5-1-2014

Bias and Precision of the Squared Canonical Correlation Coefficient Under Nonnormal Data Condition

Lesley F. Leach
Tarleton State University, leach@tarleton.edu

Robin K. Henson
University of North Texas

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Bias and Precision of the Squared Canonical Correlation Coefficient Under Nonnormal Data Conditions

Lesley F. Leach
Tarleton State University
Stephenville, TX

Robin K. Henson
University of North Texas
Denton, TX

Monte Carlo methods were employed to investigate the effect of nonnormality on the bias associated with the squared canonical correlation coefficient (R_c^2). The majority of R_c^2 estimates were found to be extremely biased, but the magnitude of bias was impacted little by the degree of nonnormality.

Keywords: canonical correlation coefficient, effect size, simulation, nonnormal, canonical correlation analysis

Introduction

Over the last several decades, the movement towards the use of effect size estimates in determining the importance of research results has intensified. This movement can be seen in the editorial policies of at least 25 educational and psychological journals (Wang & Thompson, 2007) that explicitly require the inclusion of effect sizes with statistical results. The sixth edition of the American Psychological Association Publication Manual (APA, 2001) deemed it "almost always necessary to include some measure of effect size" (p. 34) when reporting results. This shift has come with increased awareness that, when used alone to interpret results (i.e., without effect sizes or other statistics), *p*-values derived from null hypothesis significance tests (NHSTs) offer little information about the importance of results or their ability to replicate (Cumming, 2008; Henson & Smith, 2000; Kirk, 1996; Kline, 2004; Thompson, 1996, 1998). Effect size estimates offer "practical significance" information by quantifying the magnitude of a difference or relationship between variables. Consequently, numerous authors

Lesley F. Leach is an Associate Professor and Coordinator of the Educational Leadership Doctoral Program in the Department of Educational Leadership and Policy Studies. Email at: leach@tarleton.edu. Robin K. Henson is a professor of Educational Psychology in the College of Education.

and institutions have argued that effect sizes should be included with statistical results (e.g., APA, 2009, 2001; Henson, 2006; Thompson, 1996, 1998; Wilkinson & APA Task Force on Statistical Inference, 1999).

Although there are many types of effect sizes from which researchers can choose, most fall into two broad categories: (a) standardized mean difference effects and (b) measures of strength of association (Kirk, 1996; Kline, 2004; Olejnik & Algina, 2000; Onwuegbuzie, Levin, & Leech, 2003), although other statistics such as Huberty's I index certainly also qualify (Huberty & Lowman, 2000). Outside of the correlation coefficient, one of the most common effect sizes reported in the literature is R^2, likely due to the fact that R^2 is routinely provided as part of the regression output in most statistical packages (Kirk, 1996).

There has been some debate as to whether effect sizes should be included with all NHST results, even those that are not statistically significant or only for those tests that are statistically significant (Roberts & Henson, 2002; Robinson & Levin, 1997). Some researchers have gone as far as to recommend that hypothesis tests be banned entirely (e.g. Carver, 1993) and replaced with effect size estimates or other statistics (see also Harlow, Mulaik, Steiger, 1997, for a broader discussion). These views notwithstanding, there seems to be current consensus that effect sizes can add considerable value to research interpretation.

However, effect sizes are not without their limitations and can be "subject to as much abuse and misuse as are tests for statistical significance" (Onwuegbuzie, Levin, & Leech, 2003, p. 38; see also O'Grady, 1982; Robinson & Levin, 1997). Many researchers are unaware that effect size estimates can be criticized on some of the same grounds as NHSTs, including but not limited to the fact that effects can vary according to sample size and variability, and they are often impacted by the shape of the data, including departures from normality (Knapp & Sawilowsky, 2001; Onwuegbuzie & Levin, 2003; Onwuegbuzie, Levin, & Leech, 2003). As Henson (2006) noted, "If we fail to adequately understand what our effect sizes do and do not tell us, then we may fall victim to new misconceptions about our research methods" (p. 610).

R^2 Effect Size

For example, studies have shown r^2 (Wang & Thompson, 2007; Zimmerman, Zumbo, & Williams, 2003) and its analog, R^2, to often overestimate the effect found in the population (Carter, 1979; Fan, 2001; Larson, 1931; Snyder & Lawson, 1993; Thompson, 1999; Yin & Fan, 2001). By design, the ordinary least squares estimation method commonly used in regression analyses seeks to

maximize the correlation between variables resulting in the largest possible effect size. To obtain the greatest possible effect, the analyses capitalize on *all* the variance in a given sample, including the variance attributable to sampling error (Thompson & Kieffer, 2000; Wang & Thompson, 2007). Because the effect size accounts for variability unique to the sample – variance that is unlikely to be found in the population or future samples – the resulting R^2 is often a biased estimate of the effect in the population or in future samples (Roberts & Henson, 2002; Snyder & Lawson, 1993; Yin & Fan, 2001). Similar to the univariate application of R^2, studies have shown that the multivariate extension of R^2, the squared canonical correlation coefficient (R_c^2) can be positively biased due to the influence of sampling error as well (Thompson, 1990; Thorndike & Weiss, 1973).

Canonical Correlation Analysis

Like other multivariate methods, CCA has seen increased use in educational and psychological research, presumably due to its ability to limit experimentwise error rates and the fact that, by design, research studies using multivariate methods such as CCA often more accurately reflect the situations to which researchers wish to generalize (Fish, 1988; Henson, 1999; Sherry & Henson, 2005). Its primary purpose is to describe the relationship between synthetic composites of two sets of variables, although CCA can theoretically be extended to more than two variable sets.

     Like other parametric methods, CCA applies weights, called standardized canonical function coefficients, to observed variables to create synthetic variables. The measure of effect, or canonical correlation coefficient (R_c), is calculated as the simple bivariate correlation between the two synthetic variables (Campbell & Taylor, 1996; Henson, 2000; Sherry & Henson, 2005; Thompson, 1984, 1991). It is important to note that the goal of CCA is to maximize this correlation. It is in this optimization process, however, that sample-specific variation can become problematic because, although it was considered in determining the sample magnitude of effect, the sample-specific variance cannot be expected to exist in a new sample. Thus, one would anticipate a lower magnitude of effect in the population and/or replication with a new sample than R_c identifies. When squared, the canonical correlation (R_c^2) represents the proportion of variance that two synthetic CCA composites linearly share (Henson, 2000; Sherry & Henson, 2005; Thompson, 1984) and, in doing so, signifies the percentage of variability in the criterion variable set that can be explained with knowledge of the predictor variable set.

Bias in R_c Estimates

Empirical studies investigating bias in R_c estimates have found mixed results. Thorndike and Weiss (1973) first investigated the impact of sampling error on the canonical correlation using data from two different sources – clients of the Minnesota Division of Vocational Rehabilitation and data from Thorndike et al. (1968) (*N*=789 and 505, respectively). The analyses were split into two studies, and subjects from both sources were randomly split into two groups each (*n*=418 and 371 and *n*=246 and 259, respectively) for a total of four subgroups. The canonical correlations were compared to the cross-validated canonical correlations for each of the four subgroups. According to the authors, large differences demonstrated that sample-specific covariation could affect sample results.

Barcikowski and Stevens (1975) also investigated the effect of sampling error on the canonical correlation, but with results that differed from Thorndike and Weiss (1973). They selected 8 correlation matrices from the literature [two of which were from Thorndike and Weiss (1973)] and, using a procedure described by Huberty (1969), generated population matrices with the same properties as the selected datasets. Sample correlation matrices were generated from the population matrices, and canonical correlation analyses were performed. The number of variables ranged from 7 to 41, and the sample sizes ranged from 200-3000 in increments of 200; each sample size was replicated 100 times. The results indicated that the canonical correlations were "very stable under replication" (Barcikowski & Stevens, 1975, p. 362), even in the cases of small sample sizes (e.g., 100-200).

