CONVERGENT VALIDITY OF VARIABLES RESIDUALIZED BY A SINGLE COVARIATE: THE ROLE OF CORRELATED ERROR IN POPULATIONS AND SAMPLES

Kim Nimon

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

May 2013

APPROVED:

Robin K. Henson, Major Professor
Jeff Allen, Minor Professor
Becky Glover, Committee Member
Darrell Hull, Committee Member
Abbas Tashakkori, Chair of the Department Educational Psychology
Jerry R. Thomas, Dean of the College of Education
Mark Wardell, Dean of the Toulouse Graduate School

Nimon, Kim, Convergent Validity of Variables Residualized by a Single Covariate: The Role of Correlated Error in Populations and Samples. Doctor of Philosophy (Educational Research), May 2013, 55 pp., 7 tables, 11 illustrations, references, 32 titles.

This study examined the bias and precision of four residualized variable validity estimates (C0, C1, C2, C3) across a number of study conditions. Validity estimates that considered measurement error, correlations among error scores, and correlations between error scores and true scores (C3) performed the best, yielding no estimates that were practically significantly different than their respective population parameters, across study conditions. Validity estimates that considered measurement error and correlations among error scores (C2) did a good job in yielding unbiased, valid, and precise results. Only in a select number of study conditions were C2 estimates unable to be computed or produced results that had sufficient variance to affect interpretation of results. Validity estimates based on observed scores (C0) fared well in producing valid, precise, and unbiased results. Validity estimates based on observed scores that were only corrected for measurement error (C1) performed the worst. Not only did they not reliably produce estimates even when the level of modeled correlated error was low, C1 produced values higher than the theoretical limit of 1.0 across a number of study conditions. Estimates based on C1 also produced the greatest number of conditions that were practically significantly different than their population parameters.

CONVERGENT VALIDITY OF VARIABLES RESIDUALIZED BY A SINGLE COVARIATE: THE ROLE OF CORRELATED ERROR IN POPULATIONS AND SAMPLES

Controlling variance is a critical aspect of quantitative research design and analysis. One of the most powerful forms of control is the process of random assignment of individuals to groups. However, in much behavioral science research, random assignment is not possible, does not completely control for between group differences, and may not control for nuisance variables. Researchers may, therefore, employ statistical methods to control for variance in dependent and/or independent variables.

It is not uncommon for researchers to invoke statistical procedures that attempt to statistically control for variance in a dependent variable. Analysis of covariance (ANCOVA), for example, is one such procedure that may be employed to partial out variance in the dependent variable that is attributable to covariates (e.g., variables that may explain group differences prior to intervention) before examining the effect of the intervention or treatment. A researcher may not be able to randomly assign treatment groups and, therefore, uses a pretest of the dependent variable as a covariate in the analysis. These covariance correction analyses have the generally common goal of refining the analysis to evaluate intervention effects or relationships after having statistically controlled for other potential influences.

It is also not uncommon for researchers to invoke statistical procedures that attempt to statistically control for variance in an independent variable. Hierarchical multiple regression, for example, is one such procedure that may be employed to remove variance from one independent variable that is shared with others in a set. A researcher may observe that a sample has dissimilar characteristics (i.e., demographics) that are extraneous to the relations under study. Including such characteristics as independent variables has the goal of isolating the effect of the independent variable of interest after having statistically controlled for influences associated with the extraneous variables.

What is less understood is the nature of variables after such controls have been employed (Lynam, Hoyle, & Newman, 2006; Thompson, 1992; Tracz, Nelson, Newman, & Beltran, 2005). To be clear, the dependent variable used in covariance-corrected analyses such as ANCOVA has first been residualized by the covariate (e.g., pretest), which removes all variance explainable by the covariate and leaves all remaining (residual) variance to be predicted by group membership or other variables. As described by Tracz et al. (2005), “This residualized or adjusted dependent variable is no longer the same as the original dependent variable” (p. 17). Similarly, partialled independent variables used in analyses such as hierarchical multiple regression only retain variance that is unique and not in common with others in a set. As described by Lynam et al. (2006), "The crux of the problem identified by almost all critics is that it is difficult to know what construct an independent variable represents once the variance shared with other independent variables is removed" (p. 329).

Validity of a Variable Residualized by a Single Covariate

In the case of a single covariate, the validity of a residualized variable (ρ_{RZ}) can be assessed by computing a part correlation [$\rho_{Z(Y.X)}$], where the criterion variable (Z) is correlated to the unresidualized variable (Y) after removing the effects of the covariate (X) from Y as denoted by Williams and Zimmerman (1982):

$$\rho_{RZ} = \frac{\rho_{YZ} - \rho_{XY}\rho_{XZ}}{\sqrt{(1-\rho_{XY}^2)}} \quad (1)$$

Williams and Zimmerman's formula assumes population data with infallible measures and is therefore practically limited to true score correlations or correlations between observed scores with the statistical assumption of error-free data. Nonetheless, applying their formula across a wide range of correlations provides fundamental insights into understanding the validity of a variable that has been residualized by a single covariate (see Figure 1).

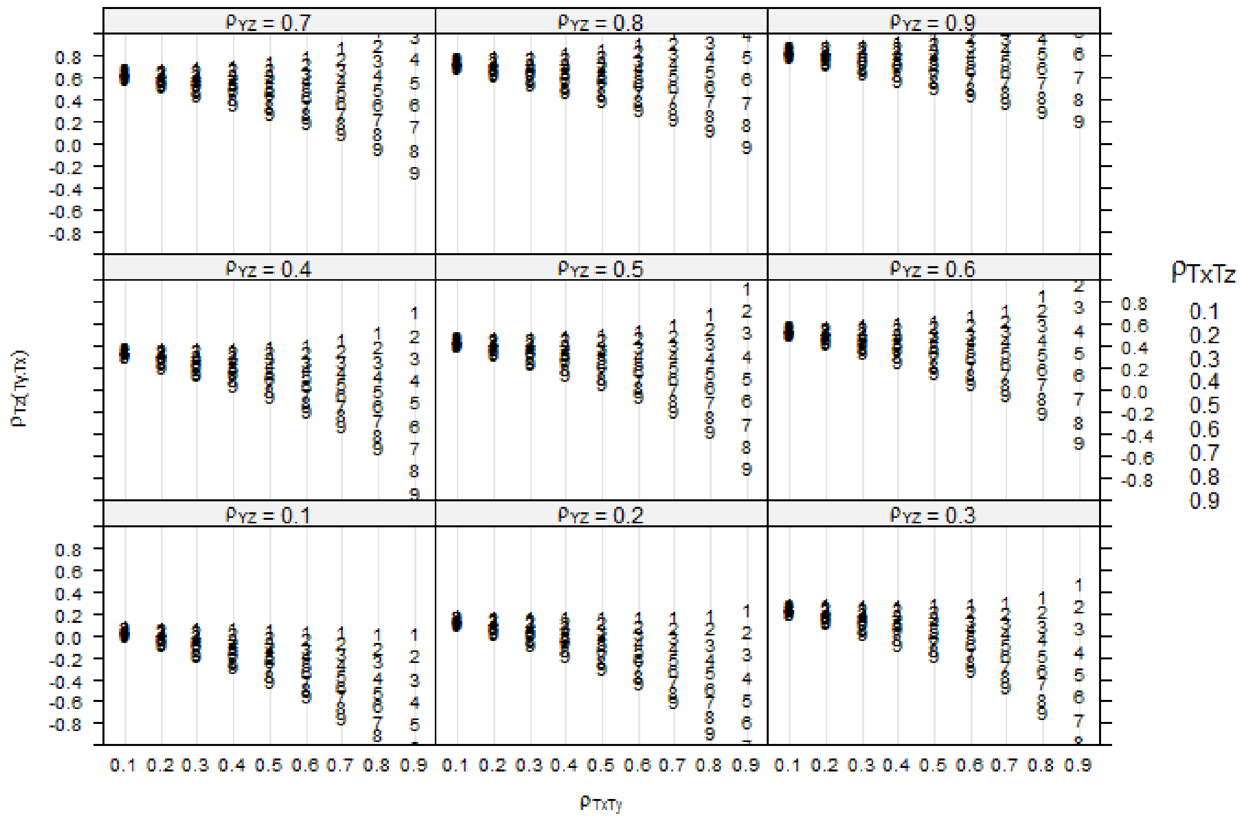

Figure 1. $\rho_{T_Z(T_Y.T_X)}$ as a function of $\rho_{T_XT_Y}$, $\rho_{T_XT_Z}$, and $\rho_{T_YT_Z}$. Decimals for $\rho_{T_XT_Z}$ values are omitted.

As illustrated in Figure 1, the validity of a residualized variable is not substantially different than the validity of the unresidualized variable given small amounts of residualization ($\rho_{T_XT_Y}$= .10). However, with higher amounts of residualization, the validity of the residualized variable is impacted by the amount of multicollinearity between the covariate and criterion variable such that the impact of the covariate-criterion relationship increases as the magnitude of the XY relationship increases. While higher degrees of multicollearity result in lower residualized coefficients, residualized validity coefficients are not always smaller than unresidualized validity coefficients. When a covariate serves to suppress irrelevant variance in the unresidualized variable, the validity of the residualized variable will be greater than the validity of the unresidualized variable.

Contemporary Validity Studies

Although the reporting of residualized validity coefficients is not commonplace in contemporary behavioral science literature, at least two studies have recently considered the validity of residualized variables. While both studies took a nomological approach to establishing the validity of a variable that had been residualized by a single covariate, they differed in their focus and analytic approach. Nimon and Henson (2010) considered the validity of a dependent variable that had been residualized by a procedure such as ANCOVA and compared unresidualized validity coefficients to residualized validity coefficients. Lynam et al. (2006) considered the validity of several independent variables that had been residualized through multiple regression and compared unstandardized regression coefficients.

Nimon and Henson (2010)

Using an established nomological network for the Beck's Depression Inventory-II (BDI-II; Beck, 1996) and the pretest-posttest design as a general framework, Nimon and Henson (2010) examined the validity of BDI-II posttest scores (Y) before and after they were regressed by BDI pretest scores (X), for the case r_{XY} = .79. They considered scores from four scales as criterion variables (Z): (a) Beck Anxiety Inventory (Beck, 1990), (b) Center for Epidemiologic Studies Depression Scale (Radloff, 1977), (c) Solitude subscale from Beck's Sociotropy-

Autonomy Scale (Clark, Steer, Beck, & Ross, 1995), and (d) Trait scale of the State Trait Anxiety Inventory (Spielberger, 1977). The dependent-criterion variable correlations ($r_{YZ[1,4]}$) reported were .65, .77, .75, and .37 respectively.

Residualized validity coefficients were computed via Eq. 1. The resulting residualized validity coefficients ($r_{RZ[1,4]}$) reported were .33, .45, .35, and .17 respectively. Comparing $r_{YZ[1,4]}$ to $r_{RZ[1,4]}$, the authors concluded that the residualized dependent variable was measuring a different construct than the unresidualized dependent variable.

Two comments regarding Nimon and Henson's (2010) study are pertinent to the present study. First, the degree of overlap between the covariate and four criterion variables approximately mirrored the relationship between the dependent and criterion variables. Second, the analyses were based on the statistical assumption that the variables were measured without error. However, given prior reliabilities reported (e.g., Beck, 1996; Beck, 1990; Radloff, 1977; Spielberger, 1977; Clark et al., 1995), such an assumption seems untenable.

Lynam, Hoyle, and Newman (2006)

Using agression and psychopathy measures, Lynam et al. (2006) examined the relations (i.e., slopes) of three pairs of subscale scores, before and after partialling a relevant subscale, to fourteen variables comprising a theoretical nomological network. With each of the three measures containing two subscales, each subscale served to residualize the other subscale within a measure. Given that the context of the analysis was multiple regression, each subscale also served to residualize the nomological variable with which the residualized subscale was being related. Therefore, instead of comparing zero-order and part correlations as in Nimon and

Henson (2010), Lynam et al. compared zero-order and partial relations to determine if raw and residualized independent variables were measuring different constructs.

Unstandardized regression coefficients were used to compare the validity of unresidualized and residualized measures. Coefficients for unresidualized measures were taken from regression analyses in which the given subscale was the only predictor. Coefficients for residualized measures were taken from regression analyses in which both subscales were entered as predictors. Statistically significant changes between the unresidualized and residualized coefficients were determined by examining the differences as a function of the standard error for the third variable effect (see MacKinnon, Krull, & Lockwood, 2000). For the first measure where the correlation between scales score was .80, 60.7% (17) of the 28 coefficients underwent statistically significant change following partialling. Of the 17, 10 (58.8%) represented significant decreases, 3 (17.7%) represented significant increases, and 4 (23.5%) represented changes in direction. For the second and third measures where the correlation between scale scores was .50, 71% of the 56 coefficients underwent significant change following partialling. Of the 40, 25 (62.5%) represented significant decreases, 10 (25%) represented significant increases, and 5 (12.5%) represented changes in direction.

Differences between the nomological networks of the unresidualized and residualized scores were assessed by computing similarity coefficients. Similarity coefficients were operationalized as intraclass correlation coefficients that indexed the similarity between groups of unstandardized regression coefficients. Higher degrees of similarity were reported for within-measure unresidualized subscale scores (Intra Class Correlations (ICCs) = .98, .94, .51) than within-measure residualized subscale scores (ICCs = .02, .53, -.46).

Based on statistically significant differences in pairs of unstandardized regression coefficients and low ICCs for within-measure residualized subscale scores, the authors concluded the unresidualized and residualized independent variables were measuring different constructs. They further indicated the three scales provided likely candidates for high levels of discrepancy given that within measure correlations ranged between .5 and .8 and coefficient alphas ranged between .47 and .89.

Three comments regarding Lynam's et al. (2006) study are pertinent to the present study. First, the process of using unstandardized regression coefficients to examine the construct validity of residualized variables warrants caution. Given that unstandardized regression coefficients are based on residualized dependent *and* residualized independent variables, using partial relations to validate a residualized independent variable seems to introduce additional interpretation issues given the construct of a residualized dependent variable is just as questionable as the construct of a residualized independent variable. Second, although the authors considered the correlation between the independent variable (X1) and its covariate (X2) in explaining differences between zero-order and partial relations, it appears they neglected to consider other factors including the relationship between the covariate and dependent variable and the relationship between the dependent variable and the independent variable. Figure 2 represents standardized regression coefficients for a number of conditions consistent with Lynam's et al. study (i.e., r_{X1X2}=.50, .80). The figure demonstrates that deviances between zero-order and partial relations are impacted by more than the relationship between the independent variable and its covariate. One can also see that by mapping Lynam's et al. findings to Figure 2 the study included conditions where the covariate served to suppress irrelevant variance in the independent variable being residualized. Third, although the authors considered the reliability of the independent variables in explaining differences between zero-order and partial relations, the estimated unresidualized and residualized relations were based on the statistical assumption that the data were error-free.

