlation values because they are in fact lowered by error of measurement. Stauffer and Mendoza (2001) suggested a formula for correcting correlation for range restriction and unreliability. Additionally, Hunter, Schmidt, and Le (2006) proposed the Hunter-Schmidt corrected correlation with attenuation for the direct and indirect restriction of range. Further, Raju et al. (2006) suggested another method for corrected correlation with attenuation based on the reliabilities of x and y in a restricted sample. Before discussing these formulas, a brief overview of basic classical test theory is presented.

Reliability

A true score can be stated in terms of the observed score and its measurement error. For example, if a researcher observed a score on a test denoted by x, and there is an error of measurement associated with it called e_x (i.e., e_x is measurement error of variable x), then the observed score of x could be written as

$$x = t_x + e_x , \qquad 12$$

where t_x is the true score for variable x.

In classical test theory, the amount of error of measurement in the variable is measured by the number called the reliability of the variable, denoted for variable x as r_{xx}. The reliability of an observed test score consists of two components: the true score and some form of error measurement (Donald, Lucy, & Asghar, 1990). Given some assumptions, it has been shown that the variance of the observed scores (σ_x^2) is equal to the variance of the true scores (σ_t^2) plus the variance of their errors of measurement (σ_e^2) (Donald et al.). This can be expressed as

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2 \qquad 13$$

Reliability can be defined as a ratio of true score variance over the observed score variance. That is, reliability, r_{xx}, is equal to

$$r_{xx} = \sigma_t^2 / \sigma_x^2 \qquad 14$$

Reliability ranges from 0 to 1, where a reliability of 1 indicates no error. Using equation 13, formula 14 can be rewritten as

$$r_{xx} = 1 - \sigma_e^2/\sigma_x^2 \qquad 15$$

Hunter and Schmidt (1990) stated that reliability measures the percentage of the observed variance, which is the true score variance. For example, if the reliability of variable x is 0.8, it is implied that 80% of the variance in variable x is due to the true score variation and the remainder 20% of the variance in variable x belongs to measurement error.

Attenuation Formulas

Correction for attenuation is a statistical procedure, according to Spearman (1904), to “rid a correlation coefficient from the weakening effect of measurement error”

(Jensen, 1998). Given the fact that the correlation between variables of interest, say x and y, is diluted by measurement error, the correction for attenuation procedure, it has been argued, provides a more accurate estimate of the correlation between variables x and y by accounting for this effect.

Before presenting corrections for attenuation, the general form of the correction approach advocated by Hunter and Schmidt (2004) for a form of meta-analysis called “validity generalization” is presented. As stated earlier, since measurement error in correlation is associated with artifacts, those artifacts can be addressed in terms of correcting correlation with attenuation (Schmidt & Huy, 2006). The following corrected correlation could be expressed as

$$\rho' = a\rho \qquad 16$$

where ρ' is the corrected population correlation with attenuation,

a is the artifact, and

ρ is the population correlation before attenuation.

Furthermore, if there is a second artifact associated with the population correlation before attenuation, the above formula could be rewritten to incorporate the additional artifact as follows:

$$\rho' = a_1 a_2\rho \qquad 17$$

where ρ' is the corrected population correlation with attenuation,

a_1 is the artifact,

a_2 is the artifact, and

ρ is the population correlation before attenuation.

Thus, if there are n artifacts associated with the correct correlation with attenuation, then the formula would become

$$\rho' = a_1 a_{2\ldots} a_n \rho \qquad 18$$

where ρ' is the corrected population correlation with attenuation,

a_1 is the artifact,

a_2 is the artifact,

a_n is the artifact, and

ρ is the population correlation before attenuation.

Based on the correlation with attenuation, Schmidt, Le, and Illies (2006) proposed a formula to include measurement error in the estimate of the parameter in an unrestricted population from the parameter in a restricted population.

Hunter-Schmidt's Correlation with Attenuation Formulas

Hunter et al. (2006) discussed the error of estimate of an unrestricted population correlation through Thorndike's case two and case three. Thorndike's (1947) case two formula is widely used in the correction for the direct range of restriction of two variables of interest, x and y, where the restricted sample correlation is known between the two variables, and the correction factor or the sample standard deviation of the unrestricted and the restricted populations are known for one of the parameters. Similarly, there is also a range restriction with attenuation formula for Thorndike's (1947) case three formula, which is used for the indirect range of restriction. In Thorndike's case three formula, the restriction occurs on the third variable, z, which is correlated to both variables of interest, x and y.

Hunter et al. (2006) developed a range restriction with attenuation formula to address the measurement error of the direct range of restriction for Thorndike's (1947) case two formula:

$$\rho^{'} = \alpha \, \rho \qquad 19$$

where α is the attenuation coefficient for the correct correlation with attenuation, and it could be expressed as

$$\alpha = \frac{U_x}{1+(U_x^2-1)\rho^2} \qquad 20$$

where U_x is the U ratio and it was defined as the ratio of $SD_{restricted}/SD_{unrestricted}$, and ρ is defined as the product of $\rho_{tx,ty}$, and (SQRT ($r_{xx}r_{yy}$)) where r_{xx} and r_{yy} are the reliabilities of variables x and y (Hunter & Schmidt, 1999). Furthermore, Hunter et al. (2006) also discussed the range restriction with attenuation for the indirect range restriction.

Hunter and Schmidt's Formula on Range Restriction with Attenuation for Indirect Range Restriction

Hunter et al. (2006) proposed a formula to range restriction with attenuation for the indirect range restriction by giving detailed instructions on how to incorporate measurement error and reliability into the existing Thorndike's formula. The steps are as follows:

Step 1: Estimating the reliability of the independent variable x in the restricted population.

$$r_{xx\ restricted} = 1 - U_x^2\,(1 - r_{xx\ unrestricted}) \qquad 21$$

where U_x is the U ratio , and

$r_{xx\ unrestricted}$ is the reliability of the unrestricted x variable.

Step 2: Estimating the range restriction on U_T used of correcting factor U_x, and the reliability of independent variable x

$$U_T = \sqrt{U_x^2(1 - r_{xx\ unrestricted})/r_{xx\ unrestricted}} \qquad 22$$

where U_x is the U ratio , and

$r_{xx\ unrestricted}$ is the reliability of the unrestricted x variable,

Step 3: Correcting for measurement error before applying restriction of range correction

$$r_{correction} = r \,/\, \sqrt{r_{xx\ unrestricted} r_{yy\ unrestricted}} \qquad 23$$

where r is the correlation before attenuation.

Step 4: Applying Thorndike case 2 formulas using equations from Step 3 and Step 2

$$r_{\ corrected\ unrestricted} = \frac{U_T r_{correction}}{\sqrt{U_T^2 r_{correction}^2 - r_{correction}^2 + 1}} \qquad 24$$

Besides Hunter and Schmidt’s (2006) formulas for correct correlation formulas, Stauffer and Mendoza (2001) suggested a new method for the range restriction with attenuation.

Stauffer and Mendoza’s Formula

Stauffer and Mendoza (2001) proposed a formula for correcting correlation for range restriction and unreliability, which was based on Thorndike’s Formula 2 with the available estimates: (1) the unrestricted predictor reliability, (2) the incident range restricted criterion reliability, and (3) the restricted correlation. Stauffer and Mendoza’s (2001) correcting correlation for range restriction and unreliability is defined as follows:

$$r_{x,y\ unrest}^{*} = \frac{U_x r_{x,y\ rest}}{\sqrt{R_{xx}}\sqrt{U_x^2 r_{x,y\ rest}^2 - r_{x,y\ rest}^2 + r_{yy}}} \qquad 25$$

where U_x is the U ratio,

$r_{x,y\ rest}$ is the correlation of the restricted sample,

R_{xx} is the unrestricted predictor reliability, and

r_{yy} is the incident range restricted criterion reliability.

Thus, since correction formula and unreliability are common in real-world situations, Stauffer and Mendoza (2001) developed a rule of thumb for determining an order to handle the corrections by looking at the nature of the reliability estimate. Further, Raju et al. (2006) developed new approach for corrected correlation based on reliabilities.

Raju, Lezotte, Fearing, and Oshima's Formula

Raju et al. (2006) proposed a method to calculate the corrected correlation based on the reliability of x and y in a restricted sample:

$$r'_{xy} = k\, r_{xy} \,/\, \sqrt{r_{xx} r_{yy} - r_{xy}^2 + k^2 r_{xy}^2} \qquad 26$$

where r_{xx} is the reliability of independent variable x,

r_{yy} is the reliability of the dependent variable y,

r_{xy} is the correlation value of x and y in a restricted sample, and

k is defined as the ratio of unrestricted true scores standard deviation over the restricted true scores standard deviation.

In other words, k is defined as follows:

$$k = SD_{tx\ unrestricted} \,/\, SD_{tx\ restricted} \qquad 27$$

Raju et al. (2006) offer a procedure for estimating corrected correlation in range restriction for unreliability when an estimate of the reliability of a predictor is not available for the unrestricted sample. It has been long recognized that measurement error in the sample will restrict the observed magnitude of a Pearson product moment; thus,

since Thorndike's day, researchers have been correcting correlation based on measurement errors and restriction in range (Mendoza & Mumford, 1987). This topic has received considerable attention in recent years, with a new correction formula for attenuation from Hunter et al. (1990) and even more recently from Raju et al. (2006). These new correction formulas help ease the measurement error for corrected correlation in the restriction of range. Meta-analysis uses different type of corrections; thus, when applying the restriction of range to meta-analysis, the meta-analysis summary would result in a better understanding of what the correlations are in the unrestricted situations.

Meta-Analysis

Meta-analysis, a procedure for summarizing empirical studies, can be employed to summarize Pearson's correlation; it can be generally defined by the following steps: (a) defining a topic area and criteria for admissible studies, (b) locating relevant primary research, (c) coding study characteristic, (d) measuring study results on a common scale, and (e) aggregating the study results and relating them to study characteristics (Matheny, Aycock, Pugh, Curlette ,& Cannella, 1985). Additionally, meta-analysis can also be used as a guide to answer the question about what a researcher should be doing to incorporate one study with another study even if the first study employs different instruments across a different range of people (Hunter & Schmidt, 1990). Furthermore, Hunter and Schmidt (2004) stated that every method of meta-analysis is based on the theory of data, and a complete theory of data includes an understanding of sampling error, measurement error, and bias sampling in case of range restriction. The estimation of sampling error in meta-analysis could be expressed as the weight average in each correlation, which is weighted by number of scores in the study (Lipsey & Wilson, 2000); accordingly

$$\bar{r} = \frac{\sum N_i r_i}{\sum N_i} \qquad 28$$

where r_i is the correlation in study i, and

N_i is the number of persons in study i.

Hunter and Schmidt (2004) stated that the corresponding variance across studies is not the usual sample variance, but the frequency-weighted average squared error. The formula could be derived in term of the above equation as follows:

$$s_r^2 = \frac{\sum N_i (r_i - \bar{r})^2}{\sum N} \qquad 29$$

where r_i is the correlation in study i,

$\bar{r}$ is the weighted average,

N_i is number of persons in study i, and

N is total number of persons.

Furthermore, other correction formulas that focus on the measurement error and bias sampling could be addressed by the earlier work of Hunter and Schmidt correction formula with attenuation on range restriction for direct and indirect restriction of range in the previous section.

Contaminated Normal

Contrary to the general belief that correlation value in a restricted sample tends to be smaller than the correlation value in the unrestricted sample, Zimmerman and Williams (2000) state that the correlation in the restricted sample is sometimes larger than the correlation in the unrestricted sample. This happens in a class called the "contaminated normal," in which the assumption of normal distribution is violated in a particular way. In a contaminated normal situation, scores of outliers increase the magnitude of the correlation (Zimmerman & Williams, 2000).

Standard Error

Cochran (1977) suggested a mean square error (MSE) formula to compare a biased estimator with an unbiased estimator or the two biased estimators that could be presented as the expected value of the square of the difference between the estimated population correlation value and the true population correlation value:

$$MSE = E(\hat{\rho} - \rho)^2 \qquad 30$$

where $\hat{\rho}$ is the estimated population correlation value and ρ is the population correlation value. Additionally, when the estimated population correlation value approaches the mean population correlation value, the MSE could be expressed in terms of the mean of the population correlation values, such as

$$MSE = E[(\hat{\rho} - m) + (m - \rho)]^2 \qquad 31$$

$$MSE = E(\hat{\rho} - m)^2 + 2(m - \rho)E(\hat{\rho} - m) + (m - \rho)^2 \qquad 32$$

Since the expected value of the difference of the estimated population correlation value and the mean of population correlation equals to zero, the MSE could be simplified to

$$MSE = E(\hat{\rho} - m)^2 + (m - \rho)^2 \qquad 33$$

Thus, the MSE could be expressed in terms of variance of $\hat{\rho}$ and bias, where

$E(\hat{\rho} - m)^2$ is the variance of $\hat{\rho}$,

$(m - \rho)^2$ is the bias,

$\hat{\rho}$ is the estimated population correlation value,

ρ is the true population correlation value, and

m is the mean of population correlation.