Thompson (1990) investigated bias in the canonical correlation that resulted in findings that conflicted with Barcikowski and Stevens (1975). Unlike the studies by Thorndike and Weiss (1973) and Barcikowski and Stevens (1975) that employed real and modeled data, respectively, Thompson used Monte Carlo methods to simulate data such that it met predetermined properties. A fully-crossed design was employed that varied the following conditions: (a) between-set correlations, (b) within-set correlations, (c) sample-size to variable ratios, and (d) variable sets. Sixty-four (i.e., 4 correlation matrices x 4 sample sizes x 4 variable sets) condition combinations were investigated. For each condition combination, 1,000 random samples were drawn and analyzed.

The ratio of subjects to variables emerged as the best predictor of bias in all six analyses (i.e., one for each of three R_c^2 deviation scores and three R_c^2 standard error deviations). Because the bivariate correlations between these values were positive, Thompson (1990) concluded that a greater number of subjects per

variable could potentially lead to less bias in $R_c{}^2$. Furthermore, Thompson found that, though the estimates of R_c were somewhat positively biased, the bias was minimal unless a ratio "as small as three to one" was used (p. 27). Finally, Thompson contended that even then the bias could be minimized in some situations if the value of R_c was moderate to large (e.g., greater than .40).

Thompson (1990) only examined the accuracy of $R_c{}^2$ when the multivariate normality assumption of CCA was met. Whereas the normality assumption is formally required only when testing the statistical significance of canonical results (Marascuilo & Levin, 1983; Sherry & Henson, 2005; Thompson, 1984), when normality is not met, distribution shapes must still be reasonably comparable. If not, entries in the matrix of association used to derive canonical estimates may be attenuated, which could compromise the results including the magnitude of the effect (Thompson, 1984). Studies have shown, however, that few educational and psychological datasets are exactly normally distributed (Blair, 1981; Bradley, 1968, 1982; Micceri, 1989; Pearson & Please, 1975) and, as such, there is a need to investigate the performance of CCA under nonnormal data conditions to inform the use of CCA in applied studies.

Purpose of the Study

As a result of the equivocal prior findings and the lack of investigation of nonnormal distributional conditions, this study compared the degree of bias associated with the squared canonical correlation coefficient ($R_c{}^2$) gained from distributions possessing varying degrees of nonnormality to that found with multivariate normal distributions. Additional study factors were included to explore potential bias in this multivariate effect size across common conditions and to allow comparison with prior studies. Monte Carlo simulation methodology was used to fulfill this purpose.

Methodology

Design

A fully-crossed design was employed in this study, manipulating the following conditions: (a) distribution shape, (b) variable sets, (c) sample sizes, (d) correlation matrices with varied between- and within-set correlations. See Table 1 for the conditions and their respective levels. Six distribution shapes were investigated, as well as 4 variable sets, 4 sample sizes, and 7 correlation matrices

(manipulating both the between- and within-set correlations), resulting in a total of 672 manipulated conditions. Five-thousand samples were drawn for each condition for a total of 3,360,000 canonical analyses

Table 1: Summary of Data Conditions Manipulated in the Study

Data condition		Levels Manipulated	
Distribution shape		$k = -1, 0, 1, 3, 5, 8$	
Variable Sets		6 + 6 (v=12) 4 + 4 (v=8) 4 + 2 (v=6) 10 + 2 (v=12)	
Sample size: variable ratio		3:1, 10:1, 25:1, 40:1	
	Matrix	Between-set correlation	Within-set correlation
	A	0	0
	B	.1 (small)	.3 (moderate)
	C	.1 (small)	.5 (large)
Correlation matrices	D	.3 (moderate)	.3 (moderate)
	E	.3 (moderate)	.5 (large)
	F	.5 (large)	.3 (moderate)
	G	.5 (large)	.5 (large)

Note. k denotes univariate kurtosis. The various variable sets are denoted in the following manner: no. of variables in the predictor set + no. of variables in the criterion set (total number of variables in both sets).

Multivariate normality` The shapes of the distributions were manipulated to facilitate comparison of results under normal theory to those found under multivariate nonnormal data conditions. Specifically, this study examined the impact of varying levels of kurtosis (k) on the squared canonical correlation coefficient. Five multivariate nonnormal datasets were generated such that all marginal distributions in each dataset possessed the following levels of univariate kurtosis: (a) negligible kurtosis ($k = -1, 0, 1, 3$) and (b) moderate kurtosis ($k = 5, 8$). These value ranges are consistent with studies investigating the effect of nonnormality on other sample statistics (e.g., Curran, West, & Finch, 1996; Olsson, Foss, Troye, & Howell, 2000).

It is unrealistic to expect that multivariate datasets seen in practical applications would typically possess equal univariate kurtoses across the marginal

distributions (Yuan & Bentler, 1997). But, for the sake of clarity and ease of interpretation, this procedure was used in this study study as it has been in past investigations (e.g., Curran, West, & Finch, 1996; Fouladi, 2000; Nevitt & Hancock, 2001; Olsson, Foss, Troye, & Howell, 2000). The results from these nonnormal distributions were compared to those from a multivariate normal distribution. Because tests of variances and covariances (e.g., CCA) in normal distributions have been found to be more affected by kurtosis than skewness (Mardia, Kent, & Bibby, 1979), skewness was held constant at symmetrical (i.e., skewness = 0).

Variable sets   We incorporated the following variable sets [denoted as the number of variables in the predictor variable set + the number in the dependent set]: (a) 6+6 (v=12), (b) 4 + 4 (v=8), (c) 4 + 2 (v=6), (d) 10 + 2 (v=12). These sets replicate the variable sets used by Thompson (1990) and represent sets that one would likely see in behavioral studies.

Sample size to variable ratios   Sample size to variable ratios of 3, 10, 25, and 40 per variable were chosen to represent those likely seen in behavioral research. They are consistent with other studies investigating the accuracy of canonical correlation results (see, for example, Thompson, 1990).

Correlation matrices   Six combinations of small, moderate, and large within- and between-set correlations made up the population correlation matrices in addition to a "null" model with all correlations equal to zero. Cohen's (1988) conventions for values of r to correspond to his d benchmarks were used to determine the entries in the correlation matrix (r =.1, .3, and .5 indicating small, medium, and large effects, respectively).

It is important to note that the benchmarks provided by Cohen (1988) were not intended to be used as rigid criteria for determining result importance. Effects should always be considered in the context of the study from which they result as well as the broader literature to determine if they indicate a small, moderate, or large effect. In this article, we use the wording *small*, *moderate*, and *large* only to refer to the various effects; our choice of wording does not indicate that the various magnitudes will always represent small, moderate, and large effects, respectively. Furthermore, Cohen's effect size rules of thumb were originally presented for use in univariate contexts. In multivariate contexts, one could conceivably expect larger effects as a result of the additional variance made available for prediction by multiple dependent variables. There is little research to

support these guidelines for multivariate outcomes, however, so the univariate approximations were used in this study.

Varying combinations of between- and within-set correlations were used to define the correlation matrices (excluding the null model with $r_b=0$ and $r_w=0$). Within-set correlations were limited to moderate (.3) and large (.5) correlations because, in a typical CCA analysis, one would often expect the correlations within the variable sets to be moderately, if not highly, correlated. Between-set correlations would likely possess a wider range, and, as such, we chose to use small (.1), medium, and large correlations in this study. The combinations for the various population correlation matrices are presented in Table 1.

Data Generation and Analysis

Populations of data were randomly generated that mirrored the correlation matrices at the kurtosis levels previously specified. A total of 42 multivariate populations ($N=100,000$ each) were created (i.e., all paired combinations of the 6 distribution shapes [1 multivariate normal and 5 kurtotic] and the 7 correlation matrices). See Appendix A for information regarding the data generation procedure.

Sample canonical analyses were performed using SAS® (SAS Institute, Inc., Cary, NC, www.sas.com) version 9.1.3 syntax. The variance explained (R_c^2) for each of the first three canonical functions was computed. The accuracy of R_c^2 was then calculated as the difference between the sample R_c^2 and population R_c^2 values. The average level of accuracy, or bias, of the R_c^2 estimates was calculated as the mean of the accuracy values for each condition combination, and the precision of the R_c^2 estimates was represented by the standard deviation of the respective accuracy values. Bias was considered to be extreme if it exceeded $\pm.30$ R_c^2; bias was considered to be minimal (and thus acceptable) if it was less than or equal to $\pm.30$ R_c^2.