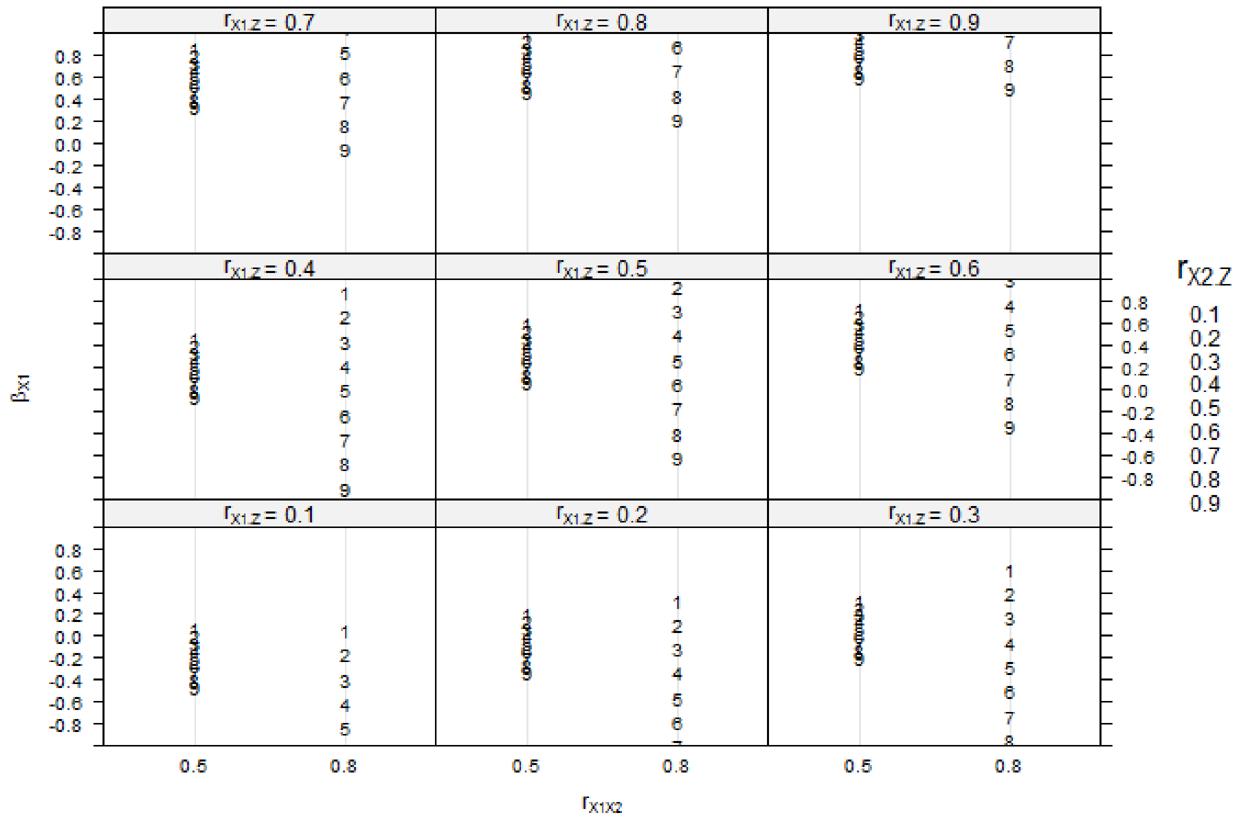

Figure 2. β_{X1} when Y is regressed on X1 and X2 as a function of $r_{X1.X2}$, $r_{X2.Z}$, and $r_{X1.Z}$. Decimals for $r_{X2.Z}$, values are omitted.

Conclusions from Contemporary Validity Studies

The validity studies reviewed inform the present study in several ways. First, they identify a void in understanding how error impacts the validity of residualized variables. Quantifying the error in sample-based unresidualized validity coefficients (i.e., zero-order correlations) across a number of conditions has recently been examined by Zimmerman (2007). However, extensions to residualized validity coefficients have yet to be considered; consequently, this is the focus of the present study. Second, the present study considers the

Williams and Zimmerman (1982) formula the appropriate foundation for computing the validity of a dependent or independent variable that has been residualized by a single covariate. Although Lynam et al. (2006) employed a partial relations approach using unstandardized regression coefficients, support for correlating residualized independent variables to nonresidualized nomological variables can be found in Winne (1983). Not only does the latter approach eliminate the confounding factor of a residualized nomological variable, it also yields correlation coefficients that are more readily interpretable than unstandardized regression coefficients. Third, the studies reviewed provide insight into the extent of residualization and the multicollinearity between the criterion and nomological variable that should be considered in the present study. The Lynam et al. study indicates a need for a range of multicollinearity that is broader than simply mirroring the level of residualization reported in Nimon and Henson (2010).

Sources of Error in Residualized Validity Coefficients

There are several potential sources of error that can bias residualized validity coefficients. In population data, error is introduced when the assumptions of perfect reliability and experimental independence are not met. In general, unreliability of data stems from the limited psychometric capacity of inventories (e.g., Wechsler Adjust Intelligence Scale) to perfectly assess underlying constructs (e.g., intelligence). Experimental dependence results when common factor or systemic error occurs in a repeated measures design. For example, when individuals are given parallel forms of a measure that has a common response bias, the assumption of experimental independence is violated and correlation between errors results. In the case of population data, the bias associated with measurement and correlated error in residualized validity coefficients can be corrected by applying Zimmerman and Williams' (1997) general correction for attenuation at the population level to each observed-score correlation in Eq. 1:

$$\rho(T_X T_Y) = \frac{\rho_{XY}}{\sqrt{\rho_{XX'}\rho_{YY'}}} - \frac{\rho(E_X E_Y)\sqrt{(1-\rho_{XX'})(1-\rho_{YY'}}}{\sqrt{\rho_{XX'}\rho_{YY'}}} \tag{2}$$

Zimmerman and Williams's formula is based on the expected-value true-score theory without the assumption of experimental independence (Winne & Belfry, 1982). Winne and Belfry demonstrated that when the term for correlated error scores is subtracted from the classical correction for attenuation, Zimmerman and Williams's formula result yields lower estimates when error scores are positively correlated, and higher estimates when error scores are negatively correlated, as compared to estimates from Spearman's (1904) formula. Figure 3 demonstrates the magnitude of the term for correlated error scores across a range of correlated error and reliability parameters. The role correlated error plays in attenuating or disattenuating a correlation coefficient increases as reliability decreases. Note that when reliability is .5, the term for correlated error scores equals the magnitude of the correlated error.

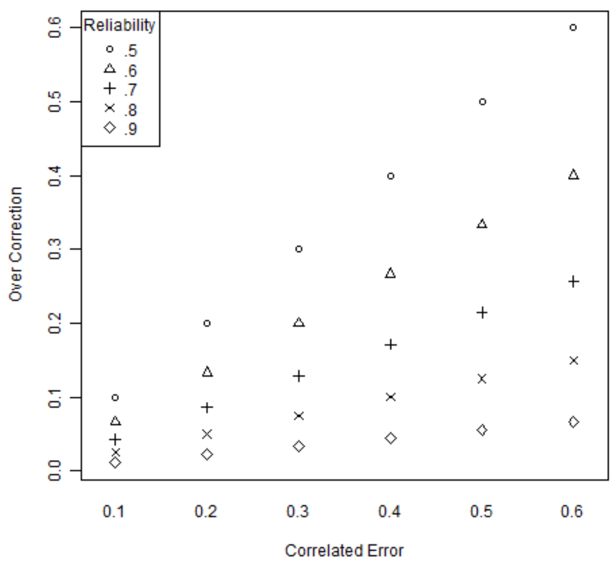

Figure 3. Overcorrection in Spearman's (1904) correlation as a function of correlated error $\rho(E_X E_Y)$ and reliability ($\rho_{XX} = \rho_{YY}$).

While applying Eq. 2 to the correlations in Eq. 1 accounts for error in population-based residualized variable validity coefficients, it does not consider all sources of error in samples.

Sample-based residualized variable validity estimates have two additional sources of error. In addition to sample error associated with each of the correlations in Eq. 1 (cf. Wang & Thompson, 2007), spurious correlations between true and error scores introduce error into sample-based residualized validity coefficients. The error imposed by correlations between true and error scores is unique to samples given that classical test theory regards an individual's observed score as the sum of a true and error score. Recently, Zimmerman (2007) adapted the general correction for attenuation at the population level formula (Eq. 2) for sample data:

$$r_{T_X T_Y} = \left(\frac{r_{XY} - r_{E_X E_Y}\sqrt{1 - r_{XX'}}\sqrt{1 - r_{YY'}}}{\sqrt{r_{XX'} r_{YY'}}} - \frac{r_{T_X E_Y}\sqrt{1 - r_{XX'}}}{\sqrt{r_{XX'}}} - \frac{r_{T_Y E_X}\sqrt{1 - r_{YY'}}}{\sqrt{r_{YY'}}}\right) \quad (3)$$

Eq. 3 not only corrects for unique and common measurement error, it also corrects for correlations between true and error scores. Applying Eq. 3 to the correlations in Eq. 1 provides a theoretical model for sample-based residualized validity coefficients. Although such a derivation of Eq. 1 may not be practical if only observed scores are available, it clarifies the role error plays in residualized validity coefficients. By means of simulations, all the quantities in such a derivation can be given explicit values, and the resulting effect of error can be observed across multiple study conditions. Such simulations build on the work of Zimmerman (2007) who investigated the effect of correlated error in bivariate sample correlations.

Effect of Correlated Error in Bivariate Sample Correlations

Using Monte Carlo methods, Zimmerman (2007) examined properties of the Spearman (1904) correlation formula under conditions in which correlations between error scores exist as a population parameter as well as when correlated errors occur by chance in random samples. Zimmerman also investigated the properties of the sample form of the correction for attenuation formula that models correlated errors (Eq. 3) under simulated correlated error conditions and compared the results to observed score correlations and Spearman correlations.

Zimmerman found that while increasing sample size reduces the values of correlations between errors scores that arise randomly in samples, increasing sample size does not reduce the effect of correlated error at the population level. Confirming the relationships depicted in Figure 3, Zimmerman found that the effect of correlated error on Spearman's (1904) correlation as applied to sample data was close to zero when $\rho(E_X E_Y)$ was equal to zero and became larger as $\rho(E_X E_Y)$ increased and ρ_{XX} decreased independent of sample size. He also demonstrated the bias and precision of estimating the effect of correlated error on Spearman's correlation using the right-hand term of Eq. 2 across a number of study conditions ($\rho_{T_X T_Y}$ = .60; ρ_{XX} = .75, .95; $\rho_{E_Y E_Z}$= 0, .10, .20; *n* = 25, 100, 400). His findings indicated the estimated discrepancy becomes more accurate and precise as sample size increases. With a sample size of 400, there was virtually no difference between the simulated and predicted results for the conditions reported. When considering variability, his study showed the effect of correlated error was less dependent on $\rho(E_X E_Y)$ and ρ_{XX} and more dependent on sample size. Consistent with the distribution of correlations among error score components, the standard deviations of the effect were close to $1/\sqrt{n-1}$ for the study conditions reported.

Zimmerman simulated a number of study conditions ($\rho_{T_X T_Y}$ = .70, .90; ρ_{XX}=.70, .95; $\rho_{E_Y E_Z}$= 0, .10, .20; *n* = 25, 50) to investigate the properties of the sample form of the correction for attenuation formula that models correlated errors (Eq. 3). Means and standard deviations of sample correlations between observed scores were computed and compared to correlations based on Spearman's (1904) formula and Eq. 3. Across the study conditions reported, both corrected correlations came closer to true score population correlations than observed score correlations.

Within the limits of rounding error, the two corrected correlations performed comparably when being compared to true score correlations for the condition of uncorrelated error. However, correlations based on Eq. 3 had less variability than Spearman correlations. In the case of correlated error, correlations based on Eq. 3 had less bias and were more precise than correlations based on Spearman's formula.

Purpose

This study extends research concerning the validity of residualized variables to determine which correction formulas provide the most precise and least biased estimates, in the presence of correlated error in population and sample data. Such information is critical if researchers are to parsimoniously and accurately assess the validity of residualized variables. Therefore, this study investigated the bias and precision of sample-based residualized variable validity estimates (r_{RZ}) under conditions where correlated errors exist in the population as well as by chance in random samples. The estimates were compared to true score population values ($\rho_{\mathrm{T_R T_Z}}$) to identify cases in which r_{RZ} estimates lead to incorrect conclusions regarding the true magnitude of the relationship between the criterion variable (Z) and the residualized variable ($Y - \hat{Y}$). Recognizing the value of assessing the convergent validity of a residualized variable, the study focused on conditions in which validity values based on population true scores were approximately between .3 and .9.

This study considers four r_{RZ} estimates. The first is an uncorrected estimate based on Williams and Zimmerman's formula (1982) as applied to sample data:

$$r_{RZ:C0} = \frac{r_{YZ} - r_{XY}r_{XZ}}{\sqrt{(1 - r_{XY}^2)}} \qquad \text{C0}$$

As noted by Williams and Zimmerman, the formula is a simple part correlation [i.e., $r_{Z(Y.X)}$], where Z is correlated to Y after removing the effects of X from Y.

The second applies Spearman's (1904) correction for attenuation based on reliability coefficients formula to each observed-score correlation in Eq. C0. Assuming independence of errors, Spearman's formula corrects for uncorrelated measurement error by dividing the observed score coefficient by the square root of the product of reliabilities (Wetcher-Hendricks, 2006). Spearman's correction as applied to $r_{RZ:C0}$ takes the form:

$$r_{RZ:C1} = \frac{r_{XX'}r_{YZ} - r_{XY}r_{XZ}}{\sqrt{r_{ZZ'}r_{XX'}}\sqrt{r_{YY'}r_{XX'} - r_{XY}^2}} \quad \text{C1}$$

The third applies Zimmerman and Williams' (1997) general correction formula for attenuation at the population level to each observed-score correlation in Eq. C0. In contrast to Spearman's (1904) formula, Zimmerman and Williams' formula does not assume independence of errors. In addition to dividing the observed score coefficient by the square root of the product of reliabilities, Zimmerman and Williams' formula corrects for correlated error. Zimmerman (2007) indicated the formula cannot be applied to sample data because it does not correct for correlations among true and error score components that can occur in sample data. It is included in the present study based on prior work (Wetcher-Hendricks, 2006) that adapted Zimmerman and Williams' formula to estimate corrected part correlations on sample data. Zimmerman and Williams' correction for attenuation as applied to $r_{RZ:C0}$takes the form:

$$r_{RZ:C2} = \frac{\begin{array}{c} r_{XX'}(r_{YZ} - r_{E_YE_Z}\sqrt{1 - r_{YY'}}\sqrt{1 - r_{ZZ'}}) \\ -(r_{XY} - r_{E_XE_Y}\sqrt{1 - r_{XX'}}\sqrt{1 - r_{YY'}}) \\ *\,(r_{XZ} - r_{E_XE_Z}\sqrt{1 - r_{XX'}}\sqrt{1 - r_{ZZ'}}) \end{array}}{\sqrt{r_{ZZ'}r_{XX'}}\sqrt{r_{YY'}r_{XX'} - (r_{XY} - r_{E_XE_Y}\sqrt{1 - r_{XX'}}\sqrt{1 - r_{YY'}})^2}} \quad \text{C2}$$

The fourth applies Zimmerman's (2007) general correction for attenuation at the sample level to each observed-score correlation in Eq. C0. In addition to correcting for measurement error without assuming independence of errors, the formula corrects for spurious correlations among true and error score components that can occur in sample data. Zimmerman's correction for attenuation as applied to $r_{RZ:C0}$takes the form:

$$r_{RZ:C3} = \frac{\begin{array}{c}\frac{r_{YZ} - r_{E_Y E_Z}\sqrt{e_{YY'}}\sqrt{e_{ZZ'}}}{\sqrt{r_{YY'}r_{ZZ'}}} - \frac{r_{T_Y E_Z}\sqrt{e_{YY'}}}{\sqrt{r_{YY'}}} - \frac{r_{T_Z E_Y}\sqrt{e_{ZZ'}}}{\sqrt{r_{ZZ'}}} \\ -\left(\frac{r_{XY} - r_{E_X E_Y}\sqrt{e_{XX'}}\sqrt{e_{YY'}}}{\sqrt{r_{XX'}r_{YY'}}} - \frac{r_{T_X E_Y}\sqrt{e_{XX'}}}{\sqrt{r_{XX'}}} - \frac{r_{T_Y E_X}\sqrt{e_{YY'}}}{\sqrt{r_{YY'}}}\right) \\ * \left(\frac{r_{XZ} - r_{E_X E_Z}\sqrt{e_{XX'}}\sqrt{e_{ZZ'}}}{\sqrt{r_{XX'}r_{ZZ'}}} - \frac{r_{T_X E_Z}\sqrt{e_{XX'}}}{\sqrt{r_{XX'}}} - \frac{r_{T_Z E_X}\sqrt{e_{ZZ'}}}{\sqrt{r_{ZZ'}}}\right)\end{array}}{\sqrt{1 - \left(\frac{r_{XY} - r_{E_X E_Y}\sqrt{e_{XX'}}\sqrt{e_{YY'}}}{\sqrt{r_{XX'}r_{YY'}}} - \frac{r_{T_X E_Y}\sqrt{e_{XX'}}}{\sqrt{r_{XX'}}} - \frac{r_{T_Y E_X}\sqrt{e_{YY'}}}{\sqrt{r_{YY'}}}\right)^2}} \qquad \text{C3}$$

Zimmerman (2007) noted that equations based on true and error scores could not be used for practical purpose since only observed scores are available. Whereas true score equivalents and error scores are usually unknown without the aid of structural equation modeling on item level data, studies conducted by Wetcher-Hendricks (2006) are notable exceptions. Wetcher-Hendricks used historical data (weather records and baseball statistics) as true score equivalents and derived error scores by subtracting true scores from sample data. Her study demonstrated that part correlations based on the equivalent of formula C2 were closer to true score part correlations when compared to part correlations based on formulas C0 and C1. Noting that covariances may also exist between true scores and error scores, she identified the need for future research to expand formula C2 to make it applicable to situations in which covariances exist between error and true scores. Formula C3 is such a derivation. Formula C3 is therefore included in the current study to examine the precision and variability of residualized variable validity estimates after they have been corrected for measurement error, correlations between errors, and spurious correlations among true and error score components.