Furthermore, the standard error can be derived from the MSE as the square root (SQRT) of the MSE, where the standard error equals to the SQRT (MSE).

Summary

Restriction of range is a frequent topic in education and other areas where there is a selection process and criterion-related validity using Pearson's correlation is desired. Thorndike's (1947) and Alexander et al.'s (1984) formulas use the original Pearson's correlation formula to develop their corrected correlation formulas for restriction of range. As stated earlier, Thorndike's formula uses known variance for the unrestricted situation to obtain an estimate of the corrected correlation, whereas Alexander et al.'s formula with unknown unrestricted variance uses the Cohen ratio formula to obtain the population variance, and then it uses the population variance in Thorndike's Formula 2 in order to obtain the estimated corrected correlation. Hunter and Schmidt (2004) stated that the majority of correlation values were not true correlation values because they were in fact lowered by error of measurement. Thus, because of unreliability, since Thorndike's day, researchers have been correcting correlation based on measurement errors and restriction in range (Mendoza, Hart, & Powell, 1994). Furthermore, correcting correlation based on measurement errors has received considerable attention in recent years with new correction formulas for attenuation from Hunter and Schmidt (1990), Mendoza et al. (1994), and most recently from Raju et al. (2006). Those new correction formulas help ease the measurement error for corrected correlation in the restriction of range. For example, works on corrected correlation in restriction of range with attenuation have been shown through Hunter and Schmidt's formulas for direct and indirect restriction of range with attenuation, Stauffer and Mendoza's formula, and Raju et al.'s formula for

corrected correlation with attenuation. Also, meta-analysis uses different types of corrections; thus, applying the restriction of range to meta-analysis would result in a better understanding of what correlations are in unrestricted situations. This literature review included restriction of range formulas, restriction of range with attenuation formulas, meta-analysis, and contaminated normal as a background of the following study, which is presented in manuscript form for this dissertation.

References

Alexander, R. A., Alliger, G. M., & Hanges, P. J. (1990). Correcting for range restriction when the population variance is unknown. *Applied Psychological Measurement*, *8*, 431–437.

Alexander, R. A., Hanges, P. J., & Alliger, G. M. (1984). Correcting for restriction of range in both x and y when the unrestricted variances are unknown. *Applied Psychological Measurement*, *8*(4), 431–437.

Cochran, G. W. (1977). *Sampling techniques*. New York: John Wiley & Sons.

Donald, A., Lucy, C. J., & Asghar, R., (1990). *Introduction to research in education*. Forth Worth, Texas: Harcourt Brace College Publisher.

Goodwin, D. L., & Leech. L. N. (2006). Understanding correlation: Factors that affect the size of r. *The Journal of Experimental Education*, *74*(3), 251–266.

Hunter, J. E., Schmidt, F. L., Le, H. (2006). Implication of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, *91*(3), 594–612.

Hunter, J. E., & Schmidt. L. F. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage Publications.

Hunter, J. E., & Schmidt, L. F. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Thousand Oaks, CA: Sage Publications.

Jensen, A.R. (1998). *The g factor: The science of mental ability*. Mahwah, NJ: Lawrence Erlbaum & Associates.

Lipsey, M. W., & Wilson, D. B. (2000). *Practical meta-analysis*. Applied social research methods series, 34. Thousand Oaks, CA: Sage Publications.

Matheny, K. B., Aycock, D. W., Pugh, J. L., Curlette, W. L., & Cannella, K. A. S. (1986). Stress coping: Qualitative and quantitative synthesis with implications for treatment. *The Counseling Psychologist*, *14*(4), 499–549.

Mendoza, J. L, Hart, D. E., & Powell, A. (1994). Explanations for accuracy of the general multivariate formulas in correcting for range restriction. *Applied Psychological Measurement*, *18*, 355–367.

Mendoza, J. L., & Mumford, M. (1987). Corrections for attenuation and range restriction on the predictor. *Journal of Educational Statistics*, *12*(3), 282–293.

Oleksandr, S. C., & Deniz S. O. (1999). How selective are psychology graduate programs? The effect of the selection ratio on GRE scores validity. *Educational and Psychological Measurement*, *59*, 951–961.

Pearson, K. (1903). Mathematical contribution to the theory of evolution-XI. On the influence of natural selection on the variability and correlation of organ. *Phil. Trans. Royal Soc. of London, Series A*, 200, pp 1–66.

Raju, N. S., Lezotte, D. V., Fearing, B. K., & Oshima, T. C. (2006). A note on correlations corrected for unreliability and range restriction. *Applied Psychology Measurement*, *30*(2), 145–149.

Rodgers, J. L., & Nicewander, W. L. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*, 59–66.

Rovine, M. J., & Von, E. A. (1997). A 14th way to look at a correlation coefficient: Correlation as the proportion of matches. *The American Statistician*, *51*, 42–46.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded topology. *Journal of Applied Psychology*, *85*(1), 112–118.

Schmidt, F. L., & Huy, L. (2006). Increasing the accuracy of corrections for range restriction: Implication for selection procedure validities and other research results. *Personal Psychology*, *59*, 281–305.

Schmidt, F. L., Le, H., & Illies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods*, *8*, 206–224.

Stauffer, M. J., & Mendoza, L. J. (2001). The proper sequence for correcting correlation coefficients for range restriction and unreliability. *Psychometrika*, *66*, 63–68.

Thorndike, R. L. (1947). *Research problems and techniques (Rep. No. 3 AAF Aviation Psychology Program Research Reports)*. Washington, DC: U.S. Government Printing Office.

Zimmerman, D. W., & Williams, R. H. (2000). Restriction of range and correlation in outlier-prone distribution. *Applied Psychological Measurement*, *24*(2), 267–280.

CHAPTER 2

A COMPARISON OF ROBUSTNESS OF THORNDIKE'S AND AN ADAPTATION OF COHEN'S FORMULA TO CORRECT FOR RESTRICTION OF RANGE

The focus of this research was to compare Alexander et al.'s formula to Thorndike's formula and to population values using Monte Carlo simulation when the assumption of a normal distribution is violated in a particular way. Many research studies involving Pearson's correlations are conducted in settings where one of the two variables has a restricted range in a sample. For example, this occurs when tests are used for selecting people for employment or admission to school. Often after selection, there is interest in correlating the selection variable, which has a restricted range, to a criterion variable in order to obtain criterion-related validity using Pearson's correlation. Since the restriction of range situation occurs in many settings and Pearson's correlation is fundamental to many statistical procedures, the accuracy of correction approaches for restriction of range can relate to many statistical procedures. A particular statistical procedure for which these correction procedures have potential application is validity generalization, which is a form of meta-analysis.

A unique feature of the current study is the way in which the robustness situations are defined. They are defined from the perspective of a statistical classification problem for two groups and one variable on which to make the classification decision. Group 1 has a normal distribution on variable X and is referred to as distribution 1. Likewise, group 2 has been measured on variable X and has a normal distribution, designated as distribution 2. Both distributions have the same population standard deviations but different population means.

In this classification problem, after the classification decision is made, there are two potential errors of classification: (1) an individual from group 1 is misclassified into group 2 and (2) an individual from group 2 is misclassified into group 1. There could be a desire to identify individuals who are predicted to be in group 1 when the classification rule is applied to a new group of people. Suppose, for example, that the task was to predict people leaving a job based on the supervisor's rating after two years of employment. The supervisor's ratings of employees would be obtained after one year of employment, and the employees' work statuses in the organization would be obtained at the end of the second year. Then, a classification rule would be obtained to predict which employees would stay and which employees would leave. Given two normal distributions and one predictor variable, the cut point for the classification rule on rating variable, X, would be halfway between the mean of the two distributions without consideration of prior probabilities. Now, a new group of employees comes along, and it is desired to study the correlations between those predicted to stay with another variable of interest, for example, Y. Those predicted to stay make a mixed distribution of group 1 and group 2. In particular, those individuals from distribution 1 below the cut point constitute group 1, along with those individuals from group 2 who have been misclassified because they are below the cut point. An illustration of the classifying problem is shown by Figure 1, Figure 2, and Figure 3 presented in Appendix F. In this fashion, the robustness situations for the restriction of range situations are defined for the Monte Carlo simulation.

Two other contributions of this research are to (1) compare the accuracy of Thorndike's (1947) case 2 formula to Alexander et al.'s (1984) formula based on point estimates of correlations and (2) to compare the standard errors of these two approaches.

In his article, Alexander et al. (1984) compares estimates of correlations based on Cohen's approach to population correlation values and not to estimates from Thorndike's formula. After a brief review of the literature in the background section, I identify the exogenous variables to define the simulated situations and provide a more detailed description of the robustness situations that will be described in the methodology section.

Background

Thorndike's Direct Restriction of Range Formula

Thorndike (1947) introduced the corrections for restriction of range. Thorndike's Formula 2 estimates the population correlation between two variables of interest, X and Y, when the restriction of range has occurred on variable X. Thorndike's Formula 2 is shown as follows:

$$r_u = \left((r_{rest}) \left(\frac{SD_u}{SD_{rest}} \right) \right) / \, \mathrm{sqrt} \left(1 - r_{rest}^2 + r_{rest}^2 \left(\frac{SD_u^2}{SD_{rest}^2} \right) \right) \qquad 30$$

where r_u is the corrected correlation for the unrestricted restriction of range,

r_{rest} is the restricted range correlation,

SD_u is the unrestricted standard deviation, and

SD_{rest} is the restricted standard deviation.

The ratio of SD_u over SD_{rest} is defined as the U ratio or the correcting factor. The U ratio for Thorndike's formula is defined as the ratio of the unrestricted standard deviation over the restricted standard deviation. Thorndike's Formula 2 is designed for situations where the required information about restricted standard deviation to unrestricted standard deviation is on the same variable that had restricted range. In addition, Thorndike's (1947) correction formula is well known for utilizing known

variance for the unrestricted situation to obtain an estimate of the corrected correlations. Oleksandr and Deniz (1999) provides a case when Thorndike's (1947) Formula 2 was used in Graduate Record Examination (GRE) validation studies in which the researcher was interested in estimating the corrected correlation from an unrestricted sample from the correlation of the restricted sample. Additionally, Thorndike's (1947) case two formula is widely employed in personnel selection for employment, when the test scores used for new applicants are related to job performances (Viswesvaran, Ones, & Schmidt, 1996; Schmidt & Hunter, 1998). In this study, selection was made on the scores; thus, the range of scores was restricted in the sample. Even though the correlation between test score and job performance could have been obtained from the restricted sample, the researcher still wanted to know the correlation in the unrestricted sample (Henriksson & Wolming, 1998; Hunter & Schmidt, 1990). Hence, Thorndike's (1947) Formula 2 has been used in many instances in real-world situations ranging from educational research to employment, and it has been shown to produce close estimates of the correlation in the population (Oleksandr & Deniz, 1999).

The Cohen Ratio

Cohen (1959) based his formula on the sample variance in the restricted sample over the square of difference between the sample mean in the restricted sample and the point of truncation in the restricted sample. The formula for the Cohen ratio can be presented as

$$\text{Cohen ratio} = \frac{\mathrm{SD}^2_{\mathrm{rest}}}{(\mathrm{X}' - \mathrm{X}_{\mathrm{c}})^2} \qquad 31$$

where $\mathrm{SD}^2_{\mathrm{rest}}$ is the variance of the restricted sample,

X' is the mean of the restricted sample , and

X_c is the point of truncation in the restricted sample.

Alexander et al. (1984) points out that Cohen's ratio has an advantage over others' formulas because it needs only the variance of the restricted sample, the mean of the restricted sample, and the highest observed X_c in the restricted sample. The value X_c can be obtained from an ascending-ordered array where all X scores could be stored in an unordered array, and the unordered array could be sorted by an efficient sorting algorithm named quick sort (Horowitz & Sahni, 2000). Thus, once the Cohen ratio is determined, the z-score and table restricted normal standard deviation, named SD_{tab}, could be obtained from the Cohen table, which has three columns (see Appendix B). The first column is the Cohen ratio, the next column is the SD_{tab}, and the last column is the z-score. The SD_{tab}, and z-score would be found from the Cohen table by performing a table look-up of the Cohen ratio.