Analysis of variance (ANOVA) was used to identify the influence of each condition on the variability of the accuracy values. The accuracy – i.e., the differences between the population R_c^2s and the sample R_c^2s – acted as the dependent variable (DV) whereas the four conditions made up the independent variables. Only main effects were considered in this study. Main effects were evaluated based on statistical significance of the F tests (alpha<.02 - value determined using the Bonferroni correction) as well as from η^2 and ω^2 effect size values (ω^2 was included as a theoretical adjustment for sampling error).

Results

Accuracy and Precision of R_c^2

Results from the first three functions were analyzed for each correlation matrix. All function II and III results were found to be extremely biased across all correlation combinations. For this reason and for the sake of brevity, only results from the first functions are reported and discussed in the present article. Second and third function results are available from the authors upon request.

The bias and precision (*SD*) of the sample R_c^2 accuracy values for correlation matrices **A** through **G** are presented in Tables A1 through A7 in Appendix B, respectively. Note that all of the condition combinations for correlation matrices **A** and **C** produced extremely biased R_c^2 accuracy values. Likewise, all but two of the combinations (97.92% of 96 cases) for correlation matrix **B** produced R_c^2 values that were extremely biased.

Correlation matrix **D** produced extremely biased accuracy values in only 22.92% of the 96 condition combinations. All condition combinations for correlation matrix **D** with sample size to variable ratios greater than or equal to 10:1 produced minimal amounts of bias. Conversely, the majority of the bias (91.67% of 21 cases) with a sample size to variable ratio of 3:1 were found to be extreme. Only two of the 3:1 *n:v* ratio condition combinations produced minimal bias; all other cases met the criteria to be considered extremely biased.

Similar results were found with the correlation matrix **E** results. In this case, 43.75% of the 96 condition combinations produced extreme bias. As a general rule, the condition combinations that were found to possess minimal levels of bias had sample size to variable ratios greater than or equal to 25:1. Unlike the other correlation matrices, results from all condition combinations in correlation matrices **F** and **G** were found to contain minimal bias.

The average bias and precision values by the various condition levels are presented in Table 2. As demonstrated in the table, bias generally decreased as the sample size to variable ratio increased. The most dramatic decrease in bias was seen in the difference in bias between the sample sizes of 3:1 and 10:1 (mean difference of .19). Differences between subsequent sample size to variable ratios were comparatively small. Bias values varied across the other condition combinations.

Table 2: Descriptive Statistics for Function I Bias by Sample Size to Variable Ratio, Variable Set, Univariate Kurtosis Level, and Correlation Matrix

Condition	*M*	*SD*	*n*
Sample size: variable ratio (*n:v*)			
3:1	0.26	0.17	840,000
10:1	0.07	0.07	840,000
25:1	0.03	0.04	840,000
40:1	0.02	0.03	840,000
Variable Set			
6 + 6	0.10	0.14	840,000
4 + 4	0.10	0.14	840,000
4 + 2	0.09	0.01	840,000
10 + 2	0.09	0.13	840,000
Expected Kurtosis (*k*)			
-1	0.09	0.14	560,000
0	0.09	0.13	560,000
1	0.09	0.14	560,000
3	0.09	0.14	560,000
5	0.10	0.14	560,000
8	0.10	0.14	560,000
Correlation Matrix			
A ($r_b = 0$, $r_w = 0$)	0.15	0.16	480,000
B ($r_b = .1$, $r_w = .3$)	0.13	0.16	480,000
C ($r_b = .1$, $r_w = .5$)	0.14	0.16	480,000
D ($r_b = .3$, $r_w = .3$)	0.07	0.11	480,000
E ($r_b = .3$, $r_w = .5$)	0.10	0.13	480,000
F ($r_b = .5$, $r_w = .3$)	0.03	0.06	480,000
G ($r_b = .5$, $r_w = .5$)	0.04	0.08	480,000

The precision of results, or standard deviation of the accuracy values, appeared to increase (i.e., the *SD* value decreased) as the sample size to variable ratio increased. Although decreased standard errors would be expected with

increased sample size, a dramatic difference in precision was detected between results with a sample size to variable ratio of 3:1 versus results from an *n:v* ratio of 10:1 (difference of .10).The precision of results was varied across variable sets. The (6+6), (4+4), and (10+2) variable sets produced results with roughly equal amounts of precision (*SD*=.14, .14, and .13, respectively). But the (4+2) variable set saw extremely precise results overall (*SD*=.01). Normality (or nonnormality) of the distributions seemed to matter little in the precision of results. The results had approximately the same precision regardless of the value of kurtosis (ranged from .13-.14). However, the precision of results varied by correlation matrix. Correlation matrices **F** and **G** saw greater precision (.06 and .08, respectively) than matrices **A**, **B**, **C**, **D**, and **E** (ranged from .11-.16). Because correlation matrices **F** and **G** had higher between- and within-set correlations, these results suggest that higher between- and within-set correlations may influence the precision of R_c^2 estimates. But these results should be taken tentatively because they are based on descriptive analyses alone; further exploration is needed.

Explanation of Variability in R_c^2 Bias

An analysis of variance (ANOVA) was run to determine which of the study factors could account for the variability in the accuracy values. The ANOVA summary table for the function I results can be found in Table 3.

Table 3: ANOVA Summary Table for Explanation of the Sources of Variation in Function I R_c^2 Bias

| Source of Variation	SS	df	MS	F	p	η^2	ω^2
Expected univariate kurtosis (k)	24.90	5	4.98	702.0	<.001	<.001	<.001
Sample size: variable ratio (n:v)	31776.88	3	10592.29	1493068.8	<.001	0.51	0.51
Variable set	8.14	3	2.71	382.3	<.001	<.001	<.001
Correlation matrix	7060.75	6	1176.79	165878.3	<.001	0.11	0.11
Error	23836.75	3359982	0.01				
Total	62707.42	3359999					

The Levene's test for homogeneity of variance was statistically significant as might be expected given the large number of simulated conditions, *F*(671, 3359328) = 2413.64, *p* < .001. Upon visual inspection of the variances by condition, we determined that the variances were roughly homogenous and

therefore most likely met the assumption of homogeneity of variance in this balanced design. Furthermore, the equal variances assumption is primarily related to the Type I error rate involved with the F tests. Because statistical significance of the ANOVA results was not our primary interest, meeting this assumption was less of a concern for this study.

All five conditions produced statistically significant main effects with $p<.001$ in all cases. It is apparent from examination of the η^2 values, however, that not all of the main effects were noteworthy. The sample size to ratio variable explained the greatest amount of variation in function I bias, accounting for 51% ($\eta^2=.51$; $\omega^2=.51$) of the variation in the DV. The only other condition that had a somewhat notable effect on the DV was the correlation matrix, and it displayed a considerably weaker relationship with the DV than the sample size to variable ratio. The correlation matrix variable explained 11% ($\eta^2=.11$; $\omega^2=.11$) of the variation in function I bias.