Methods

A program was written in R to generate simulated data across a set of study conditions to examine the bias and precision of a set of residualized variable validity estimates. True score population validity coefficients were compared to uncorrected estimates based on formula C0 in addition to the corrected estimates based on formulas C1, C2, and C3. Bias and precision were compared by study condition (see Table 1).

Table 1

Study Conditions

$\rho_{T_R T_Z}$					Correlated	Error Condition			Sample
$\rho_{T_R T_Z}$:	$\rho_{T_X T_Y}$	$\rho_{T_X T_Z}$	$\rho_{T_Y T_Z}$	Reliability[a]	Error[b]	$\rho_{E_X E_Y}$	$\rho_{E_X E_Z}$	$\rho_{E_Y E_Z}$	*n*
.27	.8	.8	.8	.6	.1	Yes	No	No	25
.53	.8	.6	.8	.7	.3	No	Yes	No	100
.93	.8	.3	.8	.8	.6	No	No	Yes	400
.40	.6	.8	.8	.9		Yes	Yes	Yes	
.55	.6	.6	.8						
.78	.6	.3	.8						
.59	.3	.8	.8						
.65	.3	.6	.8						
.74	.3	.3	.8						

Note. [a]$\rho_{XX'} = \rho_{YY'} = \rho_{ZZ'}$. [b]$\rho_{E_X E_Y} = \rho_{E_X E_Z} = \rho_{E_Y E_Z}$.

Residualized Variable Validity Coefficients

Consistent with the aim of the study, the study simulated conditions of $\rho_{T_R T_Z}$ range between .27 and .93. Nine true score residualized variable validity coefficients were simulated based on a moderately high level of convergent validity between the unresidualized and criterion variable ($\rho_{T_Y T_Z}$ level of .8), three levels of covariate adjustment ($\rho_{T_X T_Y}$ levels of .3, .6, and .8), and three levels of multicollinearity between the covariate and criterion variables ($\rho_{T_Y T_Z}$ levels of .3, .6, and .8). Correlations of .3, .6, and .8 approximate an equidistance of shared variance with the unresidualized variable (9%, 36%, 64%) and represent a broad range of covariate adjustment and multicollinearity.

Reliability

Four levels of score reliability were modeled (.9, .8, .7, .6). A reliability value of .90 is considered a cutoff value when scores are used for important clinical and/or educational decisions; .8 is considered a minimally accepted value for general research (Henson, 2001). Reliabilities of .7 and .6 were also considered given the former is often considered a lower bound of acceptable reliability (Henson, 2001) and the latter is not unusual in early research (Zimmerman, 2007). For the purpose of parsimony, the reliability of the unresidualized, covariate, and criterion variables were modeled as equal.

Correlated Error and Error Condition

Three levels of correlated errors were modeled (.1, .3, .6). The three levels are respectively considered weak, moderate, and strong levels of correlated error (Herting, 2002). Four correlated error conditions were modeled: (a) correlated error between covariate (X) and unresidualized variable (Y) only, (b) correlated error between covariate (X) and criterion variable (Z) only, (c) correlated error between unresidualized (Y) and criterion variable (Z) only, (d) correlated error between X and Y, X and Z, and Y and Z.

Sample Size

Three levels of sample size were simulated (*n*s = 25, 100, 400). The levels were selected for three reasons. First, the three levels are consistent with Zimmerman (2007) and provide the potential for comparisons to be made between the results of the present study and Zimmerman's work. Second, the sample sizes represent a broad range of theoretical distributions of sample correlation coefficients among error score components. Based on sampling theory (Fisher, 1915) and Zimmerman's (2007) simulation results, the distribution of sample correlation coefficients among error score components (e.g., $r_{E_X E_Y}$, $r_{T_X E_X}$, $r_{T_Y E_Y}$, $r_{T_X E_Y}$, $r_{T_Y E_X}$), is expected to yield approximate standard deviations of $1/\sqrt{n-1}$. For this study, it is anticipated that the correlations among error components will yield standard deviations of ~.20, ~.10, ~.05 as a function of increasing sample size. Third, the median sample size of 100 is consistent with the median subject-to-variable ratio in social science research as reported by Kieffer, Reese, and Thompson (2001).

Simulation

The fully nested combination of the study's conditions is 1,296 (i.e., 9 x 4 x 3 x 4 x 3 or 432 [9 x 4 x 3 x 4] populations x 3) Code was written in R (see appendix) to generate the implied covariance matrix for each of the study's conditions. Values for the covariance matrix were based on study parameters as indicated in Table 2. In order to not confound the magnitude of error variance with the magnitude of observed score variance, the study was designed such that the variance of population observed scores was 1.0 (cf. Corder-Bolz, 1978). Consistent with classical test theory, correlations between population true and error scores were designed to be 0 (cf. Croker & Algina, 1986).

Table 2

Population Covariance Matrix

	T_X	T_Y	T_Z	E_X	E_Y	E_Z
T_X	ρ_{XX}	$\rho_{T_XT_Y}\sigma_{T_X}\sigma_{T_Y}$	$\rho_{T_XT_Z}\sigma_{T_X}\sigma_{T_Z}$	0	0	0
T_Y	$\rho_{T_XT_Y}\sigma_{T_X}\sigma_{T_Y}$	ρ_{YY}	$\rho_{T_YT_Z}\sigma_{T_Y}\sigma_{T_Z}$	0	0	0
T_Z	$\rho_{T_XT_Z}\sigma_{T_X}\sigma_{T_Z}$	$\rho_{T_YT_Z}\sigma_{T_Y}\sigma_{T_Z}$	ρ_{ZZ}	0	0	0
E_X	0	0	0	$1 - \rho_{XX}$	$\rho_{E_XE_Y}\sigma_{E_X}\sigma_{E_Y}$	$\rho_{E_XE_Z}\sigma_{E_X}\sigma_{E_Z}$
E_Y	0	0	0	$\rho_{E_XE_Y}\sigma_{E_X}\sigma_{E_Y}$	$1 - \rho_{YY}$	$\rho_{E_YE_Z}\sigma_{E_Y}\sigma_{E_Z}$
E_Z	0	0	0	$\rho_{E_XE_Z}\sigma_{E_X}\sigma_{E_Z}$	$\rho_{E_YE_Z}\sigma_{E_Y}\sigma_{E_Z}$	$1 - \rho_{ZZ}$

The covariance matrix was passed to the function `mvrnorm` (Venables & Ripley, 2000) to generate population data. Resulting true and error scores were added to form population observed scores. To minimize the standard error of simulation, a population the size of 5,000 x *n* (i.e., 25, 100, 400) was generated for each of the study's conditions (cf. Wang & Thompson, 2007). The resulting dataset was randomly divided into 5,000 samples in preparation for analyses.

Analyses

Across the 1,296 conditions, validity estimates based on formulas C0, C1, C2, and C3 were computed for each of the samples. Bias was computed by subtracting the true score population value $\rho_{T_RT_Z}$from the sample estimate. Positive discrepancies reflect validity estimates that overestimate population parameters, while negative discrepancies reflect sample values that underestimate population parameters (cf. Wang & Thompson, 2007). Precision was determined by calculating the Standard Deviation (SD) of the four validity estimates across the 1,296 study conditions.

An analysis of variance (ANOVA) was used to provide information regarding the effect of each condition on the mean bias and mean precision of the validity estimates. The independent variables included: (a) level of covariate adjustment, (b) level of multicollinearity between the covariate and criterion variable, (c) level of reliability, (d) level of correlated error, (e) correlated error conditions, and (f) level of sample size. Estimated marginal means (EMMs) and η^2 were computed using main and interaction effects in the simulation design to predict bias and precision in the estimates.

Results

Analyses of the 432 (9 x 4 x 3 x 4) populations revealed the populations yielded the desirable population parameters. Analyses of the correlations between true scores and error scores (e.g., $r_{T_X E_X}$, $r_{T_X E_Y}$, $r_{T_X E_Z}$) revealed these so-called nuisance correlations were normally distributed with a *M* of 0 and a *SD* of $1/\sqrt{n-1}$, as expected (cf. Charles, 2005; Zimmerman, 2007). Analyses of the differences between the $r_{RZ:C3}$ validity estimates and the corresponding $r_{T_R T_Z}$ values had a *M* and *SD* of 0, as expected (cf. Zimmerman). As authors (e.g., Nimon, Zientek, & Henson, 2012; Charles, 2005; Zimmerman, 2007) have noted that Spearman's correction for attenuation sometimes produces values greater than 1.00, the number of values for the four validity estimates were also examined. In the case of $r_{RZ:C0}$and $r_{RZ:C3}$, each of the 1,296 study conditions yielded the full number of estimates (i.e., 5,000). However, in the case of $r_{RZ:C1}$and $r_{RZ:C2}$, a number of impossible values (i.e., NAs) were generated when the correction for attenuation for r_{XY} produced a value greater than 1.00. The NA values resulted in unequal cell sizes across the study conditions and an in-balanced design, as some study conditions resulted in no validity estimates.

For $r_{RZ:C2}$, the study conditions that appeared to affect the number of NA values were *n,* $\rho_{T_R T_Z}$, reliability, and related interactions (see Table 3). The EMMs for the main effects showed that across the study conditions, the greatest number of NA values involved conditions where: *n* = 25, reliability = .60, and $\rho_{T_X T_Y}$ = .80 (see Table 4). In analyzing the main effects within the context of the related interactions, the EMMs for the *n* x $\rho_{T_R T_Z}$ x reliability interaction produced findings consistent with the main effects. The interaction EMMs also showed that when *n* = 25 and $\rho_{T_X T_Y}$ = .80, the range of number of NA values dropped nearly in half when reliability increased from .60 to .70, and even further when reliability increased to .80. The remaining EMMs revealed little to no occurrence of NA values (see Figure 4).

For $r_{RZ:C1}$, the study conditions that appeared to affect the number of NA values were reliability, $\rho_{T_R T_Z}$, correlated error location (CEloc), level of correlated error (CElev), *n*, and related interactions (see Table 3). The EMMs for the main effects show that across the study conditions, the greatest number of NA values involved conditions where: reliability = .60, $\rho_{T_X T_Y}$ = .80, correlated error existed between E_X and E_Y (either solely or when all error components were similarly correlated), CElev = .60, and *n* = 25 (see Table 4). However, these main effects should be interpreted within the context of related interaction effects (see Figures 5a, 5b, 5c). According to the interaction effects, the largest number of NA values occurred in study conditions where reliability = .60, $\rho_{T_X T_Y}$ = .80, correlated error existed between E_X and E_Y, CElev = .60, and *n* = 400. These study conditions resulted in no valid estimates being provided for $r_{RZ:C1}$ (see Figure 5c). Therefore, in order to provide a balanced design, analyses on the bias and precision of the four validity estimates were modeled omitting the study condition where CElev = .60, resulting in a fully crossed design of 864 study conditions (9 x 4 x 2 x 4 x 3).

Table 3

Statistics for Number of $r_{RZ:C1}$ *and* $r_{RZ:C2}$ *NA Values within the 1296 Study Conditions*

Statistic/source	$r_{RZ:C1}$	$r_{RZ:C2}$
Descriptive Statistics		
Median	0.00	0.00
M	298.85	28.79
SD	850.03	96.14
Skewness	3.79	4.42
Kurtosis	14.68	20.35
η^2 for study condition factors		
$\rho_{TR.TZ}$	10.73%	9.84%
Reliability (rel)	10.98%	8.09%
CE Level (CElev)	6.10%	0.28%
CE Location (CEloc)	7.60%	0.57%
n	1.46%	17.01%
$\rho_{TR.TZ}$ x rel	7.64%	7.70%
$\rho_{TR.TZ}$ x CElev	4.44%	0.28%
rel x CElev	4.87%	0.22%
$\rho_{TR.TZ}$ x CEloc	6.42%	0.52%
rel x CEloc	6.78%	0.38%
CElev x CEloc	6.09%	0.28%
$\rho_{TR.TZ}$ x *n*	0.46%	18.30%
rel x *n*	0.60%	14.88%
CElev x *n*	0.03%	0.52%
CEloc x *n*	0.06%	1.03%
$\rho_{TR.TZ}$ x rel x CElev	3.79%	0.20%
$\rho_{TR.TZ}$ x rel x CEloc	4.63%	0.29%
$\rho_{TR.TZ}$ x CElev x CEloc	4.44%	0.30%
rel x CElev x CEloc	4.86%	0.23%
$\rho_{TR.TZ}$ x rel x *n*	0.21%	13.74%
		(table continues)
$\rho_{TR.TZ}$ x CElev x *n*	0.26%	0.50%
rel x CElev x *n*	0.25%	0.41%
$\rho_{TR.TZ}$ x CEloc x *n*	0.10%	0.90%
rel x CEloc x n	0.08%	0.65%
CElev x CEloc x *n*	0.03%	0.53%
$\rho_{TR.TZ}$ x rel x CElev x CEloc	3.80%	0.21%
$\rho_{TR.TZ}$ x rel x CElev x *n*	1.08%	0.36%
$\rho_{TR.TZ}$ x rel x CEloc x *n*	0.61%	0.47%
$\rho_{TR.TZ}$ x CElev x CEloc x *n*	0.27%	0.54%
rel x CElev x CEloc x n	0.25%	0.41%
$\rho_{TR.TZ}$ x rel x CElev x CEloc x *n*	1.08%	0.37%

Note. Because there is only one case (i.e., the number of estimates) per cell, the factorial analysis of variance degrees of freedom error is 0, and the η^2 values sum to 100%, within rounding error.

Table 4

Estimated Marginal Means for Main Effects and Theoretical Distribution (TD) of r

	NAs[a]		Bias[b]		Precision[b]				
Factor/level	$r_{RZ:C1}$	$r_{RZ:C2}$	$r_{RZ:C0}$	$r_{RZ:C1}$	$r_{RZ:C0}$	$r_{RZ:C1}$	$r_{RZ:C2}$	$r_{RZ:C3}$	TD
$\rho_{TR.TZ}$									
.27	682.57	70.10	0.02	-0.01	0.08	0.28	0.18	0.07	
.40	188.25	13.72	-0.03	-0.01	0.08	0.18	0.13	0.06	
.53	680.63	72.08	-0.12	0.05	0.09	0.33	0.19	0.07	
.55	188.36	13.71	-0.10	0.01	0.08	0.19	0.14	0.07	
.59	27.34	1.95	-0.11	0.00	0.07	0.14	0.10	0.06	
.65	26.35	1.80	-0.13	0.01	0.08	0.13	0.11	0.06	
.74	27.25	1.65	-0.17	0.02	0.08	0.13	0.10	0.06	
.78	188.32	13.74	-0.20	0.05	0.08	0.22	0.12	0.05	
.93	680.61	70.39	-0.33	0.16	0.07	0.49	0.17	0.04	
rel									
.60	724.10	71.58	-0.20	0.06	0.09	0.43	0.21	0.06	
.70	376.98	32.81	-0.15	0.04	0.08	0.26	0.15	0.06	
.80	89.44	10.08	-0.11	0.02	0.08	0.16	0.11	0.06	
.90	4.88	0.69	-0.06	0.00	0.07	0.09	0.08	0.06	
CElev									
.10	89.56	34.54	-0.14	0.02	0.08	0.20	0.14	0.06	
.30	221.33	29.66	-0.13	0.05	0.08	0.26	0.13	0.06	
.60	585.67	22.18	NA	NA	NA	NA	NA	NA	
CEloc									
$\rho_{EY.EZ}$	65.06	36.13	-0.09	0.10	0.08	0.19	0.14	0.06	
$\rho_{EX.EZ}$	64.16	35.96	-0.16	-0.06	0.08	0.18	0.14	0.06	
$\rho_{EX.EY}$	533.41	21.54	-0.15	0.02	0.08	0.28	0.14	0.06	
All	532.78	21.54	-0.12	0.05	0.08	0.27	0.13	0.06	
n									
25	442.56	84.83	-0.14	0.03	0.14	0.42	0.27	0.10	0.20
100	244.62	1.55	-0.13	0.03	0.07	0.19	0.10	0.05	0.10
400	209.38	0.00	-0.13	0.03	0.03	0.10	0.04	0.02	0.05

Note. [a]Study Conditions = 1,296. [b]Study Conditions = 864. TD = Theoretical distribution of *r* (i.e., $1 / [n - 1]^{.5}$).