Cohen's Ratio in Alexander et al.'s Formula

Thorndike's (1947) case two formula assumes that the variance of variables in the unrestricted area is known, whereas in Alexander et al.'s formula, the unrestricted variance is unknown. With the unknown variance in the unrestricted sample, Alexander suggested a method for estimating the unrestricted standard deviation using the Cohen ratio. Alexander et al. (1984) used the Cohen ratio in the previous equation to obtain the z-score and the SD_{tab} (as seen in the Cohen table in Appendix B). Once the SD_{tab} value was obtained, Alexander et al.'s (1984) formula was then employed to estimate the unrestricted standard deviation as follows:

$$SD' = SD_{obs} \,/\, SD_{tab} \qquad 32$$

where SD' is the unrestricted standard deviation,

SD_{obs} is the observed standard deviation in the restricted sample, and

SD_{tab} is the tabled restricted normal standard deviation in restricted sample.

Once the estimates of the unrestricted standard deviation, the observed standard deviation, and the restricted correlation are available, the estimate corrected correlation could be calculated using Thorndike's Formula 2. An example of the Monte Carlo computer simulation program of the comparison of the mean observed and corrected correlations for different population's correlation and truncation values is given in Table 2. Even though the unrestricted variance was estimated through Alexander et al.'s (1984) formula, in every case except for the truncation value at −2.0, the results of the estimated corrected correlation values were closer to the true ρ correlation values than the observed correlation values. Likewise, in every case except for the severe cut (i.e., +1.0), the estimated corrected correlation value was an overestimate for truncation values (i.e., –2.0, –1.5, –1.0, –0.5, 0, +0.5). The overestimations were either 0.01 or 0.02. Thus, the results show that Alexander et al.'s formula is a worthwhile estimate of corrected correlation in a real-world situation where more than often, the variance of the unrestricted sample is not given and the researcher could estimate it using the Cohen ratio.

Table 2

Computer Simulation Results: Comparison of Mean Observed and Corrected Correlations for ρ and Truncation Points

	ρ			
Truncation	.20	.40	.60	.80
–2.0	0.19 (0.21)	0.38 (0.42)	0.58 (0.62)	0.78 (0.81)
–1.5	0.18 (0.21)	0.36 (0.41)	0.55 (0.62)	0.76 (0.81)
–1.0	0.16 (0.21)	0.33 (0.41)	0.51 (0.61)	0.72 (0.81)
–0.5	0.14 (0.21)	0.29 (0.41)	0.46 (0.61)	0.68 (0.81)
µ	0.12 (0.21)	0.25 (0.42)	0.41 (0.61)	0.62 (0.81)
+0.5	0.10 (0.21)	0.22 (0.41)	0.36 (0.61)	0.56 (0.80)
+1.0	0.09 (0.19)	0.19 (0.39)	0.31 (0.59)	0.50 (0.79)

Note. Corrected values are in parentheses. Data based on N = 60 with 5,000 replications. This table was taken from Alexander et al. (1984). Permission is found in Appendix D.

Thorndike's and Alexander et al.'s Variances in the Unrestricted and Restricted Situations

In both Thorndike's and Alexander et al.'s correction formulas, values for the variances in the restricted and the unrestricted situations are required. For both formulas, the variance in restricted situations was a sample estimate. In the Monte Carlo simulation, the difference between the two approaches was that in Thorndike's formula, the variance in the unrestricted situation was the population variance from the exogenous, whereas in Alexander et al.'s approach, the population variance was estimated based on the sample variance in the restricted situation. The particular method that was used to create the robustness situations was the non-normal distributions for predicted group membership in a classification problem. Shown in Appendix C is a table that shows the sources for the variances in the restricted and unrestricted situations for Thorndike's and Alexander et al.'s correction formulas.

Non-Normal Situations as a Result of a Classification Problem

The problem of classifying people into groups is formed on the basis of a set of measurements, such as the aptitude tests and personal inventory scores, which often arises in applied research psychology or social science (Johnson & Wichern, 1998; Kuncel, Hezlett, & Ones, 2004). The particular classification problem employed to define non-normal distribution for robustness situations is a two-group classification situation with one predictor variable. Group 1 has a normal distribution on variable X and is referred to as distribution 1. Likewise, group 2 has been measured on variable X also and has a normal distribution that will be designated as distribution 2. Both distributions have the same population standard deviations but different population means. In this classification problem, after the classification decision is made, there are two errors of classification; more particularly, an individual from group 1 is misclassified into group 2, and an individual from group 2 is misclassified into group 1.Suppose it is desired to study the individuals predicted to be in group 1, that is, those individuals below a cut point on the predictor variable. These individuals are a mixture of group 1 and group 2, which has a non-normal distribution.

Standard Error

Standard error is the square root of the mean square error, which can be defined as the expected value of the square of the difference between the estimated population value and the true population value. Cochran (1977) presented a mean square error formula (i.e., MSE) that can be expressed as the expected value of the square of the difference between the estimated population correlation value and the true population correlation value as:

$$\text{MSE} = E(\hat{\rho} - \rho)^2 \qquad 33$$

The MSE could be rewritten in terms of the mean of population correlation value m as:

$$\text{MSE} = E(\hat{\rho} - m)^2 + (m - \rho)^2 \qquad 34$$

where $E(\hat{\rho} - m)^2$ is the variance of $\hat{\rho}$,

$(m - \rho)^2$ is the bias,

$\hat{\rho}$ is the estimated population correlation value,

ρ is the true population correlation value, and

m is the mean of population correlation.

Thus, the standard error estimate could be derived from the square root of the mean square error such as

$$\text{Standard error} = \sqrt{\text{MSE}} \qquad 35$$

Degree of Closeness of Two Correlations

Cohen (1987) suggested using Fisher's z transformation in the computing of the degree of closeness of the two correlations by taking the absolute value of the difference of the two z-values that were derived from one-half of the natural log of the quotient of (1 + r) and (1 - r); the result was then compared with Cohen's effect size to find out the magnitude of the closeness of the correlations:

$$d = |z_1 - z_2| \qquad 36$$

$$z = 0.5 \ln\left(\frac{1+r}{1-r}\right) \qquad 37$$

where z is the Fisher's z transformation,

and ln is the natural log.

Ultimately, the practical significance of the differences of two correlation corrections compared to population values depends on how the differences affect conclusions drawn from applied research.

Skewness

Skew can be defined as an expected value of the averaged cubed deviation from the mean divided by the standard deviation cubed as $E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right]$, where μ is the population mean and σ is the population standard deviation (Groeneveld & Meeden, 1984). If the skew value is greater than zero, then there is a positive skew, whereas negative skew occurs when the result is less than zero. Additionally, it is symmetric, or "no skew," when the result is zero.

Research Questions

The significance of the current study is to contribute an additional understanding of the corrections for the effect of restriction of range on correlation. Because Thorndike's Formula 2 uses known population unrestricted variance and Alexander et al.'s formula has to estimate the unrestricted variance by utilizing the Cohen (1959) ratio formula, I expected that Thorndike's Formula 2 would provide better precision than Alexander et al.'s formula. I investigated four research questions related to the restriction of range:

1. How accurate in terms of point estimates of correlations are Thorndike's (1947) and Alexander et al.'s (1984) formulas to correct for restriction of range for an original normal distribution that has been truncated?

2. How do the standard errors of Thorndike's corrected correlations compare to Alexander et al.'s corrected correlations for an original normal distribution that has been truncated?
3. In the robustness situations, how accurate in terms of point estimates of correlations are Thorndike's (1947) and Alexander et al.'s (1984) formulas to correct for restriction of range?
4. In the robustness situations, how do the standard errors of Thorndike's (1947) corrected correlations compare to the standard errors of Alexander et al.'s (1984) corrected correlations?

Methodology

The robustness situations for restriction of range situations are defined for the Monte Carlo simulation from the perspective of a statistical classification problem for two groups and one variable on which to make the classification decision. Group 1 has a normal distribution on variable X and is referred to as distribution 1. Likewise, group 2 has also been measured on variable X and has a normal distribution, designated as distribution 2. Both distributions have the same population standard deviations of 1.0 but different population means. Group 1 has a population mean of 10.0, and group 2 has the following population means: 11.0, 12.0, and 14.0. The cut points for classifying an observation into group 1 and group 2 are the average of the population mean of group 1 and the population mean of group 2. The values of the cut points are 10.5, 11.0, and 12.0. The corresponding z-score for the cut points are 0.5, 1.0, and 2.0, referenced from distribution 1. From the viewpoint of sample size, two classifications are considered, one

in which group 1 and group 2 have a sample size of 60, and the other in which the sample size for both groups is 120.

For the Monte Carlo simulation, the above classification problem forms the basis for defining areas 1 and 2. Area 1 is defined as the truncated normal distribution for group 1 only below the cut point. Area 2 is defined as members of group 2 below the cut point or those individuals predicted to be in group 1. Figure 1, Figure 2, and Figure 3 in Appendix F illustrate a graphical presentation of both areas 1 and 2 and the combined areas 1 and 2.

Table 3 and Table 4 represent the expected number of pairs of scores in area 1 after the truncation and the expected number of pairs of scores in areas 1 and 2 after truncation. When comparing Thorndike's and Alexander et al.'s approaches on estimating the corrected correlations in the restriction of range, I measured the accuracy in terms of point estimates of correlations of the two approaches for restriction of range in area 1. Similarly, I compared the point of estimates of correlation between Thorndike's formula and Alexander et al.'s formula for restriction of range in the robustness situations defined by combining areas 1 and 2. When comparing the point of estimates of correlations of the two approaches, I investigated different error terms based on the two approaches. These error terms are the differences between the observed correlation, corrected correlation, and population correlation. Defining an error term in this way facilitates summarizing error terms across situations with different populations. Eight error terms have been identified: (a) error 1 for the Thorndike's (1947) observed correlation value in area 1, (b) error 2 for the Thorndike's corrected correlation value in area 1 with a known variance, (c) error 3 for Alexander et al.'s (1984) observed

correlation value for area 1, (d) error 4 for Alexander et al.'s corrected correlation value for area 1 with an unknown variance, (e) error 5 for Thorndike's observed correlation value for area 1 and area 2, (f) error 6 for Thorndike's corrected correlation value for areas 1 and 2 with a known variance , (g) error 7 for Alexander et al.'s observed correlation value in area 1 and area 2, and (h) error 8 for Alexander et al.'s corrected correlation value with an unknown variance in areas 1 and 2.

Table 3

The Expected Number of Elements in Area 1 after Truncation

Mean X	Mean Y	Truncation	Z Score	Area 1	N	No. of Elements Expected in Area 1
10	11	10.5	0.5	0.69	60	41.4
10	12	11.0	1.0	0.84	60	50.4
10	14	12.0	2.0	0.975	60	58.5
10	11	10.5	0.5	0.69	120	82.8
10	12	11.0	1.0	0.84	120	100.8
10	14	12.0	2.0	0.975	120	117.0

Note. N is the number of elements before truncation. No. of elements expected in Area 1 is the number of pairs of scores (x,y) in sample distribution 1 after truncation.

Table 4

The Areas Defining the Truncation and Robustness Situations for Areas 1 and 2

Mean X	Mean Y	Truncation	Z-Score	Area 1	Area 2	N	No. of Elements Expected in Area 1	No. of Elements Expected in Area 2
10	11	10.5	0.5	0.69	0.31	60	41.4	18.6
10	12	11.0	1.0	0.84	0.16	60	50.4	9.6
10	14	12.0	2.0	0.975	0.025	60	58.5	1.5
10	11	10.5	0.5	0.69	0.31	120	82.8	37.2
10	12	11.0	1.0	0.84	0.16	120	100.8	19.2
10	14	12.0	2.0	0.975	0.025	120	117.0	3.0

Note. N is the number of elements before truncation. No. of elements expected in Area 1 is the number of pairs of scores (x,y) in sample distribution 1 after truncation. No. of elements expected in Area 2 is the number of pairs of scores (x,y) in sample distribution 2 after truncation.

I used a Monte Carlo simulation to investigate the mean of estimated error terms for each of the error terms based on the truncation point defined by the average of the population means of distribution one and distribution two for area 1 and for the combined areas 1 and 2.

Research Design

The four research questions previously presented are now stated in terms of area 1 and area 2.

1. How accurate in terms of point estimates of correlation are Thorndike's (1947) and Alexander et al.'s (1984) formulas to correct for restriction of range in area 1?
2. How do the standard errors of Thorndike's corrected correlations compare to those of Alexander et al.'s corrected correlations for area 1?
3. In the robustness situations that are defined by combining areas 1 and 2, how accurate in terms of point estimates of correlations are Thorndike's (1947) and Alexander et al.'s (1984) formulas to correct for restriction of range?
4. In the robustness situations that are defined by combining areas 1 and 2, how do the standard errors of Thorndike's (1947) corrected correlations compare to those of Alexander et al.'s (1984) corrected correlations for area 1 and area 2 combined?