Based on these results, it is apparent that, of the five conditions manipulated in this study, the sample size to variable ratio had the largest effect on function I bias (depicted in Figure 1). And, this effect was considerable given the fact that it could explain approximately half of the function I bias variation. It is worth noting that the 3:1 sample size to variable ratio had, by far, the greatest bias of all the ratios ($M=.26$, $SD=.17$), with less bias for the 10:1, 25:1, and 40:1 ratios ($M=.07$, $SD=.07$; $M=.03$, $SD=.04$; and $M=.02$, $SD=.03$; respectively). Larger sample size to variable ratios seemed to help decrease bias in R_c^2, particularly when $n{:}v \geq 10{:}1$. The correlation matrix variable demonstrated a small, but still noteworthy effect in comparison (depicted in *Figure 2*). The correlation matrices with larger between- and within-set correlations – correlation matrices **F** and **G** ($r_b=.43$, $r_w=.50$ and $r_b=.50$, $r_w=.50$, respectively) – had less bias than matrices **A**, **B**, **C**, **D**, and **E**, leading to the conclusion that larger correlations may help decrease bias in R_c^2.
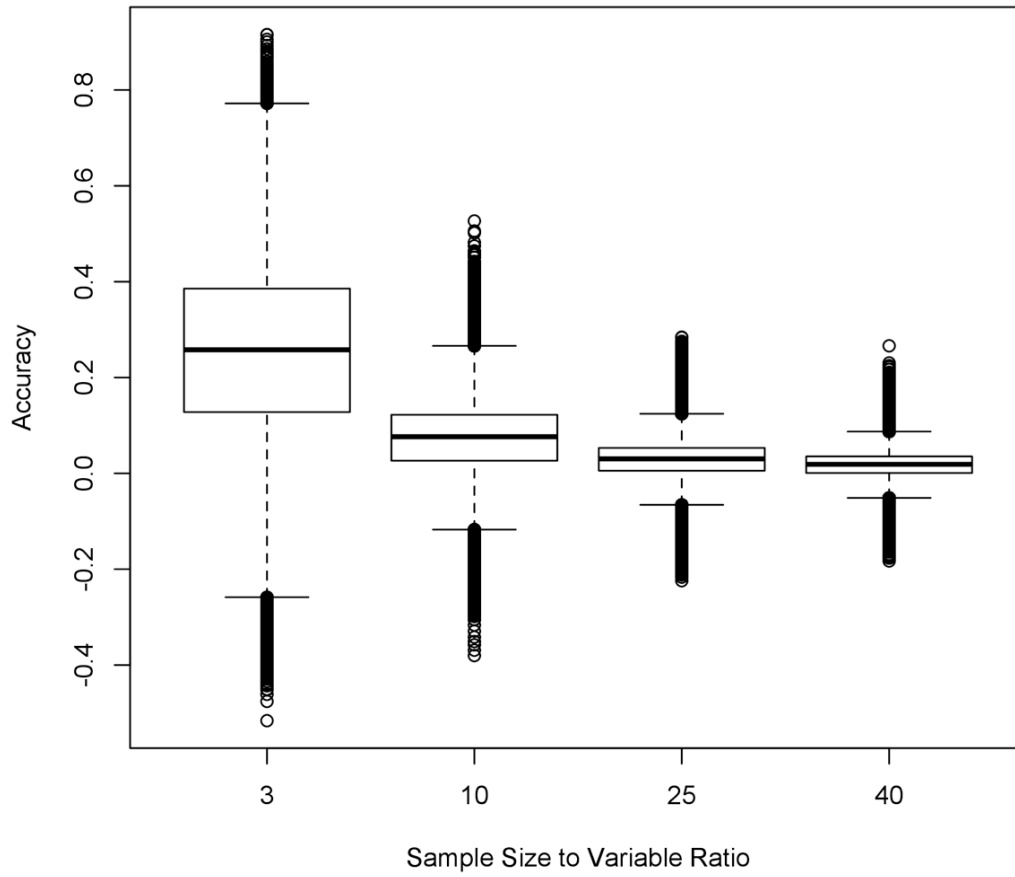
Figure 1: Boxplot of function I R_c^2 bias by sample size to variable ratio across all other conditions (N=3,360,000; n=840,000).
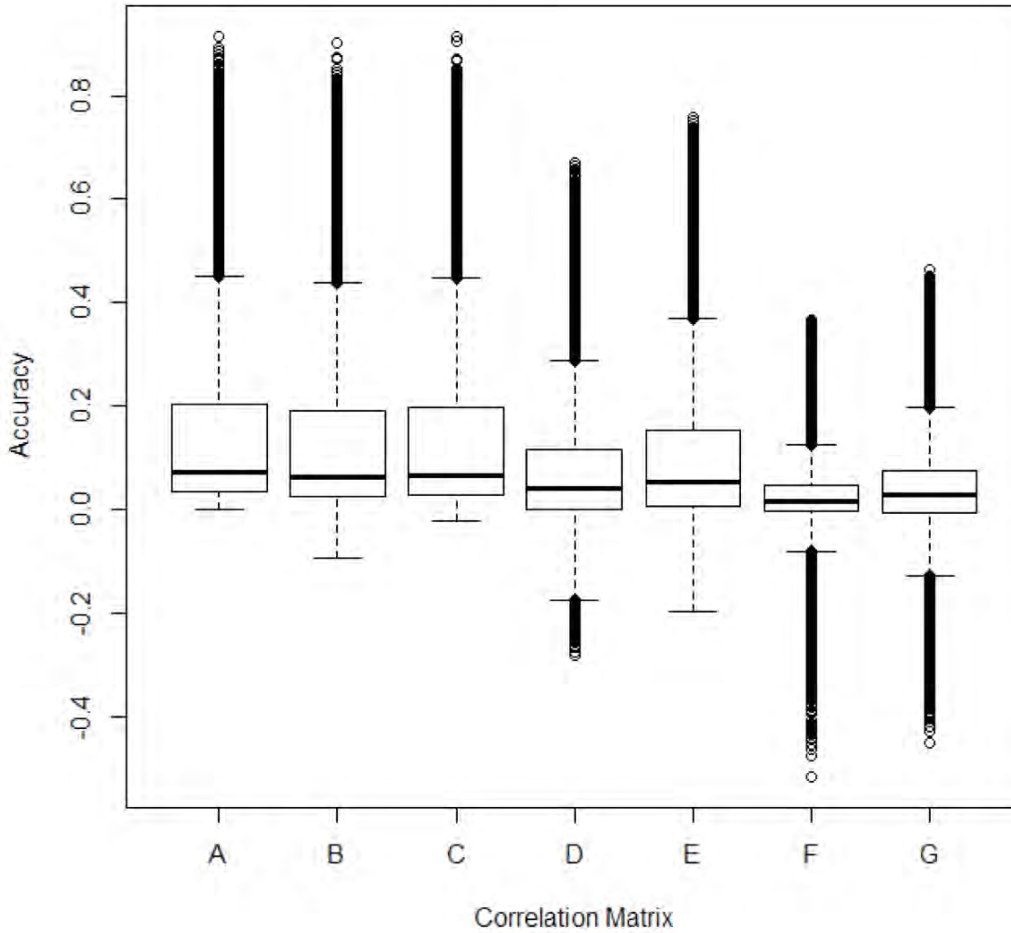
Figure 2: Boxplot of function I R_c^2 bias by correlation matrix across all other conditions (N=3,360,000; n=480,000). Correlation matrix **A** was created with r_b= 0 and r_w = 0, correlation matrix **B** with r_b = .1 and r_w = .3, correlation matrix **C** with r_b = .1 and r_w = .5, correlation matrix **D** with r_b = .3 and r_w = .3, correlation matrix **E** with r_b = .3 and r_w = .5, correlation matrix **F** with r_b = .5 and r_w = .3, and correlation matrix **G** with r_b = .5 and r_w = .5.

Conclusion

Overall, a large percentage of the first function results (47.92% of 672 total combinations) across correlation matrices provided minimal amounts of bias. With the exceptions of two minimal mean differences for the normal distribution of correlation matrix **B** data (i.e., the matrix with r_b=.10 and r_w=.30), all cases

with minimal bias were found with the correlation matrices reporting combinations of between-set and within-set correlations that were each greater than or equal to .30 (i.e., correlation matrices **D**, **E**, **F**, and **G**).

As demonstrated by the ANOVA results, the accuracy of R_c^2 was largely impacted by the sample size to variable ratio. This can be seen in the bias values; as the sample size to variable ratio increased, bias consistently decreased. This finding is not surprising given the impact of both sample size and the number of variables on the theoretical amount of sampling error present. As sample size increases, sampling error theoretically decreases. The number of variables in a model typically has the opposite effect on sampling error; as the number of variables increases, so does the theoretical amount of sampling error present in a sample. It logically follows that a larger sample size to variable ratio would likely help decrease the amount of sampling error (i.e., bias) in canonical results.