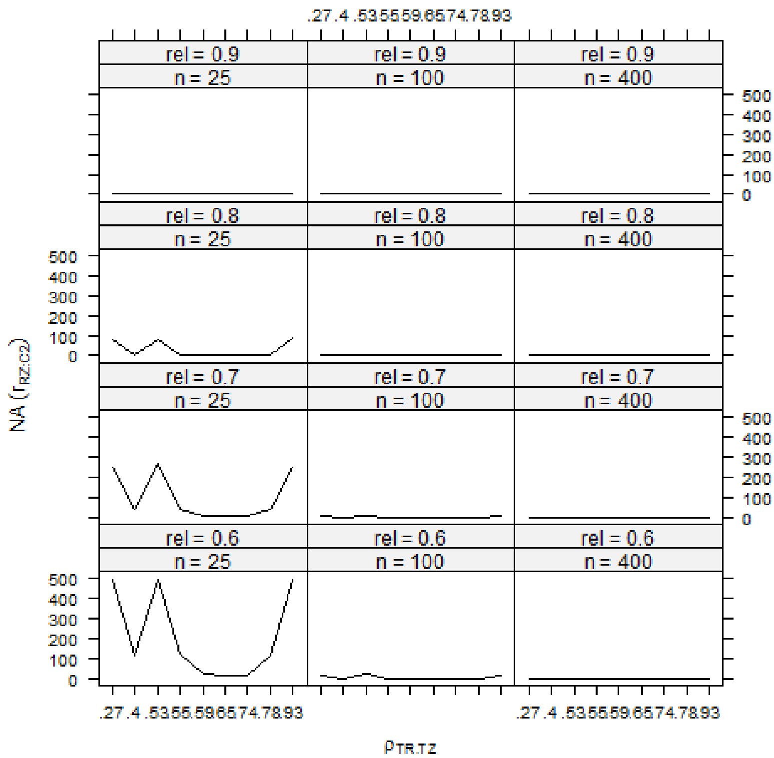

Figure 4. EMMs for number of NA nalues for $r_{\mathrm{RZ:C2}}$ across select study conditions.

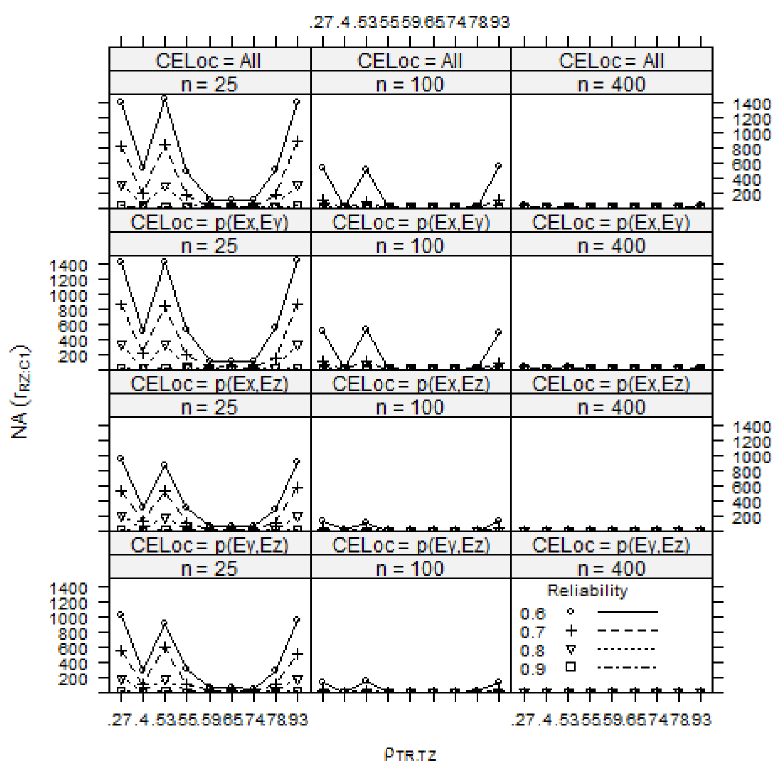

Figure 5a. EMMs for number of NA values for $r_{\mathrm{RZ:C1}}$ across select study conditions when correlated error = .1.

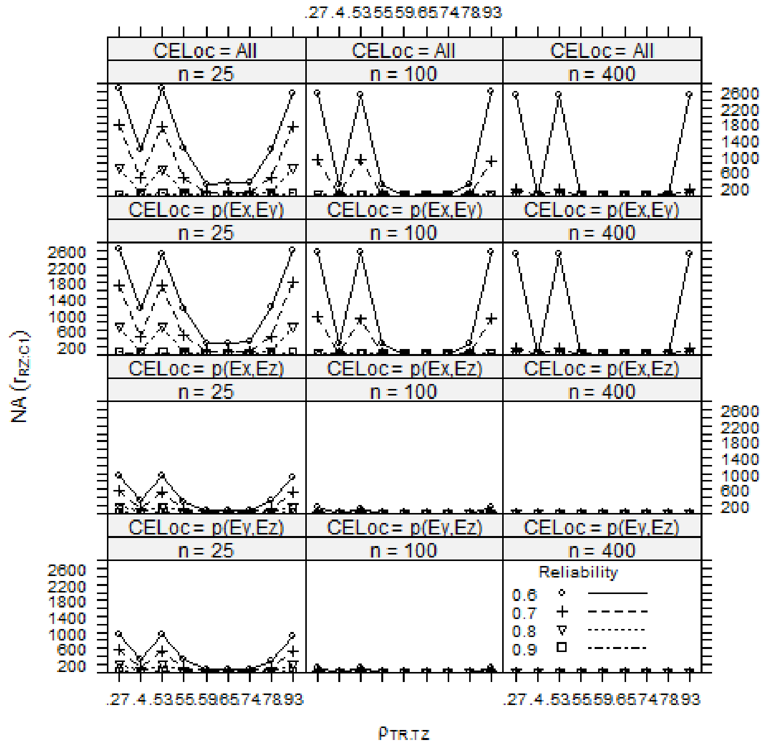

Figure 5b. EMMs for number of NA values for $r_{RZ:C1}$ across select study conditions when correlated error = .3.

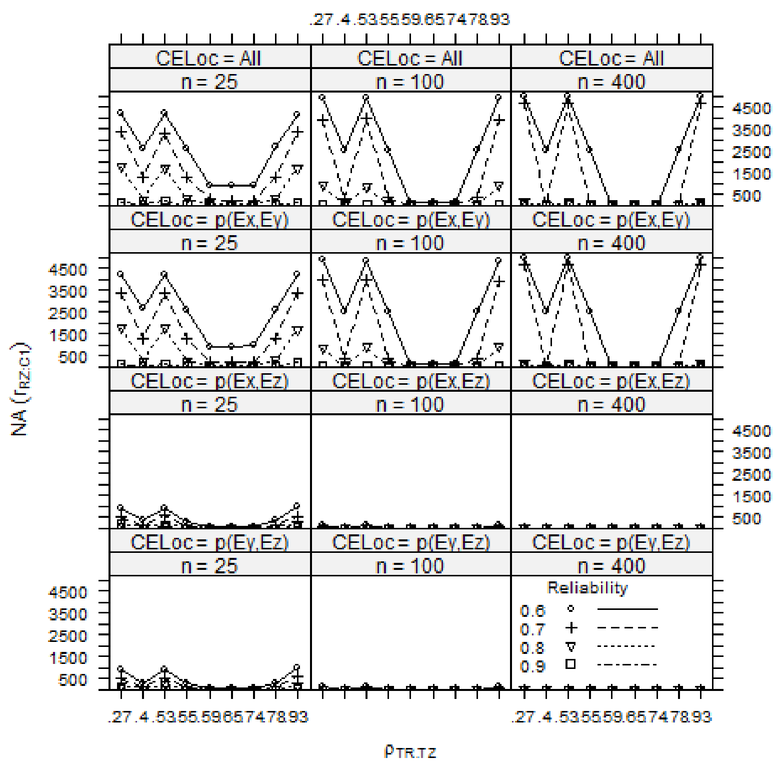

Figure 5c. EMMs for number of NA values for $r_{RZ:C1}$ across select study conditions when correlated error = .6.

Bias

Table 5 presents descriptive statistics and η^2 values across the study's factors. The greatest grand mean bias was in $r_{RZ:C0}$ followed by $r_{RZ:C1}$, while $r_{RZ:C2}$ and $r_{RZ:C3}$ demonstrated virtually no discernible grand mean bias. Across the study conditions, $r_{RZ:C1}$ demonstrated the most variance in mean bias followed by $r_{RZ:C0}$. There was little variability in the mean bias of $r_{RZ:C2}$ and $r_{RZ:C3}$ across the study conditions. The distribution of mean bias across the study conditions appeared to be negatively skewed and normal for $r_{RZ:C0}$ and $r_{RZ:C3}$, while positively skewed and non-normal for $r_{RZ:C1}$ and $r_{RZ:C2}$. Study conditions respectively explained 62.72%, 14.20%, .50%, and .83% of the variance in the bias of $r_{RZ:C0}$, $r_{RZ:C1}$, $r_{RZ:C2}$, and $r_{RZ:C3}$. Given the amount of bias detected and variance explained, only the bias of $r_{RZ:C0}$ and $r_{RZ:C1}$ were further investigated (cf. Thompson, 2006).

For $r_{RZ:C0}$, the study conditions that appeared to affect bias were $\rho_{T_R T_Z}$, reliability, CEloc, and the interaction between $\rho_{T_R T_Z}$ and reliability (see Table 5). The EMMs for the main effects show that across the study conditions, the greatest bias involved conditions where: $\rho_{T_R T_Z}$= .93, reliability = .60, and correlated error existed between E_X and E_Z or between E_X and E_Y (see Table 4). The EMMs for the $\rho_{T_R T_Z}$ and reliability interaction effect produced findings consistent with the main effects and revealed that while reliability did not play a role in the bias of $r_{RZ:C0}$ when $\rho_{T_R T_Z}$ was small (i.e., .27), the amount of negative bias increased as reliability decreased and $\rho_{T_R T_Z}$ increased (see Figure 6).

Table 5

Statistics for Bias of the 4 Validity Estimates across 864 Study Conditions

Statistic/Factor	$r_{RZ:C0}$	$r_{RZ:C1}$	$r_{RZ:C2}$	$r_{RZ:C3}$
Descriptive Statistics				
Median	-.11	<.01	<-.01	<-.01
M	-.13	.04	<-.01	<-.01
SD	.12	.16	.01	< .01
Skewness	-.84	5.00	3.45	-1.43
Kurtosis	.89	44.91	18.04	.79
η^2 for study condition factors				
$\rho_{TR.TZ}$	39.74%	2.75%	.16%	.15%
Reliability (rel)	12.38%	.72%	.02%	.00%
CE Level (CElev)	.13%	.34%	.00%	.00%
CE Location (CEloc)	3.97%	2.96%	.01%	.00%
n	.14%	.00%	.00%	.52%
$\rho_{TR.TZ}$ x rel	4.17%	1.36%	.07%	.00%
$\rho_{TR.TZ}$ x CElev	.04%	.76%	.00%	.00%
rel x CElev	.29%	2.33%	.00%	.00%
$\rho_{TR.TZ}$ x CEloc	.01%	.02%	.16%	.13%
rel x CEloc	.03%	.21%	.00%	.00%
CElev x CEloc	.60%	1.11%	.00%	.00%
$\rho_{TR.TZ}$ x *n*	.00%	.02%	.02%	.00%
rel x *n*	.98%	.79%	.00%	.00%
CElev x *n*	.00%	.01%	.00%	.00%
CEloc x *n*	.00%	.02%	.01%	.00%
$\rho_{TR.TZ}$ x rel x CElev	.00%	.47%	.00%	.00%
$\rho_{TR.TZ}$ x rel x CEloc	.01%	1.15%	.00%	.00%
$\rho_{TR.TZ}$ x CElev x CEloc	.00%	.25%	.06%	.00%
rel x CElev x CEloc	.07%	.91%	.00%	.00%
$\rho_{TR.TZ}$ x rel x *n*	.00%	.11%	.00%	.00%
$\rho_{TR.TZ}$ x CElev x *n*	.00%	.13%	.00%	.00%
rel x CElev x *n*	.15%	.32%	.00%	.00%
$\rho_{TR.TZ}$ x CEloc x *n*	.00%	.04%	.00%	.00%
rel x CEloc x n	.00%	.05%	.00%	.00%
CElev x CEloc x *n*	.00%	.02%	.00%	.00%
$\rho_{TR.TZ}$ x rel x CElev x CEloc	.00%	.52%	.00%	.00%
$\rho_{TR.TZ}$ x rel x CElev x *n*	.00%	.22%	.00%	.00%
$\rho_{TR.TZ}$ x rel x CEloc x *n*	.00%	.30%	.00%	.00%
$\rho_{TR.TZ}$ x CElev x CEloc x *n*	.00%	.12%	.00%	.00%
rel x CElev x CEloc x n	.00%	.05%	.00%	.00%
$\rho_{TR.TZ}$ x rel x CElev x CEloc x *n*	.00%	.23%	.00%	.00%
Full Model	62.72%	14.20%	.50%	.83%

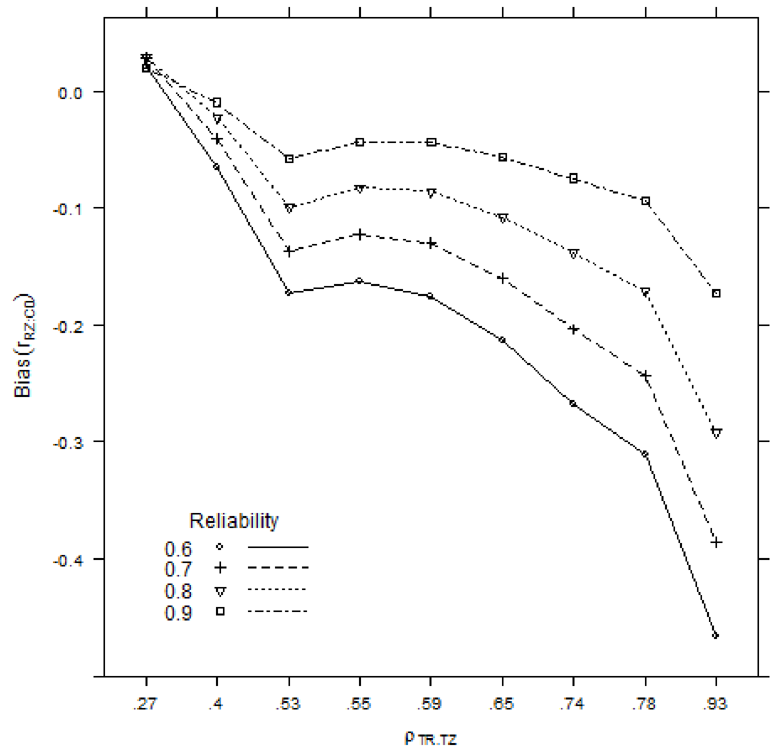

Figure 6. EMMs for bias of $r_{\mathrm{RZ:C0}}$ across select study conditions.