The research design uses a set of exogenous variables shown in Table 5 that were inputs into the Monte Carlo simulation program. Thorndike's formula uses known population variance to estimate the corrected correlation value, whereas Alexander et al.'s formula estimates the unrestricted variance from Cohen's ratio and then uses it to

calculate the corrected correlation. The particular contributions of exogenous variables used in the simulation are presented in Table 6.

Table 5

Exogenous Variables Used to Define a Situation in Monte Carlo Study

Exogenous Variables	Values
Number of Populations	2
Number of Variables	2 denoted as (X, Y)
Sample Size	60, 120
Population Mean of Distribution 1	10.0
Population Mean of Distribution 2	11.0, 12.0, 14.0
Degree of Truncation (compute by using the average of the population means of distribution 1 and distribution 2)	10.5, 11.0, and 12.0
True Underlying Population	Normal distribution
Degree of Correlation in Population	$\rho = 0.20$, $\rho = 0.40$, $\rho = 0.60$
Number of Replications	2000

Table 6

Simulated Situations Used to Produce Estimated Correlations from Thorndike's and Alexander et al.'s Approaches

Situation	Sample Size	ρ	Truncation	Areas
1	60	0.2	10.5	1, 1 & 2
2	60	0.2	11.0	1, 1 & 2
3	60	0.2	12.0	1, 1 & 2
4	60	0.4	10.5	1, 1 & 2
5	60	0.4	11.0	1, 1 & 2
6	60	0.4	12.0	1, 1 & 2
7	60	0.6	10.5	1, 1 & 2
8	60	0.6	11.0	1, 1 & 2
9	60	0.6	12.0	1, 1 & 2
10	120	0.2	10.5	1, 1 & 2
11	120	0.2	11.0	1, 1 & 2
12	120	0.2	12.0	1, 1 & 2
13	120	0.4	10.5	1, 1 & 2
14	120	0.4	11.0	1, 1 & 2
15	120	0.4	12.0	1, 1 & 2
16	120	0.6	10.5	1, 1 & 2
17	120	0.6	11.0	1, 1 & 2
18	120	0.6	12.0	1, 1 & 2

Situations 1 through 9 account for both non-robustness and robustness situations for the sample size of 60; truncation values of 10.5, 11.0, and 12.0; and population correlation values of 0.2, 0.4, and 0.6 in area 1, and in both areas 1 and 2. Likewise, situations 10 through 18 offer non-robust and robust situations for a different sample size ($N = 120$). The results for situations 1 through 18 are shown in Appendix A in terms of corrected correlations and standard errors.

Endogenous Variables in Monte Carlo Computer Simulation

The fundamental outcome variables, or endogenous variables, were the eight error estimate terms defined previously in the methodology section and shown in Table 7.

Table 7

Error Terms for Estimating Corrected Correlation Values for Thorndike's and Alexander et al.'s Formulas

Error Terms	Formula
Error estimate for Thorndike observed correlation for area 1	$\rho'_{xy,obs,T1} - \rho_{xy}$
Error estimate for Thorndike corrected correlation for area 1	$\rho'_{xy,est,T1} - \rho_{xy}$
Error estimate for Alexander et al. observed correlation for area 1	$\rho'_{xy,obs,A1} - \rho_{xy}$
Error estimate for Alexander et al. corrected correlation for area 1	$\rho'_{xy,est,A1} - \rho_{xy}$
Error estimate for Thorndike observed correlation for area 1 and area 2	$\rho'_{xy,obs,T12} - \rho_{xy}$
Error estimate for Thorndike corrected correlation for area 1 and area 2	$\rho'_{xy,est,T12} - \rho_{xy}$
Error estimate for Alexander et al. observed correlation for area 1 and area 2	$\rho'_{xy,obs,A12} - \rho_{xy}$
Error estimate for Alexander et al. corrected correlation for area 1 and area 2	$\rho'_{xy,est,A12} - \rho_{xy}$

Note. Subscript obs is for observed, est is for estimate. T1 is for Thorndike for area 1, T12 is for Thorndike for area 1 and area 2. A1 is for Alexander et al. for area 1, A12 is for Alexander et al. for area 1 and area 2. ρ' is the estimated correlation. ρ is the true correlation value.

In each replication of a simulated situation, one error term is calculated. The fundamental outcome variables for the simulated situations are shown in Table 7, and the square root of the variance of these errors constitutes the standard error.

Planned Analysis for Research Questions

For the mean point estimate of correlation for area 1 with degree of truncation values of 10.5, 11.0, and 12.0, with sample size before truncation of 60 and 120, and the number of repetitions at 2000, the simulation results are summarized across these situations for the appropriate areas. The summary includes the mean observed correlation

values after truncation, the mean of corrected correlation values, the mean of standard errors for observed correlations values, the mean of error estimates for the observed correlation values, and the mean of error estimates for the corrected correlation values. The corrected correlations are calculated using Thorndike's (1947) formula and Alexander et al.'s (1984) formula. To address the research questions, each of the different error estimates, defined previously as the fundamental outcome variables, are compared. The summary will include the standard errors of the two approaches.

Simulation Program

I used pseudo code to illustrate the Monte Carlo simulation. Pseudo code is a high-level language description of a programming task. The primary method for this effort is defined by the Monte Carlo simulation program. The program was written in R language, which is a programming language for statistical analysis, and it is available as an open-source software program (Crawley, 2007). I used R language to analyze the quantitative data and to make a prediction of the estimated corrected correlation values in the restriction of range using the given simulation's quantitative data. The R programming language library includes a random number generator function to generate data (Crawley, 2005). The random number generator takes the following input parameters for each sample distribution: the sample size, the mean of sample distribution, the sample standard deviation, and the ρ value. It gives the vector array of ordered pairs (x,y) as the output. I wrote the Monte Carlo computer program that generates the number of inputs for the model. The simulation program can perform the following tasks: (1) read the input data file, which contains the exogenous variables described in Table 5; (2) read random user's inputs for the truncation, sample size, population correlation values,

number of replications, population mean of distribution 1, population mean of distribution 2, and population standard deviation of distributions 1 and 2; (3) compute the results; and (4) write the results to a file for further analysis. A snippet of the R program on estimating the unrestricted standard deviation for Alexander et al.'s formula is found in Appendix G.

The Monte Carlo simulation program takes the exogenous variables in Table 5 as the input parameters to the program. The input parameters are as follows:

a. Sample size, defined as R language variables, containing the population sample size parameters (i.e., 60, 120) as an integer value.
b. Variables, defined as R language input variables, containing 4 elements denoted as "Mean X," "Standard deviation X," "Mean Y," and "Standard deviation Y," where Mean X contains the population mean (i.e., 10.0), Mean Y contains the population mean (i.e., 11.0, 12.0, 14.0), and the population standard deviation for distribution 1 and distribution 2 (i.e., 1).
c. Correlations, defined as R language variables, containing 3 elements, which are the constant values 0.2, 0.4, 0.6.
d. Number of repetitions, defined as R language variable, which is the constant value 2000.
e. Underlying population, defined as R language variables, containing 2 elements, which are represented as "normal-normal" distributions for distribution 1 and distribution 2.

f. Truncation location, defined as R language variable, containing the value for the cut point, which is defined as the average of the population mean of distribution 1 and the population mean of distribution 2.

Parameters defined in Table 5 are used as the input parameters to the simulation program with the particular simulated situations defined in Table 6. The simulation program produces the endogenous variables or output as shown for each situation in Appendix A.

Expected Findings

Based on previous research, I expected that differences in the approaches of Thorndike (1947) and Alexander et al. (1984) to restriction of range would be found. Given the facts that Thorndike's (1947) estimate is based on a known unrestricted variance and Alexander et al.'s (1984) formula uses Cohen's (1959) ratio for the estimate of the unrestricted variance, I expected that the results of Thorndike's formula would be closer to the population correlation values than the results of Alexander et al.'s formula for normal situations. For the non-normal situations, there was no expectation about the magnitude of the errors in predicting the correlations for either procedure. My results should help inform the researcher about the magnitudes of the correlations to be expected in the robustness situations for both approaches.

Results

Results from the computer simulation for situations 1 through 18 (see Appendix A) are shown in Tables 8, 9, 10, and 11. Table 8 gives the mean point estimates of correlations for N = 60. Table 9 gives the mean differences for point estimates for N = 60. Likewise, Tables 10 and 11 give similar values for N = 120. Tables 8, 9, 10, and 11

provide the information that is summarized in Tables 12 and 13 to address the first research question.

Table 8

Comparison of the Mean Observed and Corrected Correlation Values for Thorndike's Approach and Alexander et al.'s Approach in Area 1 for N = 60

Truncation values	ρ (N = 60, number of repetitions = 2000) 0.2		0.4		0.6	
	Thorn	Alex	Thorn	Alex	Thorn	Alex
10.5	0.134	0.134	0.284	0.284	0.456	0.456
	(0.188)	(0.206)	(0.386)	(0.413)	(0.590)	(0.615)
11.0	0.155	0.155	0.323	0.323	0.507	0.507
	(0.194)	(0.207)	(0.395)	(0.415)	(0.597)	(0.617)
12.0	0.189	0.189	0.381	0.381	0.577	0.577
	(0.202)	(0.210)	(0.404)	(0.418)	(0.603)	(0.619)

Note. Thorn is for Thorndike. Alex is for Alexander et al. N refers to sample size in distribution 1 before truncation. Corrected correlation values shown in parentheses; observed correlation values not in parentheses.

Table 9

Mean Differences of the Observed and Corrected Correlation Values for Thorndike's Approach and Alexander et al. 's Approach for Area 1 for N = 60

Truncation values	ρ (N = 60, number of repetitions = 2000) 0.2		0.4		0.6	
	Thorndike	Alexander	Thorndike	Alexander	Thorndike	Alexander
10.5	0.066	0.066	0.116	0.116	0.144	0.144
	(0.012)	(–0.006)	(0.014)	(–0.013)	(0.010)	(–0.015)
11.0	0.045	0.045	0.077	0.077	0.093	0.093
	(0.006)	(–0.007)	(0.005)	(–0.015)	(0.003)	(–0.017)
12.0	0.011	0.011	0.019	0.019	0.023	0.023
	(–0.002)	(–0.010)	(–0.004)	(–0.018)	(–0.003)	(–0.019)

Note. N refers to sample size in distribution 1 before truncation. Thorndike and Alexander et al. corrected correlation values error estimated values shown in parentheses; observed error estimated values of correlations from truncation situation not in parentheses.

Table 10

Comparison of the Mean Observed and Corrected Correlation Values for Thorndike and Alexander et al. in Area 1 Across 2000 Replications for N = 120

Truncation values	ρ (N = 120, number of repetitions = 2000) 0.2		0.4		0.6	
	Thorndike	Alex	Thorndike	Alex	Thorndike	Alex
10.5	0.138	0.138	0.287	0.287	0.459	0.459
	(0.195)	(0.202)	(0.395)	(0.406)	(0.596)	(0.607)
11.0	0.159	0.159	0.326	0.326	0.510	0.510
	(0.199)	(0.204)	(0.399)	(0.408)	(0.600)	(0.609)
12.0	0.189	0.189	0.380	0.380	0.577	0.577
	(0.201)	(0.205)	(0.402)	(0.408)	(0.602)	(0.609)

Note. Alex is for Alexander et al. N refers to sample size in distribution 1 before truncation. Corrected correlation values shown in parentheses; observed correlation values not in parentheses.

Table 11

Mean Differences of the Observed and Corrected Correlation Values For Thorndike and Alexander et al. for Area 1 for N = 120

Truncation values	ρ (N = 120, number of repetitions = 2000) 0.2		0.4		0.6	
	Thorndike	Alex	Thorndike	Alex	Thorndike	Alex
10.5	0.062	0.062	0.113	0.113	0.141	0.141
	(0.005)	(–0.002)	(0.005)	(–0.006)	(0.004)	(–0.007)
11.0	0.041	0.041	0.074	0.074	0.090	0.090
	(0.001)	(–0.004)	(0.001)	(–0.008)	(0.000)	(–0.009)
12.0	0.011	0.011	0.020	0.020	0.023	0.023
	(–0.001)	(–0.005)	(–0.002)	(–0.008)	(–0.002)	(–0.009)

Note. Alex is for Alexander et al. N refers to sample size in distribution 1 before truncation. Thorndike and Alexander et al. corrected correlation values error estimated values shown in parentheses; observed error estimated values of correlations from truncation situation not in parentheses.