The current results speak directly to the sample size needed to obtain reasonable outcomes from CCA analyses. It is apparent from the descriptive statistics in Table 2 that, across all condition combinations, the 3:1 sample size to variable ratio produced substantial bias; bias dramatically decreased when a sample size to variable ratio of at least 10:1 was used. Larger sample sizes (e.g., 25:1 and 40:1) produced even less biased results, further demonstrating the fact that larger sample size to variable ratios are ideal conditions for decreasing bias in R_c^2. This finding echoes that found by Thompson (1990), who found the sample size to variable ratio to be the best predictor of bias in the squared canonical correlation coefficient. Furthermore, Thompson found that, though the estimates of R_c were somewhat positively biased, the bias was minimal unless a ratio of "as small as three to one" was used (p. 27). This was also the case in this study. Dramatic decreases in bias were seen between the sample size to variable ratios of 3:1 and 10:1. Thompson pointed out, however, that, with a small sample size to variable ratio (e.g., $n{:}v{=}3{:}1$), the bias could be minimized in some situations if the value of R_c was moderate to large. As can be seen across the matrices, this was somewhat true for the data with higher between- and within-set correlations in this study.

None of the other conditions, including the marginal kurtosis level, notably impacted the accuracy of R_c^2. These results mirror Barcikowski and Stevens (1975) and Thompson (1990) that involved normal distributions. Keep in mind, however, that this study was limited to negligibly and moderately kurtotic distributions; data with more extreme kurtosis could have a differential effect on R_c^2 estimates.

In this study, precision of the R_c^2 values was examined only through descriptive analyses (i.e., standard deviation of the differences between the population and sample R_c^2 values) because there is only one value in each cell for all replications. Although conclusions are limited as a result of the descriptive analyses, some general comments about the precision of R_c^2 can be made. As with the accuracy of the R_c^2 estimates, it appears that the precision of R_c^2 may increase as the $n{:}v$ ratio increases. This is logical given the effect of larger samples and fewer variables on sampling error. When sample size is maximized and the number of variables are minimized, a greater $n{:}v$ ratio will likely produce more precise R_c^2 results. The pattern of results by the variable sets is somewhat unclear and needs further investigation. More often than not, the (6 + 6) and (10 + 2) produced the most precise results. And, for correlation matrices **A**, **B**, **C**, **D**, and **E**, the precision values generally remained the same for the various marginal kurtosis levels. Matrices **F** and **G** with higher between- and within-set correlations saw greater precision by comparison. But, because these results are based on descriptive analyses alone, these results should be taken tentatively and should likely only be used to inform future studies.

Recommendations for Practice

Based on the results of the study, several recommendations are warranted in the use of canonical correlation analyses in educational and psychological research. First, it is recommended that a sample size to variable ratio of at least 10:1 be used in CCA analyses to lessen the bias that may affect R_c^2 results. As was seen in the descriptive statistics presented in Table 2, under these study conditions, using an $n{:}v$ ratio of 10:1 versus 3:1 led to dramatic reductions in bias. It would not be unlikely to expect similar results in applied studies under similar conditions.

Greater sample size to variable ratios may also provide more precise results as well. Because larger sample size to variable ratios reduce bias even more, however, researchers are encouraged to use the largest sample that is available to them and the fewest variables that will adequately represent their model. Maximizing the sample size and minimizing the number of study variables will help to increase the $n{:}v$ ratio and subsequently likely reduce bias and increase precision in the results.

Second, because the univariate kurtosis level was shown to not substantially impact results, researchers can be relatively confident that, when k is homogenous across variables and within the range of -1 to 8, R_c^2 bias is not likely to be greater or less than that that would be found with results from a normal distribution.

Therefore, it is reasonable that multivariate distributions that consist of moderately kurtotic univariate distributions can be treated in the same manner as normal distributions. These results may be encouraging to applied researchers given the fact that in practice, educational and psychological distributions are rarely exactly normally distributed (Blair, 1981; Bradley, 1968, 1982; Micceri, 1989; Pearson & Please, 1975). Resulting R_c^2 values are likely to be accurate in cases even with data that are moderately kurtotic.

Recommendations should be heeded with the limitations of the study in mind, however. Because the data were simulated, we were not able to model every conceivable condition that could impact the squared canonical correlation coefficient. Further research could seek to extend this study with a larger range of population effect sizes, sample sizes, distributional shapes, and numbers of variables.

Despite its limitations, the findings from this study revealed important conditions to consider in the use of the squared canonical correlation coefficient, particularly under nonnormal data conditions. These findings and recommendations are meant to impact research practice and provide more accurate applications of canonical correlation analysis, particularly as regards the use of the squared canonical correlation coefficient.

References

American Psychological Association. (2001). *Publication Manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.

American Psychological Association. (2009*). Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Barcikowski, R. S., & Stevens, J. P. (1975). A Monte Carlo study of the stability of canonical correlations, canonical weights, and canonical variate-variable correlations. *Multivariate Behavioral Research, 10*: 353-364.

Blair, R. C. (1981). A reaction to "Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance." *Review of Educational Research, 51*: 499-507.

Bradley, J. V. (1968). *Distribution free statistical tests.* Englewood Cliffs, NJ: Prentice-Hall.

Bradley, J. V. (1982). The insidious L-shaped distribution. *Bulletin of the Psychonomic Society, 20*: 85-88.

Campbell, K. T., & Taylor, D. L. (1996). Canonical correlation analysis as a general linear model: A heuristic lesson for teachers and students. *The Journal of Experimental Education, 64*: 157-172.

Carter, D. S. (1979). Comparison of different shrinkage formulas in estimating the population multiple correlation coefficients. *Educational and Psychological Measurement, 39*: 261-266.

Carver, R. P. (1993). The case against statistical significance testing, revisited. *The Journal of Experimental Education, 61*: 287-292.

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

Cumming, G. (2008). Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science, 3*: 286-300.

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

Fan, X. (2001). Statistical significance and effect size in education research: Two sides of a coin. *The Journal of Educational Research, 94*: 275–282.

Fan, X., Felsővályi, Á., Sivo, S.A., Keenan, S.C. (2002). *SAS for Monte Carlo studies: A guide for quantitative researchers*. Cary, NC: SAS Institute.

Fish, L. J. (1988). Why multivariate methods are usually vital. *Measurement and Evaluation in Counseling and Development, 21*: 130-137.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*: 521-532.

Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling, 7*: 356-410.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum.

Henson, R. K. (1999). Multivariate normality: What is it and how is it assessed? In B. Thompson (Ed.), *Advances in Social Science Methodology*, *Vol. 5* (pp.193-211). Stamford, CT: JAI Press.

Henson, R. K. (2000). Demystifying parametric analyses: Illustrating canonical correlation analysis as the multivariate general linear model. *Multiple Linear Regression Viewpoints, 26*: 11-19.

Henson, R. K. (2006). Effect size measures and meta-analytic thinking in counseling psychology research. *The Counseling Psychologist, 34*: 601-629.

Henson, R. K., & Smith, A. D. (2000). State of the art in statistical significance and effect size reporting: A review of the APA Task Force Report and current trends. *Journal of Research and Development in Education, 33*: 285-296.

Huberty, C. J. (1969). *An empirical comparison of selected classification rules in multiple discriminant analysis*. Unpublished doctoral dissertation, University of Iowa, Iowa City.

Huberty, C. J., & Lowman, L. L. (2000). Group overlap as a basis for effect size. *Educational and Psychological Measurement, 60*: 543-563.

Kaiser, H. F., & Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika, 27*: 179-182.

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

Knapp, T. R., & Sawilowsky, S. S. (2001). Constructive criticisms of methodological and editorial practices. *The Journal of Experimental Education, 70*: 65-79.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, D.C.: American Psychological Association.

Larson, S. C. (1931). The shrinkage of the coefficient of multiple correlation. *Journal of Educational Psychology, 22*: 45-55.

Marascuilo, L. A., & Levin, J. R. (1983). *Multivariate statistics in the social sciences: A researcher's guide*. Monterey, CA: Brooks/Cole.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. London: Academic Press.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*: 156-166.

Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling, 8*: 353-377.