For $r_{RZ:C1}$, the study conditions that appeared to affect bias were the main effects of CELoc *and* $\rho_{T_R T_Z}$ as well as interaction effects involving $\rho_{T_R T_Z}$, reliability, CELoc, and CElev (see Table 3). The EMMs for the main effects show that across the study conditions, the greatest bias involved conditions where: correlated error was isolated to E_Y and E_Z and $\rho_{T_R T_Z}$ = .93 (see Table 4). To analyze the main effects with the context of relevant interactions, the EMMs associated with the four-way interaction involving $\rho_{T_R T_Z}$, reliability, CELoc, and CElev were plotted (see Figure 7). The EMMs for the four-way interaction effect produced findings somewhat inconsistent with the main effects, revealing that the greatest bias involved conditions where: correlated error existed between E_X and E_Y, $\rho_{T_R T_Z}$ = .93, and CElev = .30. The EMMs also revealed that decreases in reliability tended to increase the amount of *positive* bias except in the case where the correlated error was isolated to $\rho_{E_X E_Z}$. In the latter case, decreases in reliability tended to increase the amount of *negative* bias. When the correlated error condition was isolated to $\rho_{E_Y E_Z}$ or $\rho_{E_X E_Z}$, the amount of bias tended to be more pronounced when $\rho_{T_X T_Y}$ = .80. When the correlated error condition included $\rho_{E_X E_Y}$, the amount of bias was most pronounced when $\rho_{T_R T_Z}$= .53 | .93. The patterns of bias for the condition when CElev = .10 mirrored those when CElev = .30. However, when CElev = .10, the magnitude of bias was not as great.

Precision

Table 6 presents descriptive statistics and η^2 values across the study's factors. The greatest variability was found in $r_{RZ:C1}$, followed by $r_{RZ:C2}$, $r_{RZ:C0}$ and $r_{RZ:C3}$. The distribution of precision estimates across the study conditions appeared to be minimally skewed and slightly non-normal for $r_{RZ:C0}$ and $r_{RZ:C3}$, while positively skewed and non-normal for $r_{RZ:C1}$ and $r_{RZ:C2}$.

For $r_{RZ:C0}$, the study conditions that appeared to affect precision were *n*, reliability, and the interaction between *n* and reliability (see Table 6). The EMMs for the main effects show that across the study conditions, the greatest variability involved conditions where: *n* = 25 and reliability = .60 (see Table 4). The EMMs for the *n* x reliability interaction effect produced findings consistent with the main effects and revealed a slight interaction between *n* and reliability and indicated that lower levels of reliability had an increasingly negative effect on precision as *n* decreased (see Figure 8).

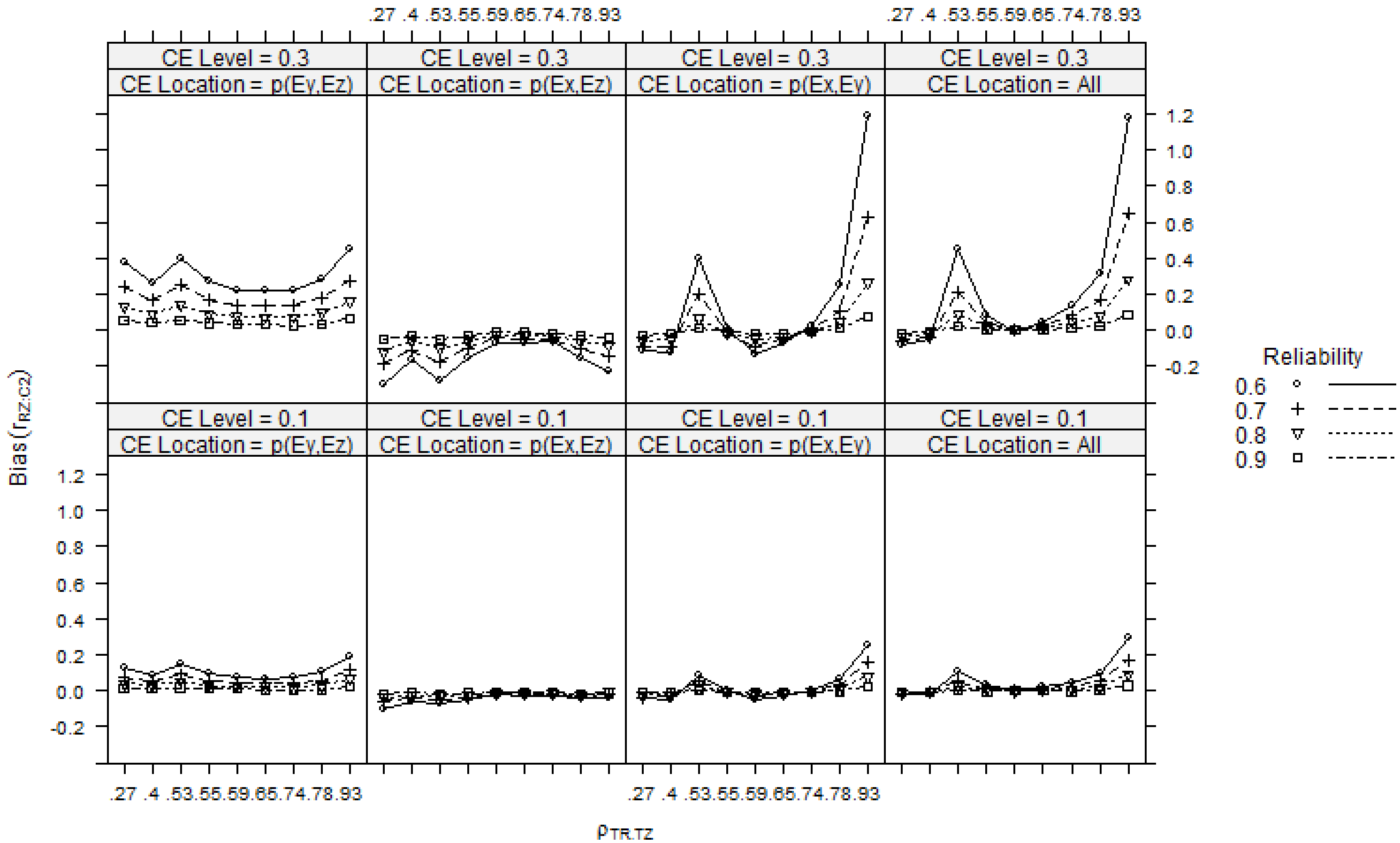

Figure 7. EMMs for bias of $r_{RZ:C1}$ across select study conditions.

Table 6

Statistics for Precision of the 4 Validity Estimates across 864 Study Conditions

Statistic/Factor	$r_{RZ:C0}$	$r_{RZ:C1}$	$r_{RZ:C2}$	$r_{RZ:C3}$
Descriptive Statistics				
Median	.07	.13	.08	.05
M	.08	.23	.14	.06
SD	.05	.30	.13	.03
Skewness	.54	3.17	2.23	.59
Kurtosis	-1.17	14.74	5.83	-1.08
η^2 for study condition factors				
$\rho_{TR.TZ}$	0.95%	14.91%	5.57%	9.04%
Reliability (rel)	3.38%	19.24%	13.57%	0.00%
CE Level (CElev)	0.08%	1.13%	0.04%	0.00%
CE Location (CEloc)	0.28%	2.36%	0.18%	0.00%
n	93.24%	20.86%	53.45%	88.35%
$\rho_{TR.TZ}$ x rel	0.29%	9.38%	2.83%	0.00%
$\rho_{TR.TZ}$ x CElev	0.00%	1.71%	0.07%	0.00%
rel x CElev	0.01%	0.99%	0.02%	0.00%
$\rho_{TR.TZ}$ x CEloc	0.07%	3.70%	0.35%	0.00%
rel x CEloc	0.05%	2.17%	0.16%	0.00%
CElev x CEloc	0.07%	0.97%	0.08%	0.00%
$\rho_{TR.TZ}$ x *n*	0.26%	1.74%	5.75%	2.57%
rel x *n*	1.01%	3.98%	10.13%	0.00%
CElev x *n*	0.02%	0.00%	0.02%	0.00%
CEloc x *n*	0.08%	0.05%	0.13%	0.00%
$\rho_{TR.TZ}$ x rel x CElev	0.00%	1.42%	0.20%	0.00%
$\rho_{TR.TZ}$ x rel x CEloc	0.00%	2.93%	0.62%	0.00%
$\rho_{TR.TZ}$ x CElev x CEloc	0.02%	1.72%	0.28%	0.00%
rel x CElev x CEloc	0.01%	0.79%	0.09%	0.00%
$\rho_{TR.TZ}$ x rel x *n*	0.07%	1.78%	2.72%	0.00%
$\rho_{TR.TZ}$ x CElev x *n*	0.00%	0.39%	0.09%	0.00%
rel x CElev x *n*	0.00%	0.10%	0.02%	0.00%
$\rho_{TR.TZ}$ x CEloc x *n*	0.02%	0.87%	0.43%	0.00%
rel x CEloc x n	0.01%	0.15%	0.17%	0.00%
CElev x CEloc x *n*	0.02%	0.02%	0.11%	0.00%
$\rho_{TR.TZ}$ x rel x CElev x CEloc	0.00%	1.32%	0.36%	0.00%
$\rho_{TR.TZ}$ x rel x CElev x *n*	0.00%	0.85%	0.31%	0.00%
$\rho_{TR.TZ}$ x rel x CEloc x *n*	0.01%	2.39%	1.02%	0.01%
$\rho_{TR.TZ}$ x CElev x CEloc x *n*	0.01%	0.41%	0.44%	0.00%
rel x CElev x CEloc x n	0.00%	0.28%	0.13%	0.00%
$\rho_{TR.TZ}$ x rel x CElev x CEloc x *n*	0.01%	1.41%	0.70%	0.01%

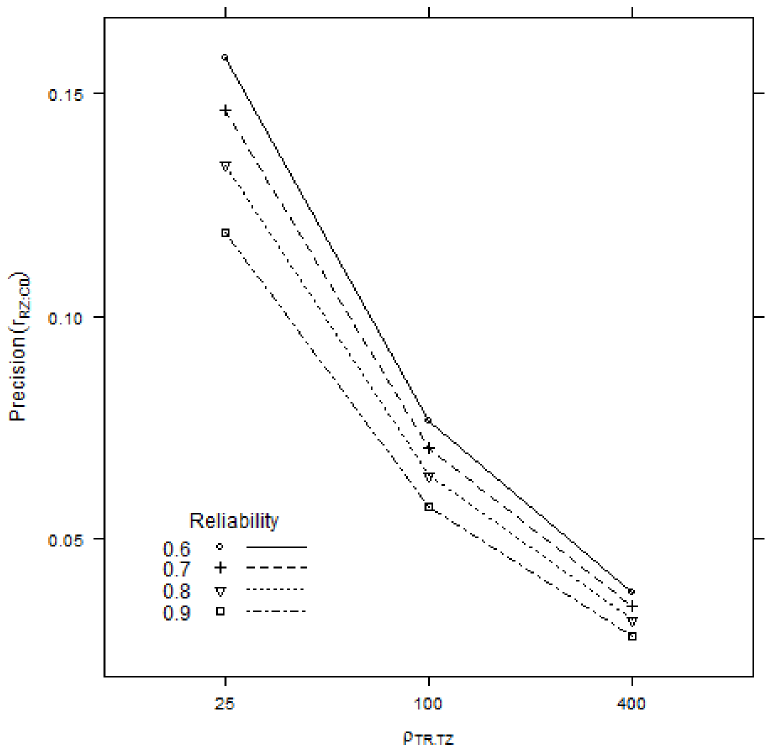

Figure 8. EMMs for precision of $r_{\mathrm{RZ:C0}}$ across select study conditions.

For $r_{RZ:C1}$, the study conditions that appeared to affect precision were the main effects of *n*, reliability, $\rho_{\mathrm{T_R T_Z}}$, CElev, CELoc, as well as interaction effects involving $\rho_{\mathrm{T_R T_Z}}$, reliability, CELoc, and *n* (see Table 6). The EMMs for the main effects show that across the study conditions, the greatest bias involved conditions where: *n* = 25, reliability = .60, and $\rho_{\mathrm{T_R T_Z}}$= .93, CElev = .30, and correlated error included E_X and E_Y (see Table 4). To analyze the main effects with the context of relevant interactions, the EMMs associated with the four-way interaction involving $\rho_{T_R T_Z}$, reliability, CELoc, and *n* were plotted (see Figure 9). The EMMs for the $\rho_{\mathrm{T_R T_Z}}$ x reliability x CELoc x *n* interaction effect produced findings somewhat consistent with the main effects, revealing that the greatest variability involved conditions where: $\rho_{\mathrm{T_R T_Z}}$ = .93, reliability = .60, correlated error existed between E_X and E_Y, and *n* = 400.

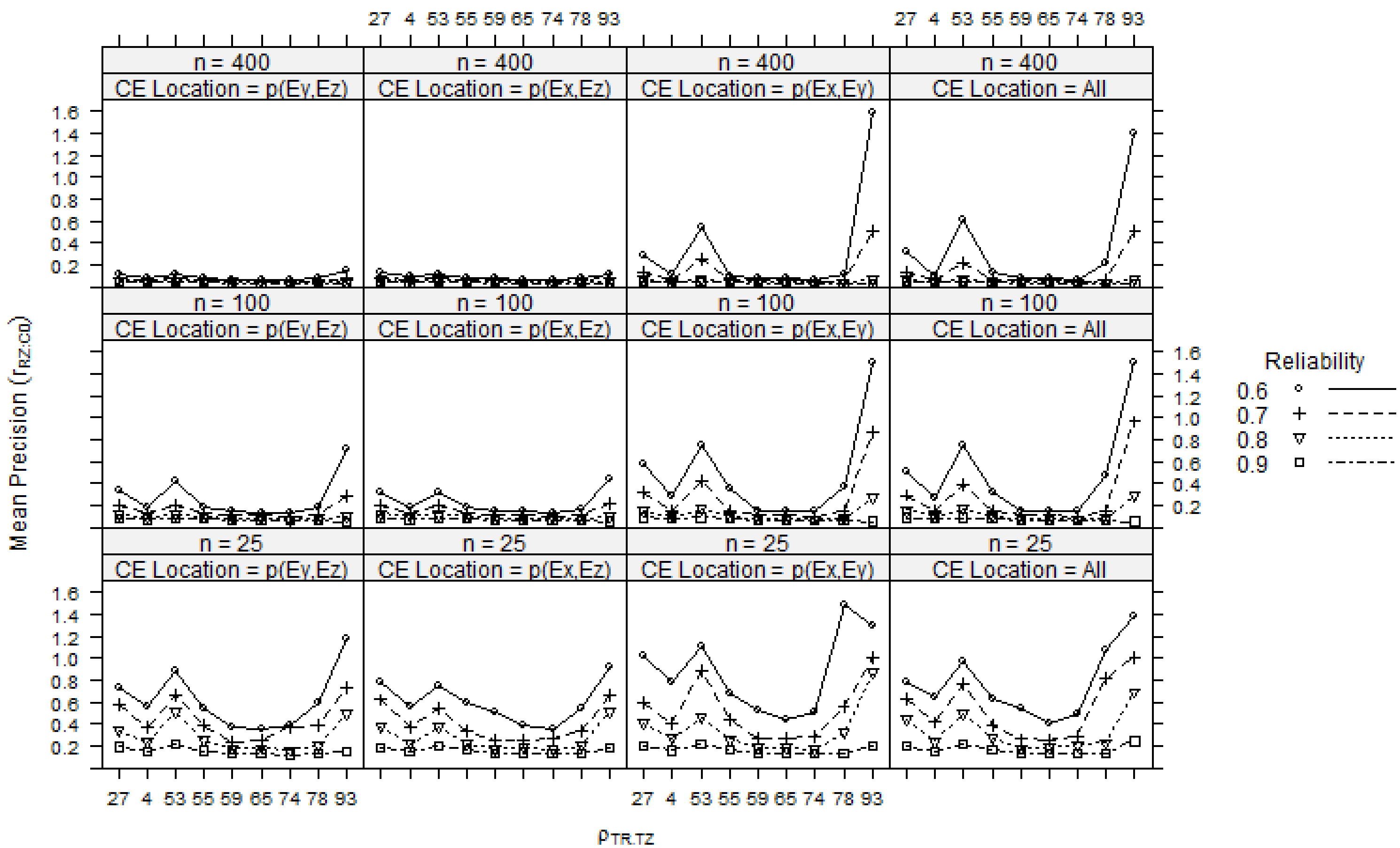

Figure 9. EMMs for precision of $r_{RZ:C1}$ across select study conditions.