Table 12

Summary of the Absolute Values of the Mean Differences of the Observed and Corrected Correlation Values for Area 1 for N = 60

Truncation values	ρ (N = 60, number of repetitions = 2000)					
	0.2		0.4		0.6	
	Thorndike	Alex	Thorndike	Alex	Thorndike	Alex
10.5	0.012	0.006	0.014	0.013	0.010	0.015
11.0	0.006	0.007	0.005	0.015	0.003	0.017
12.0	0.002	0.010	0.004	0.018	0.003	0.019
Mean	0.0067	0.0077	0.0077	0.0153	0.0053	0.0170
Mean difference		0.0010		0.0076		0.0117
Factor		1.150		2.000		3.188

Note. Alex is for Alexander et al. N refers to sample size in distribution 1 before truncation. Mean difference is the mean difference between Thorndike and Alexander et al. Factor is the factor of Alexander et al. over Thorndike. Exact value of Thorndike = 0.006556. Exact value of Alexander et al. = 0.04.

Table 13

Summary of the Absolute Values of the Mean Differences of the Observed and Corrected Correlation Values for Area 1 for N = 120

	ρ (N = 120, number of repetitions = 2000)					
	0.2		0.4		0.6	
Truncation values	Thorndike	Alex	Thorndike	Alex	Thorndike	Alex
10.5	0.005	0.002	0.005	0.006	0.004	0.007
11.0	0.001	0.004	0.001	0.008	0.000	0.009
12.0	0.001	0.005	0.002	0.008	0.002	0.009
Mean	0.0023	0.0037	0.0027	0.0073	0.0020	0.0083
Mean difference		0.0014		0.0046		0.0063
Factor		1.571		2.750		4.167

Note. Alex is for Alexander et al. N refers to sample size in distribution 1 before truncation. Mean difference is the mean difference between Thorndike and Alexander et al. Factor is the factor of Alexander et al. over Thorndike. Exact value of Thorndike = 0.002333. Exact value of Alexander et al. = 0.006444.

Results for Research Question 1

Research question 1 addresses the overall effect of Thorndike's and Alexander et al.'s approaches and compares the results with the population correlations across simulated situations for different patterns of population correlations in the set of studies for area 1. Inspection of Tables 12 and 13 shows the results from the computer simulation on the robustness of the point of estimate of the corrected correlation for Thorndike's and Alexander et al.'s formulas for truncation values of 10.5, 11.0, and 12.0; ρ values of 0.2, 0.4, and 0.6; and sample sizes of 60 and 120 in area 1. As shown in Table 12, Thorndike's approach is closer to correct than Alexander et al.'s approach in terms of point of estimates of correlations for a sample size of 60. The differences of the mean estimates for Alexander et al.'s approach and Thorndike's approach in Table 12 ranged from 0.0010 to 0.0117. The factors of Alexander et al. over Thorndike are 1.150, 2.00, and 3.188. The exact values of Thorndike and Alexander et al. are 0.007 and 0.040. In every case, Thorndike's formula is better than Alexander et al.'s formula in terms of estimating the corrected correlation values. It is also shown that in every case in Table 12, the error estimates and the exact value of Thorndike are less than the error estimates and the exact value of Alexander et al. These results appear because Thorndike is used to estimate the corrected correlation value based on the known population variance (i.e., 1 in the simulation), while Alexander et al. is used to estimate the corrected estimate correlation value based on the Cohen ratio formula. Therefore, Thorndike's estimate of the true correlation value is closer to the population correlation value than Alexander et al.'s estimate. Nevertheless, Alexander et al.'s approach appears to give very good estimates. Likewise, shown in Table 13 is the result of the computer simulation on the

robustness of the point estimate of the corrected correlation for Thorndike and Alexander et al. for truncation values of 10.5, 11.0, and 12.0; ρ values of 0.2, 0.4, and 0.6; and a sample size of 120. As shown in the table, Thorndike is closer to correct than Alexander et al. in terms of point of estimates of correlations. The differences of the mean of point of error estimate of Alexander et al. and Thorndike in Table 13 ranged from 0.0014 to 0.0063. The factors between Alexander et al. and Thorndike are 1.571, 2.750, and 4.167. The exact values of Thorndike and Alexander et al. are 0.0023 and 0.0064. In every case in Table 13, the error estimates and the exact value of Thorndike are less than the error estimates and the exact value of Alexander et al. Thus, for both sample sizes of 60 and 120 in area 1, Thorndike gives a better estimate of corrected correlations than Alexander et al.; however, Alexander et al. appears to be a reasonable approximation.

Results for Research Question 2

Research question 2 addresses the comparison of the standard errors for both Thorndike's approach and Alexander et al.'s approach. Results of the simulation on standard error estimates are summarized in Table 14. In most of the cases, Alexander et al.'s average standard error is larger than Thorndike's average standard error at each cut point except at the minimal cut (i.e., 12.0), the difference between the standard errors of the two approaches is 0.004 at most. This is to be expected because Thorndike's formula estimates the corrected correlation from a known variance, while Alexander et al.'s formula estimates the corrected correlation using the Cohen ratio. Thus, it makes the standard error of the Alexander et al.'s formula larger than the standard error of Thorndike's formula in most cases. Additionally, the results also indicate that Alexander et al.'s correction formula has closer estimates to the true correlation value with sample

size equal to 60. Thus, it matches the earlier work from Alexander et al. (1984), which indicates that his formula can produce a very close estimate of the corrected correlation for a sample size of 60. Furthermore, Table 14 also shows that the observed correlation's standard error after truncation before any correction as additional information to help see the effect of the correction factors.

Table 14

Standard Errors of Estimated Correlations for Area 1

	ρ (Area = 1, number of replications = 2000)					
Truncation	N = 60			N = 120		
values	0.2	0.4	0.6	0.2	0.4	0.6
10.5						
(1)	0.133	0.125	0.106	0.087	0.080	0.068
(2)	0.188	0.162	0.116	0.121	0.102	0.073
(3)	0.203	0.182	0.141	0.127	0.115	0.091
11.0						
(1)	0.096	0.088	0.072	0.067	0.061	0.049
(2)	0.120	0.103	0.076	0.084	0.072	0.053
(3)	0.129	0.116	0.091	0.086	0.078	0.062
12.0						
(1)	0.047	0.043	0.035	0.033	0.030	0.024
(2)	0.052	0.051	0.045	0.037	0.036	0.032
(3)	0.053	0.050	0.041	0.036	0.034	0.028

Note. N refers to sample size in distribution before truncation.

(1) Standard error of estimate of the observed correlation values after truncation.
(2) Standard error of estimate of the estimated corrected correlation values Thorndike.
(3) Standard error of estimate of the estimated corrected correlation values Alexander et al.

Results from the computer simulation for situations 1 through 18 for areas 1 and 2 are shown in Tables 15, 16, 17, and 18. These tables represent the robustness of Thorndike's and Alexander et al.'s approaches for area 1 and 2. Table 15 gives mean

point of estimates of correlations for N = 60. Table 16 gives the mean differences for point estimates for N = 60. Likewise, Tables 17 and 18 give similar values for N = 120. Tables 15, 16, 17, and 18 provide the information that is summarized in Tables 19 and 20 to answer research question 3.

Table 15

Comparison of the Observed and Corrected Correlation Values in Area 1 and Area 2 for N = 60

	ρ (N = 60, number of repetitions = 2000)					
	0.2		0.4		0.6	
Truncation values	Thorndike	Alex	Thorndike	Alex	Thorndike	Alex
10.5	0.154	0.154	0.265	0.265	0.398	0.398
	(0.224)	(0.227)	(0.373)	(0.411)	(0.536)	(0.580)
11.0	0.275	0.275	0.393	0.393	0.527	0.527
	(0.337)	(0.381)	(0.471)	(0.530)	(0.612)	(0.673)
12.0	0.254	0.254	0.420	0.420	0.591	0.591
	(0.265)	(0.282)	(0.438)	(0.460)	(0.611)	(0.634)

Note. Alex is for Alexander et al. N is sample size. Corrected correlation values shown in parentheses; observed correlation values not in parentheses.

Table 16

Mean Differences of the Observed and Corrected Correlation Values for Area 1 and Area 2 for Thorndike and Alexander et al. for N = 60

Truncation values	ρ (N = 60, number of repetitions = 2000) 0.2		0.4		0.6	
	Thorndike	Alex	Thorndike	Alex	Thorndike	Alex
10.5	0.046	0.046	0.135	0.135	0.202	0.202
	(-0.024)	(-0.047)	(0.027)	(-0.011)	(0.064)	(0.020)
11.0	-0.075	-0.075	0.007	0.007	0.073	0.073
	(-0.137)	(-0.181)	(-0.071)	(-0.130)	(-0.012)	(-0.073)
12.0	-0.054	-0.054	-0.02	-0.02	0.009	0.009
	(-0.065)	(-0.082)	(-0.038)	(-0.06)	(-0.011)	(-0.034)

Note. Alex is for Alexander et al. N is sample size. Thorndike and Alexander et al. corrected correlation/error estimated values shown in parentheses; observed error estimated values of correlations from truncation situation not in parentheses.

Table 17

Comparison of the Observed and Corrected Correlation Values in Area 1 and Area 2

Truncation values	ρ (N = 120, number of repetitions = 2000) 0.2		0.4		0.6	
	Thorndike	Alex	Thorndike	Alex	Thorndike	Alex
10.5	0.179	0.179	0.297	0.297	0.436	0.436
	(0.260)	(0.283)	(0.373)	(0.411)	(0.583)	(0.622)
11.0	0.294	0.294	0.416	0.416	0.550	0.550
	(0.360)	(0.404)	(0.497)	(0.551)	(0.636)	(0.689)
12.0	0.277	0.277	0.435	0.435	0.598	0.598
	(0.288)	(0.301)	(0.450)	(0.468)	(0.614)	(0.633)

Note. Alex is for Alexander et al. N is sample size. Corrected correlation values shown in parentheses; observed correlation values not in parentheses.

Table 18

Mean Differences of the Observed and Corrected Correlation Values for Area 1 and Area 2

Truncation values	ρ (N = 120, number of repetitions = 2000) 0.2		0.4		0.6	
	Thorndike	Alex	Thorndike	Alex	Thorndike	Alex
10.5	0.021	0.021	0.103	0.103	0.164	0.164
	(-0.060)	(-0.083)	(-0.018)	(-0.053)	(0.017)	(-0.022)
11.0	-0.094	-0.094	-0.016	-0.016	0.050	0.050
	(-0.16)	(-0.204)	(-0.097)	(-0.151)	(-0.036)	(-0.089)
12.0	-0.077	-0.077	-0.035	-0.035	0.002	0.002
	(-0.088)	(-0.101)	(-0.050)	(-0.068)	(-0.014)	(-0.033)

Note. Alex is for Alexander et al. N is sample size. Thorndike and Alexander et al. corrected correlation error estimated values shown in parentheses; observed error estimated values of correlations from truncation situation not in parentheses.

Table 19

Summary of the Absolute Values of the Mean Differences of the Observed and Corrected Correlation Values for Area 1 and Area 2 for N = 60

Truncation values	ρ (N = 60, number of repetitions = 2000) 0.2		0.4		0.6	
	Thorndike	Alex	Thorndike	Alex	Thorndike	Alex
10.5	0.024	0.047	0.027	0.011	0.064	0.020
11.0	0.137	0.181	0.071	0.130	0.012	0.073
12.0	0.065	0.082	0.038	0.060	0.011	0.034
Mean	0.075	0.103	0.045	0.067	0.029	0.042
Mean Difference		0.028		0.022		0.013
Factor		1.371		1.478		1.460

Note. Alex is for Alexander et al. N refers to sample size in distribution 1 before truncation. Mean difference is the mean difference between Thorndike and Alexander et al. Factor is the factor of Alexander et al. over Thorndike. Exact value Thorndike = 0.049889. Exact value Alexander et al. = 0.070889.

Table 20

Summary of the Absolute Values of the Mean Differences of the Observed and Corrected Correlation Values for Area 1 and Area 2 for N = 120

	ρ (N = 120, number of repetitions = 2000)					
	0.2		0.4		0.6	
Truncation values	Thorndike	Alex	Thorndike	Alex	Thorndike	Alex
10.5	0.060	0.083	0.018	0.053	0.017	0.022
11.0	0.160	0.204	0.097	0.151	0.036	0.089
12.0	0.088	0.101	0.050	0.068	0.014	0.033
Mean	0.103	0.129	0.055	0.091	0.022	0.048
Mean difference		0.026		0.036		0.026
Factor		1.260		1.649		2.150

Note. Alex is for Alexander et al. N refers to sample size in distribution 1 before truncation. Mean difference is the mean difference between Thorndike and Alexander et al. Factor is the factor of Alexander et al.'s error estimate over Thorndike's error estimate. Exact value Thorndike = 0.06. Exact value Alexander et al. = 0.089333.