O'Grady, K. E. (1982). Measures of explained variance: Cautions and limitations. *Psychological Bulletin, 92*: 766-777.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*: 241-286.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling, 7*: 557-595.

Onwuegbuzie, A. J., & Levin, J. R. (2003). Without supporting statistical evidence, where would reported measures of substantive importance lead? To no good effect. *Journal of Modern Applied Statistical Methods, 2*: 133-151.

Onwuegbuzie, A. J., Levin, J. R., & Leech, N. L. (2003). Do effect-size measures measure up?: A brief assessment. *Learning Disabilities: A Contemporary Journal, 1*(1): 37-40.

Pearson, E. S., & Please, N. W. (1975). Relation between the shape of population distribution and the robustness of four simple test statistics. *Biometrika, 62*: 223-241.

Roberts, J. K., & Henson, R. K. (2002). Correction for bias in estimating effect sizes. *Educational and Psychological Measurement, 62*: 241-253.

Robinson, D. H., & Levin, J. R. (1997). Reflections on statistical and substantive significance, with a slice of replication. *Educational Researcher, 26*(5): 21-26.

Sherry, A., & Henson, R. K. (2005). Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of Personality Assessment, 84*: 37-48.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education, 61*: 334-349.

Thompson, B. (1984). *Canonical correlation analyses: Uses and interpretation.* Thousand Oaks, CA: Sage.

Thompson, B. (1990). Finding a correction for the sampling error in multivariate measures of relationship: A Monte Carlo study. *Educational and Psychological Measurement, 50*: 15-31.

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. *Measurement and Evaluation in Counseling and Development, 24*(2): 80-96.

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement, 55*: 525-534.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 25*(2): 26-30.

Thompson, B. (1998, April). *Five methodology errors in educational research: The pantheon of statistical significance and other faux pas*. Paper presented at the annual meeting of the American Educational Research Association, San Diego.

Thompson, B. (1999). If statistical significance tests are broken/misused, what practices should supplement or replace them? *Theory & Psychology, 9*: 165–181.

Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: Strong arguments move the field. *The Journal of Experimental Education, 70*: 80-93.

Thompson, B., & Kieffer, K. M. (2000). Interpreting statistical significant test results: A proposed new "what if" method. *Research in the Schools, 7*(2): 3-10.

Thorndike, R. M., & Weiss, D. J. (1973). A study of the stability of canonical correlations and canonical components. *Educational and Psychological Measurement, 36*: 861-878.

Thorndike, R. M., Weiss, D. J., & Davis, R. V. (1968). Multivariate relationships between a measure of vocational interests and a measure of vocational needs. *Journal of Applied Psychology, 52*: 491-496.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51*: 473-481.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika, 48*, 465-471.

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

Wilkinson, L., & American Psychological Association (APA) Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*: 594-604.

Yin, P., & Fan, X. (2001). Estimating R^2 shrinkage in multiple regression: A comparison of different analytical methods. *Journal of Experimental Education, 69*: 203-224.

Yuan, K. H., & Bentler, P. M. (1997). Generating multivariate distributions with specified marginal skewness and kurtosis. In W. Bandilla and F. Faulbaum (Eds.), *SoftStat' 97-Advances in Statistical Software 6* (pp. 385-391). Stuttgart, Germany: Lucius and Lucius.

Zimmerman, D. W., Zumbo, B. D., & Williams, R. H. (2003). Bias in estimation and hypothesis testing of correlation. *Psicológica, 24*: 133-158.

Appendix A

Procedures presented by Fleishman (1978), Kaiser and Dickman (1962), Vale and Maurelli (1983) were used to generate the multivariate random distributions in this study. More extreme values of kurtosis (e.g., $k = 15, 25$) were considered, but rejected because the data generation procedure could not produce distributions that contained the desired levels of nonnormality.

Forty-two populations (i.e., one for each combination of the 6 kurtosis levels and 7 correlation matrices) were generated, and sample canonical analyses were performed using SAS® (SAS Institute, Inc., Cary, NC, www.sas.com) version 9.1.3 syntax. It is important to note that correlation matrix **F** was intended to have $r_b=.3$ and $r_w=.5$, but the resulting matrix was not of full rank. For that reason, we generated correlation matrix **F** to have $r_b=.43$ and $r_w=.5$, the correlations that were the closest to the intended values that would generate a matrix of full rank. The syntax was written by the authors using the reference by Fan, Felsővályi, Sivo, and Keenan (2002). For the sake of brevity, the syntax was not included in this article; copies can be obtained from the authors.

Several checks were incorporated into the SAS® code to insure its accuracy. First, the variables in each of the populations were checked to make sure that they truly approximated the pre-specified correlations and kurtosis levels. Second, values of the condition variables were saved for each of the 3,360,000 canonical analyses so that they could be compared with the expected values. Third, calculations for randomly selected cases were manually checked to verify their accuracy.

Appendix B

Table A1: Bias and Precision of Function I Sample R_c^2 Values for Correlation Matrix **A**

Var. Set	n: v	Expected k					
		-1	**0**	**1**	**3**	**5**	**8**
v=12 (6+6)	3.1	**.45**[a] (.09)[b]	**.45** (.09)	**.45** (.09)	**.45** (.09)	**.45** (.09)	**.45** (.10)
	10.1	**.14** (.04)	**.14** (.03)	**.14** (.04)	**.14** (.04)	**.14** (.04)	**.14** (.04)
	25.1	**.06** (.02)	**.06** (.02)	**.06** (.02)	**.06** (.02)	**.06** (.02)	**.06** (.02)
	40.1	**.04** (.01)	**.03** (.01)	**.04** (.01)	**.04** (.01)	**.04** (.01)	**.04** (.01)
v=8 (4+4)	3.1	**.41** (.12)	**.41** (.12)	**.41** (.12)	**.41** (.12)	**.41** (.12)	**.41** (.13)
	10.1	**.13** (.05)	**.13** (.05)	**.13** (.05)	**.13** (.05)	**.13** (.05)	**.13** (.05)
	25.1	**.05** (.02)	**.05** (.02)	**.05** (.02)	**.05** (.02)	**.05** (.02)	**.05** (.02)
	40.1	**.03** (.01)	**.03** (.01)	**.03** (.01)	**.03** (.01)	**.03** (.01)	**.03** (.01)
v=6 (4+2)	3.1	**.36** (.14)	**.36** (.15)	**.36** (.14)	**.36** (.14)	**.36** (.15)	**.36** (.15)
	10.1	**.11** (.05)	**.11** (.05)	**.11** (.05)	**.11** (.05)	**.11** (.05)	**.11** (.06)
	25.1	**.04** (.02)	**.04** (.02)	**.04** (.02)	**.04** (.02)	**.04** (.02)	**.04** (.02)
	40.1	**.03** (.01)	**.03** (.01)	**.03** (.01)	**.03** (.01)	**.03** (.01)	**.03** (.01)
v=12 (10+2)	3.1	**.38** (.10)	**.38** (.10)	**.38** (.10)	**.38** (.10)	**.38** (.10)	**.38** (.11)
	10.1	**.11** (.04)	**.11** (.04)	**.11** (.04)	**.12** (.04)	**.12** (.04)	**.12** (.04)
	25.1	**.05** (.02)	**.05** (.02)	**.05** (.02)	**.05** (.02)	**.05** (.02)	**.05** (.02)
	40.1	**.03** (.01)	**.03** (.01)	**.03** (.01)	**.03** (.01)	**.03** (.01)	**.03** (.01)

Note. Correlation matrix **A** was created with r_w=0 and r_b=0. *k* denotes univariate kurtosis. *n*=5,000 per cell. Bolded entries represent values that exceeded the ±.30 R_c^2 criterion for extreme bias. [a]The bias of the R_c^2 values is denoted as the mean difference between the known population value of R_c^2 and the average sample value of R_c^2 across 5,000 sample replications. [b]The precision of the R_c^2 values is denoted as the standard deviation of the accuracy values.