For $r_{RZ:C2}$, the study conditions that appeared to affect precision were *n*, reliability, $\rho_{T_R T_Z}$, and related interactions (see Table 6). The EMMs for the main effects show that across study conditions, the greatest variability involved conditions where: *n* = 25, reliability = .60, and $\rho_{T_X T_Y}$ = .80 (see Table 4). The EMMs for *n* x reliability x $\rho_{T_R T_Z}$ interaction effect produced findings consistent with the main effects and also revealed that the interaction between reliability and $\rho_{T_R T_Z}$ was most evident in conditions where *n* = 25. For *n* = 25, the effect that $\rho_{T_R T_Z}$ had on precision was strongest when reliability = .60 and weakest when reliability = .90 (see Figure 10).

For $r_{RZ:C3}$, the study conditions that appeared to affect precision were *n*, $\rho_{T_R T_Z}$, and the interaction between *n* and $\rho_{T_R T_Z}$ (see Table 6). The EMMs for the main effects show that across the study conditions, the greatest variability involved conditions where: *n* = 25 and $\rho_{T_R T_Z}$ = .27 | .53 | .55 (see Table 4). The EMMs for the interaction effect produced findings consistent with the main effects and also indicated that the effect $\rho_{T_R T_Z}$ had on precision was strongest when *n* = 25 and weakest when *n* = 400 (see Figure 11).

Discussion

Synthesis of Findings

To synthesize the results of the study, it is helpful to consider differences in bias or precision that are large enough to make a practical difference in the interpretation of validity estimates. The benchmark used for identifying practical differences in bias and precision was .20 and is supported by literature that considers the magnitude of *r* when interpreting its meaning. For example, Cohen (1998) interpreted *r*s of .10, .30, and .50 as small, medium and large, respectively. When considering convergent validity levels, Ward, Fischer, Lam, and Hall

(2009) drew distinctions between correlations greater than .70, those that ranged between .30 and .50, and those that ranged between .10 and .20.

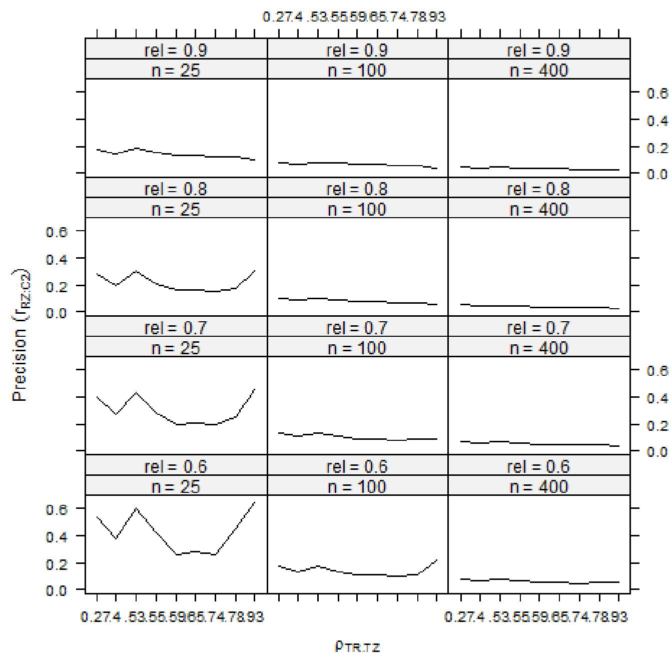

Figure 10. EMMs for precision of $r_{\mathrm{RZ:C2}}$ across select study conditions.

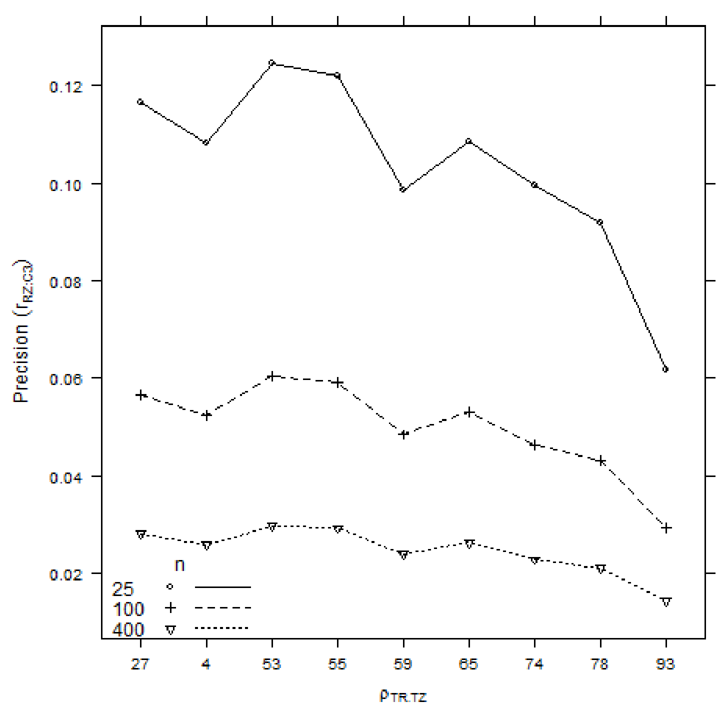

Figure 11. EMMs for precision of $r_{\mathrm{RZ:C3}}$ across select study conditions.

$r_{RZ:C3}$. Validity estimates based on Eq. C3 yielded values that were not practically significantly different than their respective population parameters, across the study conditions. This finding should not be surprising given that the derivation of $r_{RZ:C2}$ stems from Zimmerman's formula (2007, Eq. 6) which accurately yields sample correlations among true scores. Although Eq. C3 may not be used when only observed scores are available, the findings indicate that unbiased and precise estimates for r_{RZ}may be obtained when the reliability of data, correlations among error scores, and correlations between error and true scores are simultaneously modeled.

$r_{RZ:C2}$. Validity estimates based on Eq. C2, that is similar to C3 but does not model correlations between error and true scores, yielded unbiased values. However, due to the role that sample size played in the variability of correlations between error and true scores, the variability in $r_{RZ:C2}$ estimates were large enough to be practically significantly different than their respective population parameters, particularly when sample sizes were small (*n* = 25) and reliability was low (.60 | .70) or when reliability was moderate (.80) and the level of co-variation was high (.80).

$r_{RZ:C1}$. Validity estimates based on Eq. C1, that is similar to C2 but does not model correlations between error scores, yielded values that were practically significantly different than their respective population parameters under a number of conditions. Particularly when the level of correlated error was moderate (.30), reliability was low (.60), and the level of co-variation was high (.80), biased estimates occurred. If the correlated error was isolated to the covariate and criterion, as might occur if the covariate and criterion were measured at the same time but at a different time than the outcome, the bias was negative; otherwise, positive bias occurred when correlated error was isolated to the outcome and criterion, or involved the covariate and outcome. In addition, when reliability was low (.60) and the level of co-variation was high (.80), variability in $r_{RZ:C1}$ estimates was large enough to be practically significantly different than their respective population parameters. The exception to this finding was when sample sizes were large and when correlated error was isolated to the outcome and criterion or the covariate and criterion. In the latter case, there was no practically significant variability in $r_{RZ:C1}$ estimates.

$r_{RZ:C0}$. Validity estimates based on Eq. C0, that is similar to C1 does not model reliability, yielded values that were sufficiently precise to be not practically significantly different than their respective population parameters. However, under a number of conditions, $r_{RZ:C0}$ estimates were sufficiently biased to affect the interpretation of results. Particularly when: (a) $\rho_{T_R T_Z} \geq .65$ and reliability = .60, (b) $\rho_{T_R T_Z} \geq .74$ and reliability = .70, and (c) $\rho_{T_R T_Z} = .93$ and reliability = .80, there was a practically significant amount of negative bias in $r_{RZ:C0}$ estimates.

Conclusions

Based on the study's findings, it can be concluded that the performance of four equations ranked from best to worst is as follows: C3, C2, C0, and C1. Eq. C3 performed the best, yielding no estimates that were practically significantly different than their respective population parameters, across study conditions. Eq. C2 did a good job in yielding unbiased, valid, and precise results. Only in a select number of study conditions were estimates unable to be computed or produced results that had sufficient variance to affect interpretation of results. Eq. C0 fared well in producing valid, precise, and unbiased results. Although C0 suffered from bias under a number of study conditions, the variability of results were not sufficiently large to affect interpretation of results. Eq. C0 also reliably produced estimates across all study conditions (i.e., no NA values). Eq. C1 performed the worst. Not only did C1 not reliably produce estimates even when the level of modeled correlated error was low (.10), C1 produced values higher than the theoretical limit of 1.0 across a number of study conditions. Estimates based on C1 also produced the greatest number of conditions that were practically significantly different than their population parameters.

The study's findings support the benefits of using structural equation modeling (SEM) in to accurately assess the validity of a residualized variable. Presuming that equivalent SEM models can be built to model the correlated error in Eq. C2, researchers can expect such models to generally yield unbiased, valid, and precise validity estimates. However, based on the findings of this study, it seems possible that constraining nuisance correlations between error scores to zero could yield correlation coefficients greater than 1.0. It is possible that SEM models could yield a closer fit of the data if nuisance correlations were freely estimated, similar to how nuisance correlations are modeled in Eq. C3. Alternatively, such correlations might be truncated to unity or perhaps another value such as the observed score correlation.

In the event that SEM cannot be applied to a dataset for any number of reasons (e.g., insufficient sample size, lack of item level data), researchers must then choose between C0 and C1 when assessing the validity of a residualized variable. Eq. C0 would seem to be the most parsimonious choice when: (a) $\rho_{T_R T_Z} < .65$ and reliability $\geq .60$, (b) $\rho_{T_R T_Z} \leq .74$ and reliability $\geq .70$, or (c) $\rho_{T_R T_Z} < .93$ and reliability $\geq .80$, as this study found a practically insignificant amount of bias in the associated $r_{RZ:C0}$ estimates. In the absence of such conditions, a post hoc analyses of the related EMMs revealed that C1 usually outperformed C0 in producing less biased estimates, except in select conditions (see Table 7). However, in the absence of the aforementioned conditions, C1 always produced less precise estimates than C0 (see Table 7). These findings have implications for analyses (e.g., meta-analyses) that rely on validity estimates that have been corrected by C1. Unless researchers are confident there is no correlated error among the variables, they must give careful consideration to the: (a) level of correlation between a residualized variable and criterion, (b) level of reliability in the data, and (c) potential source of correlated error before selecting a validity estimate that has been corrected for attenuation by C1 over a validity estimate based on observed score correlations (i.e., C0). Without such consideration, resulting estimates may be actually further away from population parameters than estimates based on observed data.

Table 7

EMMs of Bias and Precision for $r_{RZ:C0}$ and $r_{RZ:C1}$ across Select Conditions

				Bias		Precision	
$\rho_{TR.TZ}$	rel	CElev	CEloc[a]	$r_{RZ:C0}$	$r_{RZ:C1}$	$r_{RZ:C0}$	$r_{RZ:C1}$
0.65	0.6	0.1	$\rho_{E_YE_Z}$	-0.20	**0.07**	**0.09**	0.18
0.65	0.6	0.1	$\rho_{E_XE_Z}$	-0.24	**-0.03**	**0.09**	0.20
0.65	0.6	0.1	$\rho_{E_XE_Y}$	-0.25	**-0.03**	**0.09**	0.19
0.65	0.6	0.1	All	-0.21	**0.02**	**0.09**	0.20
0.65	0.6	0.3	$\rho_{E_YE_Z}$	**-0.11**	0.22	**0.08**	0.19
0.65	0.6	0.3	$\rho_{E_XE_Z}$	-0.25	**-0.07**	**0.09**	0.20
0.65	0.6	0.3	$\rho_{E_XE_Y}$	-0.27	**-0.07**	**0.09**	0.24
0.65	0.6	0.3	All	-0.18	**0.04**	**0.08**	0.22
0.74	0.6	0.1	$\rho_{E_YE_Z}$	-0.26	**0.07**	**0.09**	0.19
0.74	0.6	0.1	$\rho_{E_XE_Z}$	-0.30	**-0.02**	**0.09**	0.18
0.74	0.6	0.1	$\rho_{E_XE_Y}$	-0.30	**0.00**	**0.09**	0.21
0.74	0.6	0.1	All	-0.27	**0.05**	**0.09**	0.19
0.74	0.6	0.3	$\rho_{E_YE_Z}$	**-0.18**	0.22	**0.08**	0.19
0.74	0.6	0.3	$\rho_{E_XE_Z}$	-0.32	**-0.07**	**0.09**	0.19
0.74	0.6	0.3	$\rho_{E_XE_Y}$	-0.30	**0.02**	**0.09**	0.27
0.74	0.6	0.3	All	-0.22	**0.14**	**0.08**	0.27
0.74	0.7	0.1	$\rho_{E_YE_Z}$	-0.19	**0.04**	**0.08**	0.13
0.74	0.7	0.1	$\rho_{E_XE_Z}$	-0.23	**-0.02**	**0.09**	0.14
0.74	0.7	0.1	$\rho_{E_XE_Y}$	-0.23	**-0.01**	**0.09**	0.13
0.74	0.7	0.1	All	-0.20	**0.02**	**0.08**	0.13
0.74	0.7	0.3	$\rho_{E_YE_Z}$	**-0.13**	0.14	**0.07**	0.21
0.74	0.7	0.3	$\rho_{E_XE_Z}$	-0.24	**-0.05**	**0.09**	0.14
0.74	0.7	0.3	$\rho_{E_XE_Y}$	-0.23	**0.00**	**0.09**	0.16
0.74	0.7	0.3	All	-0.17	**0.08**	**0.08**	0.15
0.78	0.6	0.1	$\rho_{E_YE_Z}$	-0.29	**0.10**	**0.09**	0.26
0.78	0.6	0.1	$\rho_{E_XE_Z}$	-0.35	**-0.04**	**0.09**	0.24

(table continues)