Results for Research Question 3

Research question 3 addresses the degree of accuracy in terms of the mean differences of the point estimates of correlations for Thorndike's approach and Alexander et al.'s approach in robustness situations (i.e., combined areas 1 and 2). Tables 19 and 20 show the results from the computer simulation in the robustness situations for Thorndike and Alexander et al. for truncation values of 10.5, 11.0, and 12.0; ρ values of 0.2, 0.4, 0.6; and sample size of 60 and 120. As shown in Table 19, Thorndike's approach is closer to correct than Alexander et al.'s approach in terms of mean differences when the sample size equals 60. In every summarized situation in Table 19, the mean differences of Thorndike's approach are smaller than Alexander et al.'s approach. Additionally, in every summarized situation, the mean difference of the two methods is never larger than

0.028; thus, I believe that Alexander et al.'s method is worthwhile. Likewise, Table 20 shows the results of the computer simulation in the robustness situations for Thorndike and Alexander for truncation values of 10.5, 11.0, and 12.0; ρ values of 0.2, 0.4, 0.6; and sample size of 120. As shown in Table 20, Thorndike's approach is closer to correct than Alexander et al.'s approach in terms of point of estimates of correlations. In every summarized situation, Thorndike's approach is better than Alexander et al.'s approach in term of estimating the population correlation values. These results appear because Thorndike's approach uses for estimating the corrected estimate correlation value based on the known parameters population variance (i.e., 1), while Alexander et al.'s approach uses the corrected estimate correlation value based on the Cohen ratio. Here, the restricted standard deviations are "thrown off" by the non-normal situations, which results in the Cohen ratio working with less accurate information when predicting the unrestricted variance. However, the main point is the magnitude of these differences, particularly in regard to Alexander et al.'s estimates and population values as compared to Thorndike's approach in robustness situations ranged from 1.37 to 2.15 larger factor for Alexander et al. to population value. Furthermore, in every case, the mean difference between Alexander et al.'s and Thorndike's formulas is never larger than 0.036. This again indicates that Alexander et al.'s method is a worthwhile estimate. Furthermore, simulation results also illustrate that the degree of truncation contributes to increased correlation for a few situations when both areas 1 and 2 are mixed. As shown in Table 16 and Table 18, the mean differences of Thorndike's approach and Alexander et al.'s approach at a truncation value of 11.0 for both sample sizes of 60 and 120 are larger when compared to the other mean estimates of Thorndike's and Alexander et al.'s

approaches at other truncation values. These results appear because there are pairs of scores (x,y) from area 2 in which the x values are close to the truncation line that contribute to an increase in the uncorrected observed correlation values in robust situations. Thus, with uncorrected observed correlation values that are less representative of the nominal population correlations (i.e., ρ = .2, .4, or .6), the adjustment formulas start with less accurate data.

Results for Research Question 4

Research question 4 addresses the comparison of the standard error estimates of Thorndike's formula and Alexander et al.'s formula for the robustness situations. Results of the simulation for standard error estimates are summarized in Table 21. In almost every case, average standard error estimates for Alexander et al.'s approach are larger than the average standard error for Thorndike's approach at each cut point. Furthermore, with Alexander et al.'s formula at the most severe cut (i.e., 10.5), the difference between the two methods is never greater than 0.030 except as stated earlier in a few situations in the contaminated normal situations. Table 21 shows the observed correlations after truncation as a guide to help interpret the other standard error.

Table 21

Standard Error Estimates for Area 1 + 2

Truncation values	ρ (Area 1 + 2, number of replications = 2000) N = 60			N = 120		
	0.2	0.4	0.6	0.2	0.4	0.6
10.5						
(1)	0.164	0.155	0.144	0.118	0.106	0.094
(2)	0.235	0.212	0.178	0.169	0.144	0.112
(3)	0.265	0.237	0.196	0.186	0.156	0.122
11.0						
(1)	0.147	0.122	0.097	0.101	0.078	0.059
(2)	0.178	0.144	0.106	0.124	0.093	0.063
(3)	0.201	0.156	0.113	0.186	0.097	0.068
12.0						
(1)	0.099	0.067	0.041	0.073	0.049	0.030
(2)	0.100	0.067	0.045	0.076	0.051	0.033
(3)	0.108	0.073	0.045	0.079	0.053	0.033

Note. N refers to sample size in distribution before truncation.

(1) Standard error of the observed correlation values without correction after truncation.
(2) Standard error of the corrected correlation values Thorndike.
(3) Standard error of the corrected correlation values Alexander et al.

Discussion

After studying and comparing Thorndike's and Alexander et al.'s approaches for estimating the corrected correlation values, I summarize and discuss the results in the following three areas: The accuracy of the correction formulas, the need for the correction formulas, and the worthiness of Alexander et al.'s formula.

1. Thorndike's formula was a better estimate of the corrected correlation than Alexander et al.'s formula in both non-robustness and robustness situations for truncation values of 10.5, 11.0, 12.0; ρ values of 0.2, 0.4, 0.6; and sample sizes of 60 and 120. This is expected because Thorndike's formula uses known variance, while Alexander et al.'s formula uses the Cohen ratio to estimate the corrected correlations. The computer simulation shows that for every case in the tables, from situations 1 to 18, Thorndike's mean differences from the population values are smaller than Alexander et al.'s mean differences, and Thorndike's standard errors are lower than Alexander et al.'s standard errors for both non-robustness and robustness situations. This, once again, confirms the expectation that Thorndike's approach is better than Alexander et al.'s approach in estimating corrected correlations for both non-robustness and robustness situations, given that a population variance for the unrestricted situation is available. This is frequently not the case, so the accuracy of Alexander et al.'s formula in comparison to population values is a major focus. The advantage of Alexander et al.'s approach is that it uses the data on hand by estimating the population variance from the available restricted variance. As stated earlier, the mean differences of the estimated values compared to population values for Alexander et al.'s approach compared to Thorndike's approach in robustness situations ranging from a factor of 1.37 to 2.15, and its magnitude of mean differences ranged from 0.013 to 0.036.

2. Corrections are needed because without the corrections, the correlation results tend to be much less accurate. The corrected correlations give results closer to the ρ values for both non-robustness and robustness situations except for a few robustness situations discussed below. This has implications for summarizing correlations in meta-analysis for situations which have restricted range based on one of the observed variables.

3. As stated earlier, Thorndike's corrected correlation values were more accurate than Alexander et al.'s corrected correlation values in terms of both mean differences and standard errors. Mean differences of the correlations for Alexander et al.'s approach compared to Thorndike's approach in robustness situations ranged from 1.37 to 2.15 larger. Nevertheless, Alexander et al.'s approach, based only on estimated variances, appears to be a worthwhile correction in most of the situations that were simulated, with a few notable exceptions. The exceptions occurred for truncation value of 11.0 for sample sizes of 60 and 120. This indicates that how non-normality plays an important role in robustness situations. As shown in Table 16 and Table 18, the mean differences of Thorndike's approach and Alexander et al.'s approach at a truncation value of 11.0 are larger when compared to the mean differences of Thorndike's approach and Alexander et al.'s approach at other truncations. I believe this occurs in robustness situations because there are more x scores from area 2 next to the truncation line, which contributes to a more inaccurate estimate of the restricted variance resulting in Cohen's formula, giving a less accurate estimate of the unrestricted variance. Additionally, the degree of

skewness in mixed areas 1 and 2 is shown in Appendix E for sample sizes of 60 and 120, ranging from –0.408 to –0.766, thus it indicated that a degree of negative skewness happened in mixed areas 1 and 2 after truncation occurred at 11.0 for every ρ value of 0.2, 0.4, and 0.6. A negative-skew situation exists when there is a long tail in the negative direction; whereas, a positive-skew situation occurs in the opposite direction. Additionally, in most of the cases, the degree of negative skewness for the robustness situation (i.e, combined areas 1 and 2) tended to decrease when there was an increase in sample size, as shown in Appendix E. Furthermore, the corrected correlations for both formulas are adversely affected by observed uncorrected correlations that may be inflated. This is expected because range restriction sometimes increases correlation between variables in a contaminated normal distribution (Zimmerman & Williams, 2000).

Limitations Leading to Future Directions

One limitation of the current study is that it does not address the double truncation on both distributions 1 and 2. Alexander et al. (1990) suggested to experiment with Wells and Fruchter's (1970) approach for corrected correlation when both x and y are truncated. The current research also does not address the corrected correlation for the skew-normal distributions 1 and 2 or the skew-skew of distributions 1 and 2. This limitation could be alleviated by modifying the current distribution function in the existing computer program to produce skewed distributions in lieu of normal distributions.

Another limitation is related to the increased correlation in a contaminated normal or robustness situation, which occurs when areas 1 and 2 are mixed together. This

increased correlation could be resolved by implementing a filtering algorithm based on the detection of highly at-risk scores from area 2 that are close to the truncation line and preventing them from entering common areas 1 and 2. The filtering algorithm is based on the “divide and conquer” computational methodology in which at-risk scores from area 2 would be searched by the binary search algorithm and eliminated before reaching the common areas 1 and 2. Another limitation of the research is that the formulas investigated did not include measurement error; however, this research produces a baseline that can be replicated with an error in the variables model to compare the effect reliability corrected correlations. These limitations provide directions for future research that can build on the results of this study.

References

Alexander, R. A., Alliger, G. M., & Hanges, P. J. (1990). Correcting for range restriction when the population variance is unknown. *Applied Psychological Measurement*, *8*, 431–437.

Alexander, R. A., Carson, K. P., & Alliger, G. M. (1990). Correction for restriction of range when both x and y are truncated. *Applied Psychological Measurement*, *18*, 355–367.

Alexander, R. A., Hanges, P. J., & Alliger, G. M. (1984). Correcting for restriction of range in both x and y when the unrestricted variances are unknown. *Applied Psychological Measurement*, *8*(4), 431–437.

Cochran, G. W. (1977). *Sampling techniques*. Hoboken, NJ: John Wiley & Sons.

Cohen, J. (1987). *Statistical power analysis for the behavior sciences*. Mahwah, NJ: Lawrence Erlbaum & Associates.

Crawley, J. M. (2007). *The R Book*. Hoboken, NJ: John Wiley & Sons.

Crawley, J. M. (2005). *Statistics: An introduction using R*. Hoboken, NJ: John Wiley & Sons.

Groeneveld, R. A., & Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician*, *33*(4), 391–399.

Henriksson, W., & Wolming, S. (1998). Academic performance in four studies programs: A comparison of students admitted on the basic of GPA and SWESAT scores, with and without credits for work experience. *Scandinavian Journal of Educational Research*, *42*, 135–150.

Horowitz, E., & Sahni, E., (1990). *Fundamentals of data structures*. New York: Computer Science Press.

Hunter , J. E., & Schmidt, L. F. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage Publications.

Johnson, R. A., & Wichern, D. W. (1998). *Applied multivariate statistical analysis*. Englewood Cliffs, NJ: Prentice Hall.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*(1), 148–161.

Oleksandr, S. C., & Deniz S. O. (1999). How selective are psychology graduate programs? The effect of the selection ratio on GRE scores validity. *Educational and Psychological Measurement*, *59*, 951–961.

Pearson, K. (1903). Mathematical contribution to the theory of evolution-XI on the influence of natural selection on the variability and correlation of organs. *Phil. Trans. Royal Soc. of London, Series A*, 200, 1–66.

Thorndike, R. L. (1947). *Research problems and techniques (Rep. No. 3 AAF Aviation Psychology Program Research Reports*). Washington, DC: U.S. Government Printing Office.

Schmidt, F. L., & Hunter, L. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletins*, *124*(2), 262–274.

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel Selection. *Annual Review of Psychology*, *43*, 627–670.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, *81*, 557–574.

Wells, D. G., & Fruchter, B. (1970). Correcting the correlation coefficient for explicit restriction on both variables. *Educational and Psychological Measurement, 30*, 925–934

Zimmerman, D. W., & Williams, R. H. (2000). Restriction of range and correlation in outlier-prone distribution. *Applied Psychological Measurement*, *24*(2), 267–279.

APPENDIXES

Appendix A

Results for Simulation Situations

The following tables provide the results for simulated situations 1 to 18. For all simulated situations, the number of replications is 2000. The same random numbers generated in distribution 1 were used for area 1 by itself and for area 1 from distribution 1 when area 1 was combined with area 2. The population standard deviation in distribution 1 is 1, and the population standard deviation in distribution 2 is also 1. Thorndike stands for Thorndike's (1947) case 2 Formula. The Thorndike's case 2 Formula is used to calculate the estimated corrected correlation value in a direct restriction of range. Alexander stands for Alexander et al.'s (1984) formula which uses the Cohen's (1959) ratio formula. (1984). The Cohen's (1959) ratio is defined as the ratio of the sample variance over the difference between the sample mean and the point of truncation squared. The "Mean correlation" is the mean correlation across 2000 replications. The "Error estimate" is the mean correlation minus the population correlation. The "Std error" is the standard deviation of the estimated correlation minus the population correlation. The median is the median correlation across 2000 replications.