Table A2: Bias and Precision of Function I Sample R_c^2 Values for Correlation Matrix **B**

Var. Set	n: v	Expected k										
		-1		**0**		**1**		**3**		**5**		**8**
v=12 (6+6)	3.1	**.41**[a] (.09)[b]		**.41** (.09)		**.41** (.09)		**.41** (.09)		**.41** (.09)		**.41** (.10)
	10.1	**.11** (.05)		**.11** (.04)		**.11** (.05)		**.11** (.05)		**.12** (.05)		**.12** (.05)
	25.1	**.04** (.03)		**.04** (.04)		**.04** (.03)		**.04** (.03)		**.04** (.03)		**.04** (.03)
	40.1	**.02** (.02)		**.02** (.02)		**.02** (.02)		**.02** (.02)		**.02** (.02)		**.02** (.02)
v=8 (4+4)	3.1	**.39** (.12)		**.38** (.12)		**.38** (.12)		**.38** (.12)		**.39** (.13)		**.39** (.13)
	10.1	**.11** (.05)		**.11** (.05)		**.11** (.06)		**.11** (.06)		**.11** (.06)		**.11** (.06)
	25.1	**.04** (.03)		**.04** (.03)		**.04** (.03)		**.04** (.03)		**.04** (.03)		**.04** (.03)
	40.1	**.02** (.02)		**.02** (.02)		**.02** (.02)		**.02** (.02)		**.02** (.02)		**.02** (.03)
v=6 (4+2)	3.1	**.34** (.14)		**.34** (.15)		**.34** (.15)		**.34** (.15)		**.34** (.15)		**.34** (.16)
	10.1	**.10** (.06)		**.10** (.06)		**.10** (.06)		**.10** (.06)		**.10** (.06)		**.10** (.07)
	25.1	**.04** (.03)		**.04** (.03)		**.04** (.03)		**.04** (.03)		**.04** (.03)		**.04** (.04)
	40.1	**.02** (.02)		**.02** (.02)		**.02** (.02)		**.02** (.02)		**.02** (.02)		**.02** (.03)
v=12 (10+2)	3.1	**.36** (.10)		**.30** (.10)		**.36** (.10)		**.36** (.10)		**.36** (.11)		**.36** (.11)
	10.1	**.10** (.05)		**.08** (.06)		**.10** (.05)		**.10** (.05)		**.10** (.05)		**.10** (.05)
	25.1	**.04** (.03)		.03 (.04)		**.04** (.03)		**.04** (.03)		**.04** (.03)		**.04** (.03)
	40.1	**.02** (.02)		.02 (.03)		**.02** (.02)		**.02** (.02)		**.02** (.02)		**.02** (.02)

Note. Correlation matrix **B** was created with r_w=.3 and r_b=.1 k denotes univariate kurtosis. n=5,000 per cell. Bolded entries represent values that exceeded the ±.30 R_c^2 criterion for extreme bias. [a]The bias of the R_c^2 values is denoted as the mean difference between the known population value of R_c^2 and the average sample value of R_c^2 across 5,000 sample replications. [b]The precision of the R_c^2 values is denoted as the standard deviation of the accuracy values.

Table A3: Bias and Precision of Function I Sample R_c^2 Values for Correlation Matrix **C**

Var. Set	n: v	Expected k					
		-1	0	1	3	5	8
v=12 (6+6)	3.1	**.43**[a] (.09)[b]	**.43** (.09)	**.42** (.09)	**.43** (.09)	**.43** (.10)	**.44** (.10)
	10.1	**.13** (.04)	**.13** (.04)	**.13** (.04)	**.13** (.04)	**.13** (.04)	**.13** (.05)
	25.1	**.05** (.02)	**.05** (.02)	**.04** (.02)	**.05** (.02)	**.05** (.02)	**.05** (.02)
	40.1	**.03** (.02)	**.03** (.02)	**.03** (.02)	**.03** (.02)	**.03** (.02)	**.03** (.02)
v=8 (4+4)	3.1	**.40** (.12)	**.40** (.12)	**.40** (.12)	**.40** (.12)	**.40** (.12)	**.40** (.13)
	10.1	**.12** (.05)	**.12** (.05)	**.11** (.05)	**.12** (.05)	**.12** (.06)	**.12** (.05)
	25.1	**.04** (.03)	**.04** (.03)	**.04** (.03)	**.04** (.03)	**.04** (.03)	**.04** (.03)
	40.1	**.03** (.02)	**.03** (.02)	**.03** (.02)	**.03** (.02)	**.02** (.02)	**.03** (.02)
v=6 (4+2)	3.1	**.35** (.14)	**.35** (.14)	**.35** (.15)	**.35** (.15)	**.35** (.15)	**.35** (.15)
	10.1	**.10** (.06)	**.10** (.06)	**.10** (.06)	**.10** (.06)	**.10** (.06)	**.10** (.06)
	25.1	**.04** (.03)	**.04** (.03)	**.04** (.03)	**.04** (.03)	**.04** (.03)	**.04** (.03)
	40.1	**.02** (.02)	**.02** (.02)	**.02** (.02)	**.02** (.02)	**.02** (.02)	**.02** (.02)
v=12 (10+2)	3.1	**.37** (.10)	**.37** (.10)	**.37** (.10)	**.37** (.10)	**.37** (.10)	**.37** (.11)
	10.1	**.11** (.04)	**.11** (.04)	**.11** (.04)	**.11** (.04)	**.11** (.04)	**.11** (.05)
	25.1	**.04** (.02)	**.04** (.02)	**.04** (.02)	**.04** (.02)	**.04** (.02)	**.04** (.02)
	40.1	**.02** (.02)	**.02** (.02)	**.02** (.02)	**.02** (.02)	**.02** (.02)	**.02** (.02)

Note. Correlation matrix **C** was created with r_w=.5 and r_b=.1. *k* denotes univariate kurtosis. *n*=5,000 per cell. Bolded entries represent values that exceeded the ±.30R_c^2 criterion for extreme bias. [a]The bias of the R_c^2 values is denoted as the mean difference between the known population value of R_c^2 and the average sample value of R_c^2 across 5,000 sample replications. [b]The precision of the R_c^2 values is denoted as the standard deviation of the accuracy values.

Table A4: Bias and Precision of Function I Sample R_c^2 Values for Correlation Matrix **D**

Var. Set	n: v	Expected k					
		-1	0	1	3	5	8
v=12 (6+6)	3.1	.15[a] (.09)[b]	**.16** (.09)	.15 (.09)	**.16** (.09)	**.16** (.09)	**.17** (.09)
	10.1	.04 (.06)	.04 (.06)	.04 (.06)	.04 (.06)	.05 (.06)	.05 (.06)
	25.1	.02 (.04)	.02 (.04)	.02 (.04)	.02 (.04)	.02 (.04)	.02 (.04)
	40.1	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)
v=8 (4+4)	3.1	**.19** (.12)	**.19** (.12)	**.19** (.12)	**.20** (.13)	**.20** (.13)	**.21** (.13)
	10.1	.05 (.08)	.05 (.08)	.05 (.08)	.05 (.08)	.05 (.09)	.06 (.09)
	25.1	.02 (.05)	.02 (.05)	.02 (.05)	.02 (.05)	.02 (.06)	.02 (.06)
	40.1	.01 (.04)	.01 (.04)	.01 (.04)	.01 (.04)	.01 (.04)	.01 (.05)
v=6 (4+2)	3.1	**.22** (.15)	**.22** (.15)	**.22** (.15)	**.22** (.16)	**.23** (.16)	**.23** (.16)
	10.1	.06 (.10)	.06 (.09)	.06 (.10)	.06 (.10)	.06 (.10)	.07 (.11)
	25.1	.02 (.06)	.02 (.06)	.02 (.06)	.02 (.06)	.02 (.06)	.03 (.07)
	40.1	.01 (.05)	.01 (.05)	.01 (.05)	.01 (.05)	.02 (.05)	.02 (.05)
v=12 (10+2)	3.1	**.20** (.11)	**.20** (.11)	**.20** (.11)	**.20** (.11)	**.20** (.11)	**.22** (.11)
	10.1	.05 (.07)	.05 (.07)	.05 (.07)	.06 (.07)	.06 (.07)	.07 (.07)
	25.1	.02 (.04)	.02 (.04)	.02 (.04)	.02 (.04)	.02 (.05)	.03 (.05)
	40.1	.01 (.04)	.01 (.03)	.01 (.03)	.01 (.03)	.02 (.04)	.02 (.04)

Note. Correlation matrix **D** was created with r_w=.3 and r_b=.3. *k* denotes univariate kurtosis. *n*=5,000 per cell. Bolded entries represent values that exceeded the ±.30 R_c^2 criterion for extreme bias. [a] The bias of the R_c^2 values is denoted as the mean difference between the known population value of R_c^2 and the average sample value of R_c^2 across 5,000 sample replications. [b] The precision of the R_c^2 values is denoted as the standard deviation of the accuracy values.