$\rho_{TR.TZ}$	rel	CElev	CEloc[a]	Bias $r_{RZ:C0}$	Bias $r_{RZ:C1}$	Precision $r_{RZ:C0}$	Precision $r_{RZ:C1}$
0.78	0.6	0.1	$\rho_{E_XE_Y}$	-0.34	**0.06**	**0.10**	0.46
0.78	0.6	0.1	All	-0.31	**0.10**	**0.09**	0.45
0.78	0.6	0.3	$\rho_{E_YE_Z}$	**-0.21**	0.28	**0.08**	0.31
0.78	0.6	0.3	$\rho_{E_XE_Z}$	-0.38	**-0.15**	**0.10**	0.29
0.78	0.6	0.3	$\rho_{E_XE_Y}$	-0.33	**0.26**	**0.09**	0.86
0.78	0.6	0.3	All	**-0.26**	0.31	**0.09**	0.73
0.78	0.7	0.1	$\rho_{E_YE_Z}$	-0.23	**0.06**	**0.08**	0.16
0.78	0.7	0.1	$\rho_{E_XE_Z}$	-0.28	**-0.03**	**0.09**	0.17
0.78	0.7	0.1	$\rho_{E_XE_Y}$	-0.26	**0.03**	**0.09**	0.22
0.78	0.7	0.1	All	-0.24	**0.05**	**0.08**	0.20
0.78	0.7	0.3	$\rho_{E_YE_Z}$	**-0.16**	0.18	**0.07**	0.21
0.78	0.7	0.3	$\rho_{E_XE_Z}$	-0.30	**-0.10**	**0.09**	0.17
0.78	0.7	0.3	$\rho_{E_XE_Y}$	-0.26	**0.11**	**0.09**	0.30
0.78	0.7	0.3	All	-0.21	**0.17**	**0.08**	0.48
0.93	0.6	0.1	$\rho_{E_YE_Z}$	-0.45	**0.19**	**0.09**	0.62
0.93	0.6	0.1	$\rho_{E_XE_Z}$	-0.51	**-0.04**	**0.10**	0.49
0.93	0.6	0.1	$\rho_{E_XE_Y}$	-0.49	**0.25**	**0.09**	0.79
0.93	0.6	0.1	All	-0.47	**0.29**	**0.09**	0.91
0.93	0.6	0.3	$\rho_{E_YE_Z}$	**-0.36**	0.45	**0.08**	0.74
0.93	0.6	0.3	$\rho_{E_XE_Z}$	-0.56	**-0.23**	**0.10**	0.50
0.93	0.6	0.3	$\rho_{E_XE_Y}$	**-0.48**	1.19	**0.09**	2.12
0.93	0.6	0.3	All	**-0.42**	1.18	**0.08**	1.95
0.93	0.7	0.1	$\rho_{E_YE_Z}$	-0.37	**0.12**	**0.08**	0.36
0.93	0.7	0.1	$\rho_{E_XE_Z}$	-0.43	**-0.02**	**0.09**	0.33
0.93	0.7	0.1	$\rho_{E_XE_Y}$	-0.40	**0.16**	**0.08**	0.47
0.93	0.7	0.1	All	-0.39	**0.17**	**0.08**	0.55
0.93	0.7	0.3	$\rho_{E_YE_Z}$	-0.30	**0.27**	**0.07**	0.36
0.93	0.7	0.3	$\rho_{E_XE_Z}$	-0.47	**-0.15**	**0.09**	0.30
0.93	0.7	0.3	$\rho_{E_XE_Y}$	**-0.38**	0.63	**0.08**	1.12
0.93	0.7	0.3	All	**-0.34**	0.65	**0.08**	1.10
0.93	0.8	0.1	$\rho_{E_YE_Z}$	-0.28	**0.07**	**0.07**	0.20
0.93	0.8	0.1	$\rho_{E_XE_Z}$	-0.33	**-0.01**	**0.07**	0.23
0.93	0.8	0.1	$\rho_{E_XE_Y}$	-0.30	**0.08**	**0.07**	0.29
0.93	0.8	0.1	All	-0.29	**0.09**	**0.07**	0.24
0.93	0.8	0.3	$\rho_{E_YE_Z}$	-0.23	**0.16**	**0.06**	0.22
0.93	0.8	0.3	$\rho_{E_XE_Z}$	-0.36	**-0.08**	**0.08**	0.21
0.93	0.8	0.3	$\rho_{E_XE_Y}$	-0.28	**0.26**	**0.07**	0.51
0.93	0.8	0.3	All	**-0.26**	0.28	**0.06**	0.46

Note. Estimates with less bias and variability are bolded.

Limitations

The study's findings must be taken within the context of the study's limitations and delimitations. First, the study only considered moderately high level of convergent validity between the outcome and criterion variable ($\rho_{T_Y T_Z}$= .80), equal levels of reliability across the outcome, covariate, and criterion variables, and four correlated error conditions, where correlated error was isolated to one of three pairs of variables as well as across all variables. No conclusions can be made, therefore, regarding the performance of C0, C1, and C2, across study conditions with lower levels of convergent validity, disparate levels of reliability across the outcome, covariate, and criterion variables, and other correlated error conditions. Second, given that C1 was unable to reliably provide estimates when the level of correlated error = .60, it is not clear how C0, C1, and C2 will compare in the presence of a strong amount of correlated error, beyond understanding the conditions when estimates based on C1 and C2 cannot be computed. Third, this study only considered a single covariate, where in practice researchers may use multiple covariates in their analyses. Fourth, the study was limited by the performance of the equations which included in the case of C1 and C2 yielding estimates that were greater than theoretical limits. It is possible that C1 and C2 might have performed better had components of the formula been forced within theoretical bounds (e.g, $\frac{r_{XZ}}{\sqrt{r_{ZZ} r_{XX}}}$). However, the findings from this study are nonetheless useful as they help identify cases when such estimates may occur.

Recommendations for Future Research

Beyond considering conditions that were not simulated in the study, the findings from this study present a number of recommendations for future research. The first recommendation relates to corrected correlations that yield values beyond $\pm$ 1. The second recommendation considers what data should be included in research reports.

First, further research is needed to determine how best to handle corrected correlations that result in values greater than or less than their theoretical limit. While conventional wisdom (e.g., Onwuegbuzie et al., 2004) suggests that such values should be truncated to unity, Nimon et al. (2012) found that a "truncated correlation of 1.00 may be a less accurate estimate of the true score correlation than its observed course counterpart" (p. 3). It would, therefore, be helpful to empirically define the conditions when a corrected correlation coefficient should be truncated to unity or its observed course counterpart. Future research could then use such knowledge to create variants of C1 and C2 that force components of the equations (e.g, $\frac{r_{XZ}}{\sqrt{r_{ZZ}r_{XX}}}$). to be within theoretical bounds (i.e., $\pm$ 1).

Second, the findings of the study demonstrated the impact correlated error has on residualized validity estimates that result from C1. It would help inform future research if reports based on SEM analyses would identify the amount of correlated error between variables. While this would necessitate researchers conducting additional analyses (e.g., creating phantom variables to model correlated error at the scale level, analyzing residual correlations), such knowledge would help begin to build a body of knowledge regarding the level of correlated error that exists across sources of data. This recommendation builds on the work of Zientek and Thompson (2009) that recommends that correlation matrices be provided to help improve research reports and allow for secondary analyses. Given the findings from this study as well as Fan's (2003) who demonstrated that OLS and SEM approaches to correcting for measurement error yield nearly identical results, it would also seem advantageous for researchers to report both observed score correlations and correlations that have been corrected for measurement error (either via SEM or Eq. C1) so that comparisons could be made between the sets. As noted by Zimmerman, if a correlation that has only been corrected for measurement error is much greater than a correlation resulting from observed scores, it is probably inaccurate as it does not consider correlated error or nuisance correlations between true and error scores.

Implications for Applied Research

The results from this study provide two helpful data points for applied researchers. First, in the event that a corrected correlation results in a value greater than unity or the calculation of residualized validity coefficient results in an impossible value, the researcher should consider the role that correlated error may be playing in their analyses and investigate means (e.g., SEM) to model the correlated error. Second, the study demonstrates that in the presence of highly reliable data, the validity of a residualized variable may be accurately and parsimoniously modeled with observed score data, even in the presence of correlated error.