Situation 1

Exogenous Variables

Population correlation = 0.2,   Sample size = 60,

Population mean of distribution 1 = 10,   Population mean of distribution 2 = 11,

Truncation value = 10.5

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.134	0.188	0.134	0.206	0.154	0.224	0.154	0.247
Error Est	0.066	0.012	0.066	-0.006	0.046	-0.024	0.046	-0.047
Std Error Est	0.133	0.188	0.133	0.203	0.164	0.235	0.164	0.265
Median	0.138	0.192	0.138	0.212	0.177	0.261	0.177	0.298

Situation 2

Exogenous Variables

Population correlation = 0.2,   Sample size = 60,

Population mean of distribution 1 = 10,   Population mean of distribution 2 = 12,

Truncation value = 11.0

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.155	0.194	0.155	0.207	0.275	0.337	0.275	0.381
Error Est	0.045	0.006	0.045	-0.007	-0.075	-0.137	-0.075	-0.181
Std Error Est	0.096	0.120	0.096	0.129	0.147	0.178	0.147	0.201
Median	0.160	0.199	0.16	0.209	0.304	0.372	0.304	0.423

Situation 3

Exogenous Variables

Population correlation = 0.2, Sample size = 60,

Population mean of distribution 1 = 10, Population mean of distribution 2 = 14,

Truncation value = 12.0

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.189	0.202	0.189	0.210	0.254	0.265	0.254	0.282
Error Est	0.011	-0.002	0.011	-0.010	-0.054	-0.065	-0.054	-0.082
Std Error Est	0.047	0.052	0.047	0.053	0.099	0.100	0.099	0.108
Median	0.200	0.207	0.200	0.217	0.218	0.238	0.218	0.246

Situation 4
Exogenous Variables

Population correlation = 0.4,   Sample size = 60,

Population mean of distribution 1 = 10,   Population mean of distribution 2 = 11,

Truncation value = 10.5

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.284	0.386	0.284	0.413	0.265	0.373	0.265	0.411
Error Est	0.116	0.014	0.116	-0.013	0.135	0.027	0.135	-0.011
Std Error Est	0.125	0.162	0.125	0.182	0.155	0.212	0.155	0.237
Median	0.289	0.396	0.289	0.424	0.291	0.414	0.291	0.457

Situation 5

Exogenous Variables

Population correlation = 0.4,   Sample size = 60,

Population mean of distribution 1 = 10,   Population mean of distribution 2 = 12,

Truncation value = 11.0

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.323	0.395	0.323	0.415	0.393	0.471	0.393	0.530
Error Est	0.077	0.005	0.077	-0.015	0.007	-0.071	0.007	-0.130
Std Error Est	0.088	0.103	0.088	0.116	0.122	0.144	0.122	0.156
Median	0.327	0.402	0.327	0.419	0.419	0.499	0.419	0.561

Situation 5

Exogenous Variables

Population correlation = 0.4, Sample size = 60,

Population mean of distribution 1 = 10, Population mean of distribution 2 = 12,

Truncation value = 11.0

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.323	0.395	0.323	0.415	0.393	0.471	0.393	0.530
Error Est	0.077	0.005	0.077	-0.015	0.007	-0.071	0.007	-0.130
Std Error Est	0.088	0.103	0.088	0.116	0.122	0.144	0.122	0.156
Median	0.327	0.402	0.327	0.419	0.419	0.499	0.419	0.561

Situation 7

Exogenous Variables

Population correlation = 0.6,   Sample size = 60,

Population mean of distribution 1 = 10,   Population mean of distribution 2 = 11,

Truncation value = 10.5

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.456	0.590	0.456	0.615	0.398	0.536	0.398	0.580
Error Est	0.144	0.010	0.144	-0.015	0.202	0.064	0.202	0.020
Std Error Est	0.106	0.116	0.106	0.141	0.144	0.178	0.144	0.196
Median	0.465	0.600	0.465	0.627	0.419	0.574	0.419	0.617

Situation 8

Exogenous Variables

Population correlation = 0.6, Sample size = 60,

Population mean of distribution 1 = 10, Population mean of distribution 2 = 12,

Truncation value = 11.0

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.507	0.597	0.507	0.617	0.527	0.612	0.527	0.673
Error Est	0.093	0.003	0.093	-0.017	0.073	-0.012	0.073	-0.073
Std Error Est	0.072	0.076	0.072	0.091	0.097	0.106	0.097	0.113
Median	0.513	0.603	0.513	0.622	0.543	0.631	0.543	0.692

Situation 9

Exogenous Variables

Population correlation = 0.6,   Sample size = 60,

Population mean of distribution 1 = 10,   Population mean of distribution 2 = 14,

Truncation value = 12.0

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.577	0.603	0.577	0.619	0.591	0.611	0.591	0.634
Error Est	0.023	-0.003	0.023	-0.019	0.009	-0.011	0.009	-0.034
Std Error Est	0.035	0.045	0.035	0.041	0.041	0.045	0.041	0.045
Median	0.587	0.606	0.587	0.625	0.600	0.614	0.600	0.636

Situation 10

Exogenous Variables

Population correlation = 0.2, Sample size = 120,

Population mean of distribution 1 = 10, Population mean of distribution 2 = 11,

Truncation value = 10.5

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.138	0.195	0.138	0.202	0.179	0.260	0.179	0.283
Error Est	0.062	0.005	0.062	-0.002	0.021	-0.060	0.021	-0.083
Std Error Est	0.087	0.121	0.087	0.127	0.118	0.169	0.118	0.186
Median	0.139	0.199	0.139	0.203	0.197	0.292	0.197	0.317

Situation 11
Exogenous Variables

Population correlation = 0.2, Sample size = 120,

Population mean of distribution 1 = 10, Population mean of distribution 2 = 12,

Truncation value = 11.0

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.159	0.199	0.159	0.204	0.294	0.360	0.294	0.404
Error Est	0.041	0.001	0.041	-0.004	-0.094	-0.16	-0.094	-0.204
Std Error Est	0.067	0.084	0.067	0.086	0.101	0.124	0.101	0.186
Median	0.161	0.201	0.161	0.207	0.313	0.383	0.313	0.432

Situation 12
Exogenous Variables

Population correlation = 0.2,   Sample size = 120,

Population mean of distribution 1 = 10,   Population mean of distribution 2 = 14,

Truncation value = 12.0

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.189	0.201	0.189	0.205	0.277	0.288	0.277	0.301
Error Est	0.011	-0.001	0.011	-0.005	-0.077	-0.088	-0.077	-0.101
Std Error Est	0.033	0.037	0.033	0.036	0.073	0.076	0.073	0.079
Median	0.192	0.203	0.192	0.209	0.280	0.287	0.280	0.305

Situation 13

Exogenous Variables

Population correlation = 0.4, Sample size = 120,

Population mean of distribution 1 = 10, Population mean of distribution 2 = 11,

Truncation value = 10.5

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.287	0.395	0.287	0.406	0.297	0.418	0.297	0.453
Error Est	0.113	0.005	0.113	-0.006	0.103	-0.018	0.103	-0.053
Std Error Est	0.080	0.102	0.080	0.115	0.106	0.144	0.106	0.156
Median	0.288	0.399	0.288	0.407	0.312	0.447	0.312	0.478

Situation 14
Exogenous Variables

Population correlation = 0.4, Sample size = 120,

Population mean of distribution 1 = 10, Population mean of distribution 2 = 12,

Truncation value = 11.0

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.326	0.399	0.326	0.408	0.416	0.497	0.416	0.551
Error Est	0.074	0.001	0.074	-0.008	-0.016	-0.097	-0.016	-0.151
Std Error Est	0.061	0.072	0.061	0.078	0.078	0.093	0.078	0.097
Median	0.329	0.403	0.329	0.411	0.429	0.513	0.429	0.566

Situation 15

Exogenous Variables

Population correlation = 0.4,   Sample size = 120,

Population mean of distribution 1 = 10,   Population mean of distribution 2 = 14,

Truncation value = 12.0

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.380	0.402	0.380	0.408	0.435	0.450	0.435	0.468
Error Est	0.020	-0.002	0.020	-0.008	-0.035	-0.050	-0.035	-0.068
Std Error Est	0.030	0.036	0.030	0.034	0.049	0.051	0.049	0.053
Median	0.384	0.404	0.384	0.412	0.438	0.451	0.438	0.471

Situation 16

Exogenous Variables

Population correlation = 0.6, Sample size = 120,

Population mean of distribution 1 = 10, Population mean of distribution 2 = 11,

Truncation value = 10.5

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.459	0.596	0.459	0.607	0.436	0.583	0.436	0.622
Error Est	0.141	0.004	0.141	-0.007	0.164	0.017	0.164	-0.022
Std Error Est	0.068	0.073	0.068	0.091	0.094	0.112	0.094	0.122
Median	0.461	0.600	0.461	0.610	0.451	0.606	0.451	0.642

Situation 17

Exogenous Variables

Population correlation = 0.6,   Sample size = 120,

Population mean of distribution 1 = 10,   Population mean of distribution 2 = 12,

Truncation value = 11.0

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.510	0.600	0.510	0.609	0.550	0.636	0.550	0.689
Error Est	0.090	0.000	0.090	-0.009	0.050	-0.036	0.050	-0.089
Std Error Est	0.049	0.053	0.049	0.062	0.059	0.063	0.059	0.068
Median	0.513	0.602	0.513	0.612	0.556	0.644	0.556	0.696

Situation 18

Exogenous Variables

Population correlation = 0.6, Sample size = 120,

Population mean of distribution 1 = 10, Population mean of distribution 2 = 14,

Truncation value = 12.0

Comparison of Observed and Corrected Correlation Values

	Area 1				Area 1 + Area 2			
	Thorndike		Alexander et al.		Thorndike		Alexander et al.	
Statistics	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr	Obs Corr	Est Corr
Mean Corr	0.577	0.602	0.577	0.609	0.598	0.614	0.598	0.633
Error Est	0.023	-0.002	0.023	-0.009	0.002	-0.014	0.002	-0.033
Std Error Est	0.024	0.032	0.024	0.028	0.030	0.033	0.030	0.033
Median	0.581	0.603	0.581	0.612	0.599	0.615	0.599	0.634

Appendix B

Cohen's Table

Cohen Ratio	SD_{tab}	z-score
0.109	.993	3.00
0.113	.992	2.95
0.116	.991	2.90
0.120	.990	2.85
0.124	.989	2.80
0.128	.987	2.75
0.132	.986	2.70
0.137	.984	2.65
0.141	.982	2.60
0.146	.980	2.55
0.151	.978	2.50
0.156	.975	2.45
0.161	.972	2.40
0.167	.969	2.35
0.172	.966	2.30
0.178	.963	2.25
0.184	.959	2.20
0.190	.955	2.15
0.197	.951	2.10
0.203	.946	2.05
0.210	.942	2.00
0.217	.936	1.95

0.224	.931	1.90
0.231	.926	1.85
0.239	.920	1.80
0.247	.914	1.75
0.254	.907	1.70
0.263	.901	1.65
0.271	.894	1.60
0.279	.886	1.55
0.288	.879	1.50
0.296	.871	1.45
0.305	.863	1.40
0.314	.855	1.35
0.323	.847	1.30
0.332	.838	1.25
0.342	.830	1.20
0.351	.821	1.15
0.361	.812	1.10
0.370	.803	1.05
0.380	.794	1.00
0.389	.784	0.95
0.399	.775	0.90
0.409	.765	0.85
0.419	.756	0.80
0.429	.746	0.75
0.438	.736	0.70

0.448	.726	0.65
0.458	.717	0.60
0.468	.707	0.55
0.477	.697	0.50
0.487	.688	0.45
0.497	.678	0.40
0.506	.668	0.35
0.516	.659	0.30
0.525	.649	0.25
0.534	.640	0.20
0.544	.630	0.15
0.553	.621	0.10
0.562	.612	0.05
0.571	.603	0.00
0.580	.594	-0.05
0.588	.585	-0.10
0.597	.576	-0.15
0.605	.568	-0.20
0.614	.559	-0.25
0.622	.551	-0.30
0.630	.542	-0.35
0.638	.534	-0.40
0.646	.526	-0.45
0.653	.518	-0.50
0.661	.510	-0.55

0.668	.503	-0.60
0.675	.495	-0.65
0.682	.488	-0.70
0.689	.481	-0.75
0.696	.473	-0.80
0.703	.466	-0.85
0.709	.460	-0.90
0.716	.453	-0.95
0.722	.446	-1.00
0.728	.440	-1.05
0.734	.433	-1.10
0.740	.427	-1.15
0.746	.421	-1.20
0.751	.415	-1.25
0.757	.409	-1.30
0.762	.403	-1.35
0.767	.398	-1.40
0.772	.392	-1.45
0.777	.387	-1.50
0.782	.381	-1.55
0.787	.376	-1.60
0.791	.371	-1.65
0.796	.366	-1.70
0.800	.361	-1.75
0.804	.356	-1.80

0.809	.352	-1.85
0.813	.347	-1.90
0.817	.343	-1.95
0.820	.338	-2.00
0.825	.334	-2.05
0.828	.329	-2.10
0.832	.325	-2.15
0.835	.321	-2.20
0.838	.317	-2.25
0.842	.313	-2.30
0.845	.309	-2.35
0.848	.306	-2.40
0.851	.302	-2.45
0.855	.298	-2.50
0.857	.295	-2.55
0.860	.291	-2.60
0.864	.288	-2.65
0.867	.285	-2.70
0.869	.281	-2.75
0.868	.278	-2.80
0.875	.275	-2.85
0.875	.272	-2.90
0.882	.269	-2.95
0.882	.266	-3.00

Note : Taken from Alexander et al. (1984). Permission is found in Appendix D.