APPENDIX

SIMULATION CODE

```
library(MASS)
library(corpcor)

path<-"C:/Users/nimonmedia/My Documents/Dissert/"

#Study Conditions

pops<-1000000
iter<-5000

ptxty<-c(.8, .6, .3)
ptxtz<-c(.8, .6, .3)
ptytz<-c(.8)
rel<-c(.6, .7, .8, .9)
ce<-c(.1, .3, .6)
ss<-c(25, 100, 400)
ec<-matrix(c(1,0,0,0,1,0,0,0,1,1,1,1),byrow=TRUE,nrow=4,ncol=3)
sc1<-expand.grid(ptxtz,ptxty,ptytz,rel,ce,ec[1,1],ec[1,2],ec[1,3])
sc2<-expand.grid(ptxtz,ptxty,ptytz,rel,ce,ec[2,1],ec[2,2],ec[2,3])
sc3<-expand.grid(ptxtz,ptxty,ptytz,rel,ce,ec[3,1],ec[3,2],ec[3,3])
sc4<-expand.grid(ptxtz,ptxty,ptytz,rel,ce,ec[4,1],ec[4,2],ec[4,3])
sc<-rbind(sc1,sc2,sc3,sc4)
ptrtz<-rep(c(.27, .53, .93, .40, .55, .78, .59,
.65,.74),length(rel)*length(ce)*nrow(ec))
sc<-cbind(ptrtz,sc)
id<-1:nrow(sc)
sc<-cbind(id,sc)
colnames(sc)<-
c("id","ptrtz","ptxtz","ptxty","ptytz","rel","ce","pexey","pexez","peyez")
sc$pexey<-sc$pexey*sc$ce
sc$pexez<-sc$pexez*sc$ce
sc$peyez<-sc$peyez*sc$ce
outp<-matrix(nrow=nrow(sc)  ,ncol=44)
colnames(outp)<-
c("id","ss","k","txty","txtz","tytz","trtz","oxoy","oxoz","oyoz","oroz",
"eyez","exey","exez","txey","txez","tyex","tyez","tzex","tzey",
"txex","tyey","tzez","rxx","ryy","rzz","rxxc","ryyc","rzzc","exx","eyy",
"ezz","ztxty","ztxtz","ztytz","ztrtz","stxty","stxtz","stytz","strtz",
"wtxty","wtxz","wtytz","wtrtz")

outp<-data.frame(outp)

outs<-matrix(nrow=iter,ncol=108)

colnames(outs)<-c(colnames(outp),"btxty","btxtz","btytz","btrtz",
                  "boxoy","boxoz","boyoz","boroz",
                  "beyez","bexey","bexez",
                  "brxx","bryy","brzz",
                  "brxxc","bryyc","brzzc",
                  "bexx","beyy","bezz",
                  "bztxty","bztxtz","bztytz","bztrtz",
                  "bstxty","bstxtz","bstytz","bstrtz",
                  "bwtxty","bwtxtz","bwtytz","bwtrtz",
                  "abtxty","abtxtz","abtytz","abtrtz",
                  "aboxoy","aboxoz","aboyoz","aboroz",
                  "abeyez","abexey","abexez",
                  "abrxx","abryy","abrzz",
```

```
                 "abrxxc","abryyc","abrzzc",
                 "abexx","abeyy","abezz",
                 "abztxty","abztxtz","abztytz","abztrtz",
                 "abstxty","abstxtz","abstytz","abstrtz",
                 "abwtxty","abwtxtz","abwtytz","abwtrtz")

outs<-data.frame(outs)

outmean<-matrix(nrow=nrow(sc)*length(ss),ncol=123)
colnames(outmean)<-c(colnames(outs),
                 "cp2trtz","cp2oroz","cp2ztrtz","cp2strtz","cp2wtrtz",
                 "cn2trtz","cn2oroz","cn2ztrtz","cn2strtz","cn2wtrtz",
                 "cnatrtz","cnaoroz","cnaztrtz","cnastrtz","cnawtrtz")
outmean<-data.frame(outmean)
outsd<-outmean

for (i in 1:nrow(sc)){

  set.seed(1234+i)
  pcov<-matrix(0,nrow=6, ncol=6)
  colnames(pcov)<-rownames(pcov)<-c("tx","ty","tz","ex","ey","ez")
  pcov["tx","tx"]<-pcov["ty","ty"]<-pcov["tz","tz"]<-sc$rel[i]
  pcov["ex","ex"]<-pcov["ey","ey"]<-pcov["ez","ez"]<-1-sc$rel[i]
  pcov["ty","tx"]<-sc$ptxty[i]*sc$rel[i]
  pcov["tz","tx"]<-sc$ptxtz[i]*sc$rel[i]
  pcov["tz","ty"]<-sc$ptytz[i]*sc$rel[i]
  pcov["ey","ex"]<-sc$pexey[i]*(1-sc$rel[i])
  pcov["ez","ex"]<-sc$pexez[i]*(1-sc$rel[i])
  pcov["ez","ey"]<-sc$peyez[i]*(1-sc$rel[i])

  ct<-nrow(pcov)
  for (j in 1:ct){
    for (k in 2:ct){
      pcov[j,k]<-pcov[k,j]
    }
  }
  pcov<-make.positive.definite(pcov)
  data1<-mvrnorm(n=pops,rep(0,ct),pcov,empirical=TRUE)
  data1<-data.frame(data1)
  ox<-data1$tx+data1$ex
  oy<-data1$ty+data1$ey
  oz<-data1$tz+data1$ez
  data1<-cbind(data1,ox,oy,oz)

  outp$k[i]<-0
  outp$id[i]<-i
  outp$oxoy[i]<-cor(data1$ox,data1$oy)
  outp$oxoz[i]<-cor(data1$ox,data1$oz)
  outp$oyoz[i]<-cor(data1$oy,data1$oz)
  outp$oroz[i]<-part(outp$oxoy[i],outp$oxoz[i],outp$oyoz[i])
  outp$txty[i]<-cor(data1$tx,data1$ty)
  outp$txtz[i]<-cor(data1$tx,data1$tz)
  outp$tytz[i]<-cor(data1$ty,data1$tz)
  outp$trtz[i]<-part(outp$txty[i],outp$txtz[i],outp$tytz[i])
  outp$eyez[i]<-cor(data1$ey,data1$ez)
  outp$exey[i]<-cor(data1$ex,data1$ey)
  outp$exez[i]<-cor(data1$ex,data1$ez)
```

```
outp$txey[i]<-cor(data1$tx,data1$ey)
outp$txez[i]<-cor(data1$tx,data1$ez)
outp$tyex[i]<-cor(data1$ty,data1$ex)
outp$tyez[i]<-cor(data1$ty,data1$ez)
outp$tzex[i]<-cor(data1$tz,data1$ex)
outp$tzey[i]<-cor(data1$tz,data1$ey)
outp$txex[i]<-cor(data1$tx,data1$ex)
outp$tyey[i]<-cor(data1$ty,data1$ey)
outp$tzez[i]<-cor(data1$tz,data1$ez)
outp$rxx[i]<-var(data1$tx)/var(data1$ox)
outp$ryy[i]<-var(data1$ty)/var(data1$oy)
outp$rzz[i]<-var(data1$tz)/var(data1$oz)
outp$rxxc[i]<-var(data1$tx)/(var(data1$tx)+var(data1$ex))
outp$ryyc[i]<-var(data1$ty)/(var(data1$ty)+var(data1$ey))
outp$rzzc[i]<-var(data1$tz)/(var(data1$tz)+var(data1$ez))
outp$exx[i]<-var(data1$ex)/var(data1$ox)
outp$eyy[i]<-var(data1$ey)/var(data1$oy)
outp$ezz[i]<-var(data1$ez)/var(data1$oz)

outp$ztxty[i]<-zim(outp$oxoy[i],outp$rxx[i],outp$ryy[i],
    outp$exey[i],outp$exx[i],outp$eyy[i],outp$txey[i],outp$tyex[i])
outp$ztxtz[i]<-zim(outp$oxoz[i],outp$rxx[i],outp$rzz[i],
    outp$exez[i],outp$exx[i],outp$ezz[i],outp$txez[i],outp$tzex[i])
outp$ztytz[i]<-zim(outp$oyoz[i],outp$ryy[i],outp$rzz[i],
    outp$eyez[i],outp$eyy[i],outp$ezz[i],outp$tyez[i],outp$tzey[i])
outp$ztrtz[i]<-part(outp$ztxty[i],outp$ztxtz[i],outp$ztytz[i])

outp$stxty[i]<-spear(outp$oxoy[i],outp$rxxc[i],outp$ryyc[i])
outp$stxtz[i]<-spear(outp$oxoz[i],outp$rxxc[i],outp$rzzc[i])
outp$stytz[i]<-spear(outp$oyoz[i],outp$ryyc[i],outp$rzzc[i])
outp$strtz[i]<-part(outp$stxty[i],outp$stxtz[i],outp$stytz[i])

outp$wtxty[i]<-wet(outp$oxoy[i],outp$rxxc[i],outp$ryyc[i],outp$exey[i])
outp$wtxtz[i]<-wet(outp$oxoz[i],outp$rxxc[i],outp$rzzc[i],outp$exez[i])
outp$wtytz[i]<-wet(outp$oyoz[i],outp$ryyc[i],outp$rzzc[i],outp$eyez[i])
outp$wtrtz[i]<-part(outp$wtxty[i],outp$wtxtz[i],outp$wtytz[i])

for (j in 1:length(ss)){
  for (k in 1:iter){

    s<-sample(1:pops,ss[j],replace=F)
    datas<-data1[s,]
    datas<-data.frame(datas)
    outs$ss[k]<-ss[j]
    outs$k[k]<-k
    outs$id[k]<-i
    outs$oxoy[k]<-cor(datas$ox,datas$oy)
    outs$oxoz[k]<-cor(datas$ox,datas$oz)
    outs$oyoz[k]<-cor(datas$oy,datas$oz)
    outs$oroz[k]<-part(outs$oxoy[k],outs$oxoz[k],outs$oyoz[k])
    outs$txty[k]<-cor(datas$tx,datas$ty)
    outs$txtz[k]<-cor(datas$tx,datas$tz)
    outs$tytz[k]<-cor(datas$ty,datas$tz)
    outs$trtz[k]<-part(outs$txty[k],outs$txtz[k],outs$tytz[k])
    outs$eyez[k]<-cor(datas$ey,datas$ez)
    outs$exey[k]<-cor(datas$ex,datas$ey)
    outs$exez[k]<-cor(datas$ex,datas$ez)
```

```
outs$txey[k]<-cor(datas$tx,datas$ey)
outs$txez[k]<-cor(datas$tx,datas$ez)
outs$tyex[k]<-cor(datas$ty,datas$ex)
outs$tyez[k]<-cor(datas$ty,datas$ez)
outs$tzex[k]<-cor(datas$tz,datas$ex)
outs$tzey[k]<-cor(datas$tz,datas$ey)
outs$txex[k]<-cor(datas$tx,datas$ex)
outs$tyey[k]<-cor(datas$ty,datas$ey)
outs$tzez[k]<-cor(datas$tz,datas$ez)
outs$rxx[k]<-var(datas$tx)/var(datas$ox)
outs$ryy[k]<-var(datas$ty)/var(datas$oy)
outs$rzz[k]<-var(datas$tz)/var(datas$oz)
outs$rxxc[k]<-var(datas$tx)/(var(datas$tx)+var(datas$ex))
outs$ryyc[k]<-var(datas$ty)/(var(datas$ty)+var(datas$ey))
outs$rzzc[k]<-var(datas$tz)/(var(datas$tz)+var(datas$ez))
outs$exx[k]<-var(datas$ex)/var(datas$ox)
outs$eyy[k]<-var(datas$ey)/var(datas$oy)
outs$ezz[k]<-var(datas$ez)/var(datas$oz)

outs$ztxty[k]<-zim(outs$oxoy[k],outs$rxx[k],outs$ryy[k],
      outs$exey[k],outs$exx[k],outs$eyy[k],outs$txey[k],outs$tyex[k])
outs$ztxtz[k]<-zim(outs$oxoz[k],outs$rxx[k],outs$rzz[k],
      outs$exez[k],outs$exx[k],outs$ezz[k],outs$txez[k],outs$tzex[k])
outs$ztytz[k]<-zim(outs$oyoz[k],outs$ryy[k],outs$rzz[k],
      outs$eyez[k],outs$eyy[k],outs$ezz[k],outs$tyez[k],outs$tzey[k])
outs$ztrtz[k]<-part(outs$ztxty[k],outs$ztxtz[k],outs$ztytz[k])

outs$stxty[k]<-spear(outs$oxoy[k],outs$rxxc[k],outs$ryyc[k])
outs$stxtz[k]<-spear(outs$oxoz[k],outs$rxxc[k],outs$rzzc[k])
outs$stytz[k]<-spear(outs$oyoz[k],outs$ryyc[k],outs$rzzc[k])
outs$strtz[k]<-part(outs$stxty[k],outs$stxtz[k],outs$stytz[k])

outs$wtxty[k]<-wet(outs$oxoy[k],outs$rxxc[k],outs$ryyc[k],outs$exey[k])
outs$wtxtz[k]<-wet(outs$oxoz[k],outs$rxxc[k],outs$rzzc[k],outs$exez[k])
outs$wtytz[k]<-wet(outs$oyoz[k],outs$ryyc[k],outs$rzzc[k],outs$eyez[k])
outs$wtrtz[k]<-part(outs$wtxty[k],outs$wtxtz[k],outs$wtytz[k])

outs$btxty[k]<-outs$txty[k]-outp$txty[i]
outs$btxtz[k]<-outs$txtz[k]-outp$txtz[i]
outs$btytz[k]<-outs$tytz[k]-outp$tytz[i]
outs$btrtz[k]<-outs$trtz[k]-outp$trtz[i]

outs$boxoy[k]<-outs$oxoy[k]-outp$txty[i]
outs$boxoz[k]<-outs$oxoz[k]-outp$txtz[i]
outs$boyoz[k]<-outs$oyoz[k]-outp$tytz[i]
outs$boroz[k]<-outs$oroz[k]-outp$trtz[i]

outs$bexey[k]<-outs$exey[k]-outp$exey[i]
outs$bexez[k]<-outs$exez[k]-outp$exez[i]
outs$beyez[k]<-outs$eyez[k]-outp$eyez[i]

outs$brxx[k]<-outs$rxx[k]-outp$rxx[i]
outs$bryy[k]<-outs$ryy[k]-outp$ryy[i]
outs$brzz[k]<-outs$rzz[k]-outp$rzz[i]

outs$brxxc[k]<-outs$rxxc[k]-outp$rxx[i]
outs$bryyc[k]<-outs$ryyc[k]-outp$ryy[i]
```

```
outs$brzzc[k]<-outs$rzzc[k]-outp$rzz[i]

outs$bexx[k]<-outs$exx[k]-outp$exx[i]
outs$beyy[k]<-outs$eyy[k]-outp$eyy[i]
outs$bezz[k]<-outs$ezz[k]-outp$ezz[i]

outs$bztxty[k]<-outs$ztxty[k]-outp$txty[i]
outs$bztxtz[k]<-outs$ztxtz[k]-outp$txtz[i]
outs$bztytz[k]<-outs$ztytz[k]-outp$tytz[i]
outs$bztrtz[k]<-outs$ztrtz[k]-outp$trtz[i]

outs$bstxty[k]<-outs$stxty[k]-outp$txty[i]
outs$bstxtz[k]<-outs$stxtz[k]-outp$txtz[i]
outs$bstytz[k]<-outs$stytz[k]-outp$tytz[i]
outs$bstrtz[k]<-outs$strtz[k]-outp$trtz[i]

outs$bwtxty[k]<-outs$wtxty[k]-outp$txty[i]
outs$bwtxtz[k]<-outs$wtxtz[k]-outp$txtz[i]
outs$bwtytz[k]<-outs$wtytz[k]-outp$tytz[i]
outs$bwtrtz[k]<-outs$wtrtz[k]-outp$trtz[i]

outs$abtxty[k]<- abs(outs$btxty[k])
outs$abtxtz[k]<- abs(outs$btxtz[k])
outs$abtytz[k]<- abs(outs$btytz[k])
outs$abtrtz[k]<- abs(outs$btrtz[k])

outs$aboxoy[k]<-abs(outs$boxoy[k])
outs$aboxoz[k]<-abs(outs$boxoz[k])
outs$aboyoz[k]<-abs(outs$boyoz[k])
outs$aboroz[k]<-abs(outs$boroz[k])

outs$abexey[k]<-abs(outs$bexey[k])
outs$abexez[k]<-abs(outs$bexez[k])
outs$abeyez[k]<-abs(outs$beyez[k])

outs$abrxx[k]<-abs(outs$brxx[k])
outs$abryy[k]<-abs(outs$bryy[k])
outs$abrzz[k]<-abs(outs$brzz[k])

outs$abrxxc[k]<-abs(outs$brxxc[k])
outs$abryyc[k]<-abs(outs$bryyc[k])
outs$abrzzc[k]<-abs(outs$brzzc[k])

outs$abexx[k]<-abs(outs$bexx[k])
outs$abeyy[k]<-abs(outs$beyy[k])
outs$abezz[k]<-abs(outs$bezz[k])

outs$abztxty[k]<-abs(outs$bztxty[k])
outs$abztxtz[k]<-abs(outs$bztxtz[k])
outs$abztytz[k]<-abs(outs$bztytz[k])
outs$abztrtz[k]<-abs(outs$bztrtz[k])

outs$abstxty[k]<-abs(outs$bstxty[k])
outs$abstxtz[k]<-abs(outs$bstxtz[k])
outs$abstytz[k]<-abs(outs$bstytz[k])
outs$abstrtz[k]<-abs(outs$bstrtz[k])
```

```
      outs$abwtxty[k]<-abs(outs$bwtxty[k])
      outs$abwtxtz[k]<-abs(outs$bwtxtz[k])
      outs$abwtytz[k]<-abs(outs$bwtytz[k])
      outs$abwtrtz[k]<-abs(outs$bwtrtz[k])
    }
  f<-paste("outs",i,sep="")
  f<-paste(f,"ss",sep="")
  f<-paste(f,ss[j],sep="")
  f<-paste(f,".csv",sep="")
  write.csv(file=paste(path,f,sep=""),outs,row.names=FALSE)
  l<-(j-1)*nrow(sc)+i
  outmean[l,c(1:108)]<-mean(outs[1:108],na.rm=TRUE)
  outsd[l,c(1:108)]<-sd(outs[1:108],na.rm=TRUE)
  outmean[l,"id"]<-outsd[l,"id"]<-i
  outmean[l,"ss"]<-outsd[l,"ss"]<-ss[j]
  outmean[l,"cp2trtz"]<-nrow(subset(outs,btrtz > .2))
  outmean[l,"cn2trtz"]<-nrow(subset(outs,btrtz < -.2))
  outmean[l,"cnatrtz"]<-iter-length(na.omit(outs$trtz))

  outmean[l,"cp2oroz"]<-nrow(subset(outs,boroz > .2))
  outmean[l,"cn2oroz"]<-nrow(subset(outs,boroz < -.2))
  outmean[l,"cnaoroz"]<-iter-length(na.omit(outs$oroz))

  outmean[l,"cp2ztrtz"]<-nrow(subset(outs,bztrtz > .2))
  outmean[l,"cn2ztrtz"]<-nrow(subset(outs,bztrtz < -.2))
  outmean[l,"cnaztrtz"]<-iter-length(na.omit(outs$ztrtz))

  outmean[l,"cp2strtz"]<-nrow(subset(outs,bstrtz > .2))
  outmean[l,"cn2strtz"]<-nrow(subset(outs,bstrtz < -.2))
  outmean[l,"cnastrtz"]<-iter-length(na.omit(outs$strtz))

  outmean[l,"cp2wtrtz"]<-nrow(subset(outs,bwtrtz > .2))
  outmean[l,"cn2wtrtz"]<-nrow(subset(outs,bwtrtz < -.2))
  outmean[l,"cnawtrtz"]<-iter-length(na.omit(outs$wtrtz))
  }
}

write.csv(file=paste(path,"outp.csv",sep=""),outp,row.names=FALSE)
write.csv(file=paste(path,"outmean.csv",sep=""),outmean,row.names=FALSE)
write.csv(file=paste(path,"outsd.csv",sep=""),outsd,row.names=FALSE)


part<-function(rxy,rxz,ryz){
  if (rxy*rxy >= .9999) return(NA)
  rrz<-(ryz-(rxy*rxz))/sqrt(1-rxy*rxy)
  return(rrz)
}

zim<-function(roxoy,rxx,ryy,rexey,exx,eyy,rtxey,rtyex){
  ztxty<-roxoy/sqrt(rxx*ryy)-(rexey*sqrt(exx)*sqrt(eyy))/sqrt(rxx*ryy)-
rtxey*sqrt(eyy)/sqrt(ryy) -rtyex*sqrt(exx)/sqrt(rxx)
  return(ztxty)
}

wet<-function(roxoy,rxx,ryy,rexey){
  wtxty<-roxoy/sqrt(rxx*ryy)-(rexey*sqrt(1-rxx)*sqrt(1-ryy))/sqrt(rxx*ryy)
  return(wtxty)
```

```
}

spear<-function(roxoy,rxx,ryy){
  stxty<-roxoy/sqrt(rxx*ryy)
  return(stxty)
}
```

REFERENCES

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II manual*. San Antonio, TX: Psychological Corporation.

Beck, A. T. (1990). *Beck Anxiety Inventory (BAI)*. San Antonio, TX: Psychological Corporation.

Clark, D. A., Steer, R. A., Beck, A. T., & Ross, L. (1995). Psychometric characteristics of revised sociotropy and autonomy scales in college students. *Behavorial Research Therapy*, *33*, 325-334.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Corder-Bolz, C. R. (1978). The evaluation of change: New evidence. *Educational and Psychological Measurement*, *38*, 959-996.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth.

Fan, X. (2003). Two approaches for correcting correlation attenuation caused by measurement error: Implications for research practice. *Educational and Psychological Measurement*, *63*, 915-930.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, *Biometrika*, *10*, 507-521.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement & Evaluation in Counseling & Development*, *34*, 177-189.

Herting, J. R. (2002). Evaluating and rejecting true mediation models: A cautionary note. *Prevention Science*, *3*, 286-289.

Kieffer, K. M., Reese, R. J., & Thompson, B. (2001). Statistical techniques employed in *AERJ* and *JCP* articles from 1988 to 1997: A methodological review. *Journal of Experimental Education, 69*, 280-309.

Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research*, *47*, 121-150.

Lynam, D. R., Hoyle, R. H., & Newman, J. P. (2006). The perils of partialling: Cautionary tales from aggression and psychopathy. *Assessment*, *13*, 328-341.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding, and suppression effect. *Prevention Science*, *1*, 173-181.

Nimon, K., & Henson, R. (2010, May). *Validity of residualized variables after pretest covariance corrections: Still the same variable?* Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

Nimon, K., Zientek, L. R., & Henson, R. (2012). The assumption of a reliable instrument and other pitfalls to avoid when considering the reliability of data. *Frontiers in Psychology, 3*(102), 1–13.

Onwuegbuzie, A. J., Roberts, J. K., & Daniel, L. G. (2004). A proposed new "what if" reliability analysis for assessing the statistical significance of bivariate relationships. *Measurement and Evaluation in Counseling and Development, 37*, 228-239.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401.

Spearman, C. (1904). The proof and measurement of associations between two things. *American Journal of Psychology*, *15*, 72-101.

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Thompson, B. (1992). Misuse of ANCOVA and related "statistical control" procedures. *Reading Psychology: An International Quarterly*, *13*, iii-xviii.

Tracz, S. M., Nelson, L. L., Newman, I., & Beltran, A. (2005). The misuse of ANCOVA: The academic and political implications of Type VI errors in studies of achievement and socioeconomic status. *Multiple Linear Regression Viewpoints, 31*(1), 16-21.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.

Wang, Z., & Thompson, B. (2007). Is the Pearson r^2 biased, and if so, what is the best correction formula? *Journal of Experimental Education*, *75*, 109-125.

Ward, C., Fischer, R., Lam, F. S. Z, & Hall, L. (2009). The convergent, discriminant, and incremental validity of scores on a self-report measure of cultural intelligence. *Educational and Psychological Measurement*, *69*, 85-105.

Wetcher-Hendrick, D. (2006). Adjustments to the correction for attenuation. *Psychological Methods*, *11*, 207-215.

Williams, R. H., & Zimmerman, D. W. (1982). Comparative validity of simple and residualized difference scores. *Psychological Reports*, *50*, 91-94.

Winne, P. H. (1983). Distortions of construct validity in multiple regression analysis. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 15*, 187-202.

Winne, P. H., & Belfry, M. J. (1982). Interpretative problems when correcting for attenuation. *Journal of Educational Measurement*, *19*, 125-134.

Zientek, L. R., & Thompson, B. (2009). Matrix summaries improve research reports: Secondary analyses using published literature. *Educational Researcher*, *38*, 343-352.

Zimmerman, D. W. (2007). Correction for attenuation with biased reliability estimates and correlated errors in populations and samples. *Educational and Psychological Measurement*, *67*, 920-939.

Zimmerman, D. W., & William, R. H. (1977). The theory of test validity and correlated errors of measurement. *Journal of Mathematical Psychology*, *16*, 135-152.