Appendix C

Thorndike's and Alexander et al.'s Variances in the Unrestricted and Restricted Situations

The following table illustrates where and how to obtain the variances in Thorndike's and Alexander et al.'s formulas.

	Thorndike	Alexander et al.
Restricted variance	Sample estimate	Sample estimate
Unrestricted variance	Population variance	Cohen's approach to use sample to estimate population variance

In Thorndike's formula, both the restricted and unrestricted variance are known from the sample estimate, and population variance whereas in Alexander et al.'s formula, the restricted variance is known from the sample estimate, and the unrestricted variance has to be estimated from the sample variance using the Cohen's formula.

Appendix D

Dear Dung Minh Tran,

Thank you for placing your order with Copyright Clearance Center.

Order Summary:
Order Date: 04/11/2011
Confirmation Number: 10340220
Items in order: 1
Order Total: $ 4.80

Appendix D Continue

Copyright Clearance Center RightsLink®

Home Create Account Help

SAGE

Title:	Correcting for Range Restriction When the Population Variance is Unknown
Author:	Ralph A. Alexander, George M. Alliger, Paul J. Hanges
Publication:	Applied Psychological Measurement
Publisher:	Sage Publications
Date:	09/01/1984

Gratis

Permission is granted at no cost for sole use in a Master's Thesis and/or Doctoral Dissertation. Additional permission is also granted for the selection to be included in the printing of said scholarly work as part of UMI's "Books on Demand" program. For any further usage or publication, please contact the publisher.

Appendix E

Mean of Skewness for Sample Size of 60, Truncation Value of 11.0, and Number of Repetitions of 10

ρ	Area 1	Areas 1 and 2
0.2	-0.249	-0.408
0.4	-0.503	-0.567
0.6	-0.667	-0.766

Mean of Skewness for Sample Size of 120, Truncation Value of 11.0, and Number of Repetitions of 10

ρ	Area 1	Areas 1 and 2
0.2	-0.352	-0.537
0.4	-0.438	-0.496
0.6	-0.441	-0.544

Appendix F

Figure 1

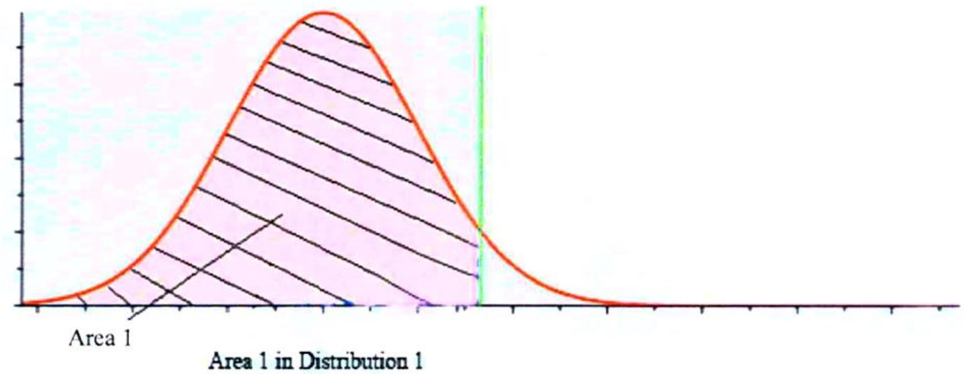


Figure 2

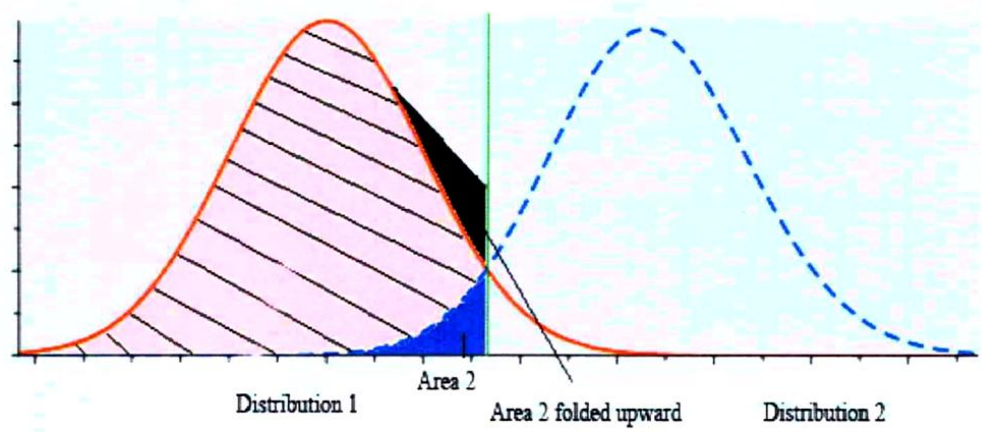
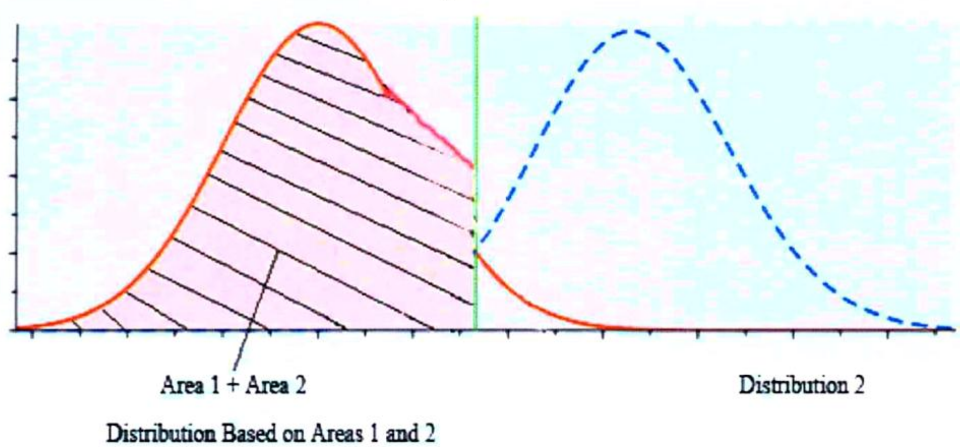


Figure 3



Appendix G

A Snippet of Calculating the Corrected Standard Deviation from the Alexander et al.'s formula Using the Cohen's Table

set up Cohen's Table = [Cohen ratio, STD tab, and Z truncation].

Table look up could be done by matching the SD tab value based on key value of the Cohen ### ratio.

See explanation in background section for computing the Cohen ratio, and the corrected

standard deviation for the Alexander’s Formula.

comments are followed the pound sign

```
cohenRatio<-
c(.109,.113,.116,.120,.124,.128,.132,.137,.141,.146,.151,.156,.161,.167,.172,.178,.184
,.190,.197,.203,.210,.217,.224,.231,.239,.247,.254,.263,.271,.279,.288,.296,.305,.314,.323
,.332,.342,.351,.361,.370,.380,.389,.399,.409,.419,.429,.438,.448,.458,.468,.477,.487,.497
,.506,.516,.525,.534,.544,.553,.562,.571,.580,.588,.597,.605,.614,.622,.630,.638,.646,.653
,.661,.668,.675,.682,.689,.696,.703,.709,.716,.722,.728,.734,.740,.746,.751,.757,.762,.767
,.772,.777,.782,.787,.791,.796,.800,.804,.809,.813,.817,.820,.825,.828,.832,.835,.838,.842
,.845,.848,.851,.855,.857
,.860,.864,.867,.869,.868,.875,.875,.882,.882)
SDtab<-
c(.993,.992,.991,.990,.989,.987,.986,.984,.982,.980,.978,.975,.972,.969,.966,.963,.959,.9
55
```

,.951,.946,.942,.936,.931,.926,.920,.914,.907,.901,.894,.886,.879,.871,.863,.855,.847,.838

,.830,.821,.812,.803,.794,.784,.775,.765,.756,.746,.736,.726,.717,.707,.697,.688,.678,.668

,.659,.649,.640,.630,.621,.612,.603,.594,.585,.576,.568,.559,.551,.542,.534,.526,.518,.510

,.503,.495,.488,.481,.473,.466,.460,.453,.446,.440,.433,.427,.421,.415,.409,.403,.398,.392

,.387,.381,.376

,.371,.366,.361,.356,.352,.347,.343,.338,.334,.329,.325,.321,.317,.313,.309,.306,.302,.298

,.295,

.291,.288,.285,.281,.278,.275,.272,.269,.266)

ZTruncation<-

c(3.00,2.95,2.90,2.85,2.80,2.75,2.70,2.65,2.60,2.55,2.50,2.45,2.40,2.35,2.30,2.25,2.20,2.

15,2.10,2.05,2.00,1.95,1.90,1.85,1.80,1.75,1.70,1.65,1.60,1.55,1.50,1.45,1.40,1.35,1.30,1.

25,1.20,1.15,1.10,1.05,1.00,0.95,0.90,0.85,0.80,0.75,0.70,0.65,0.60,0.55,0.50,0.45,0.40,0.

35,0.30,0.25,0.20,0.15,0.10,0.05,0.00,-0.05,-0.10,-0.15,-0.20,-0.25,-0.30,-0.35,-0.40,-

0.45,-0.50,-0.55,-0.60,-0.65,-0.70,-0.75,-0.80,-0.85,-0.90,-0.95,-1.00,-1.05,-1.10,-1.15,-

1.20,-1.25,-1.30,-1.35,-1.40,-1.45,-1.50,-1.55,-1.60,-1.65,-1.70,-1.75,-1.80,-1.85,-1.90,-

1.95,-2.00,-2.05,-2.10,-2.15,-2.20,-2.25,-2.30,-2.35,-2.40,-2.45,-2.50,-2.55,-2.60,-2.65,-

2.70,-2.75,-2.80,-2.85,-2.90,-2.95,-3.00)

CorrectedStd<- function (RestSample, StdObs)

Corrected Standard deviation function with input restricted sample vector, and the

standard ## deviation of the restricted sample

{

```
CohenTableVector<- data.frame (cohenRation,SDtab,ZTruncation)

dx<- 0.005

sortRestSample<-  sort (RestSample[,c(1)])

  ## sort in ascending order x values of the restricted sample

highestXValue<- sortRestSample[length (RestSample[,c(1)])]

diffDeno<- square(mean(RestSample[,c(1)]) -  highestXValue)

 ## compute square of the difference of the mean of the restricted sample and the highest

###score X value

cohenRatio<- var (RestSample[,c(1)]) / diffDeno

 ## compute the Cohen ratio by simply calculating the ratio of the variance of the

restricted

## sample and the above equation on diffDeno

 cohenZRestrictedArea1 <- 0

 cohenSDtabRestrictedArea1 <- 0

for (i in 1 : length (CohenTableVector[,c(1)]))

  {

    ## Performing the table look up on Cohen's ratio

if (( cohenRatio - CohenTableVector[,c(1)][i]) < dx)

      {

            cohenZRestrictedArea1 <- CohenTableVector[,c(3)][i]

            cohenSDtabRestrictedArea1 <- CohenTableVector[,c(2)][i]

break;

            ### exit of the for loop once the Cohen ratio is found
```

```
        }


   }
CorrectedStdValue<- stdObs / cohenSDtabRestrictedArea1;
## The corrected standard deviation in the Alexander's formula is obtained by the ratio of
the
## standard deviation of the restricted sample and the Sd tab
returnCorrectedStdValue
}
```