Table A5: Bias and Precision of Function I Sample R_c^2 Values for Correlation Matrix **E**

Var. Set	n: v	Expected k					
		-1	0	1	3	5	8
v=12 (6+6)	3.1	**.28**[a] (.10)[b]	**.28** (.10)	**.28** (.09)	**.28** (.10)	**.28** (.10)	**.30** (.11)
	10.1	.07 (.07)	.07 (07)	.07 (.07)	.07 (.07)	.07 (.07)	**.08** (.07)
	25.1	.03 (.04)	.03 (.04)	.03 (.04)	.03 (.04)	.03 (.05)	.03 (.05)
	40.1	.02 (.03)	.02 (.03)	.02 (.03)	.02 (.04)	.02 (.04)	.02 (.04)
v=8 (4+4)	3.1	**.28** (.12)	**.27** (.12)	**.27** (.13)	**.28** (.13)	**.28** (.13)	**.30** (.14)
	10.1	.07 (.08)	**.07** (.08)	**.07** (.08)	**.07** (.08)	**.08** (.08)	**.08** (.09)
	25.1	.03 (.05)	.03 (.05)	.03 (.05)	.03 (.05)	.03 (.05)	.03 (.06)
	40.1	.02 (.04)	.02 (.04)	.02 (.04)	.02 (.04)	.02 (.04)	.02 (.04)
v=6 (4+2)	3.1	**.27** (.16)	**.26** (.15)	**.26** (.15)	**.27** (.16)	**.27** (.16)	**.28** (.17)
	10.1	**.07** (.09)	**.07** (.09)	**.06** (.09)	**.07** (.09)	**.07** (.10)	**.08** (.10)
	25.1	.02 (.06)	.03 (.06)	.02 (.06)	.03 (.06)	.03 (.05)	.03 (.06)
	40.1	.02 (.05)	.01 (.05)	.02 (.05)	.02 (.05)	.02 (.05)	.02 (.05)
v=12 (10+2)	3.1	**.27** (.11)	**.27** (.11)	**.27** (.11)	**.28** (.11)	**.27** (.11)	**.29** (.12)
	10.1	**.07** (.07)	**.07** (.07)	**.07** (.07)	**.08** (.07)	**.08** (.07)	**.08** (.07)
	25.1	.03 (.04)	.03 (.04)	.03 (.04)	.03 (.04)	.03 (.04)	.03 (.05)
	40.1	.02 (.03)	.02 (.03)	.02 (.03)	.02 (.03)	.02 (.03)	.02 (.04)

Note. Correlation matrix **E** was created with r_w=.5 and r_b=.3. *k* denotes univariate kurtosis. *n*=5,000 per cell. Bolded entries represent values that exceeded the ±.30 R_c^2 criterion for extreme bias. [a] The bias of the R_c^2 values is denoted as the mean difference between the known population value of R_c^2 and the average sample value of R_c^2 across 5,000 sample replications. [b] The precision of the R_c^2 values is denoted as the standard deviation of the accuracy values.

Table A6: Bias and Precision of Function I Sample R_c^2 Values for Correlation Matrix **F**

Var. Set	n: v	Expected k					
		-1	0	1	3	5	8
v=12 (6+6)	3.1	.03[a] (.02)[b]	.03 (.02)	.02 (.02)	.03 (.02)	.03 (.02)	.04 (.02)
	10.1	.01 (.01)	.01 (.02)	.01 (.01)	.01 (.02)	.01 (.02)	.01 (.02)
	25.1	<.01 (.01)	<.01 (.01)	<.01 (.01)	<.01 (.01)	<.01 (.01)	.01 (.01)
	40.1	<.01 (.01)	<.01 (.01)	<.01 (.01)	<.01 (.01)	<.01 (.01)	<.01 (.01)
v=8 (4+4)	3.1	.07 (.07)	.06 (.07)	.06 (.07)	.07 (.07)	.07 (.07)	.08 (.07)
	10.1	.02 (.04)	.02 (.04)	.02 (.04)	.02 (.05)	.02 (.05)	.03 (.05)
	25.1	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)
	40.1	<.01 (.02)	<.01 (.02)	<.01 (.02)	.01 (.02)	.01 (.02)	.01 (.03)
v=6 (4+2)	3.1	.10 (.13)	.10 (.12)	.10 (.12)	.11 (.12)	.11 (.12)	.13 (.12)
	10.1	.03 (.08)	.03 (.08)	.02 (.07)	.03 (.08)	.03 (.08)	.04 (.08)
	25.1	.01 (.05)	.01 (.05)	.01 (.05)	.01 (.05)	.01 (.05)	.02 (.05)
	40.1	.01 (.04)	.01 (.04)	.01 (.04)	.01 (.04)	.01 (.04)	.01 (.04)
v=12 (10+2)	3.1	.09 (.07)	.08 (.07)	.08 (.06)	.09 (.07)	.09 (.07)	.10 (.07)
	10.1	.02 (.04)	.02 (.04)	.02 (.04)	.03 (.04)	.03 (.04)	.04 (.05)
	25.1	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)	.02 (.03)
	40.1	.01 (.02)	.01 (.02)	.01 (.02)	.01 (.02)	.01 (.02)	.01 (.02)

Note. Correlation matrix **F** was created to have r_w=.3 and r_b=.5, but limitations with the data generation procedures required us to create a correlation matrix with r_w=.43 and r_b=.5. *k* denotes univariate kurtosis. *n*=5,000 per cell. Bolded entries represent values that exceeded the ±.30 R_c^2 criterion for extreme bias. [a] The bias of the R_c^2 values is denoted as the mean difference between the known population value of R_c^2a\ and the average sample value of R_c^2 across 5,000 sample replications. [b] The precision of the R_c^2 values is denoted as the standard deviation of the accuracy values.

LEACH & HENSON

Table A7: Bias and Precision of Function I Sample R_c^2 Values for Correlation Matrix **G**

Var. Set	n: v	Expected k					
		-1	0	1	3	5	8
v=12 (6+6)	3.1	.08[a] (.06[b])	.08[a] (.06)	.08 (.06)	.08 (.06)	.08 (.06)	.09 (.06)
	10.1	.02 (.04)	.02 (.04)	.02 (.04)	.03 (.04)	.03 (.04)	.03 (.04)
	25.1	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)
	40.1	.01 (.02)	.01 (.02)	<.01 (.02)	.01 (.02)	.01 (.02)	.01 (.02)
v=8 (4+4)	3.1	.11 (.10)	.10 (.10)	.10 (.10)	.11 (.10)	.11 (.10)	.13 (.10)
	10.1	.03 (.06)	.03 (.06)	.03 (.06)	.03 (.06)	.04 (.07)	.04 (.07)
	25.1	.01 (.04)	.01 (.04)	.01 (.04)	.01 (.04)	.02 (.04)	.02 (.05)
	40.1	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.04)
v=6 (4+2)	3.1	.13 (.14)	.12 (.13)	.12 (.13)	.13 (.14)	.14 (.14)	**.16** (.15)
	10.1	.03 (.09)	.03 (.08)	.03 (.09)	.04 (.09)	.04 (.09)	.05 (.09)
	25.1	.01 (.06)	.01 (.06)	.01 (.06)	.02 (.06)	.02 (.06)	.02 (.06)
	40.1	.01 (.04)	.01 (.04)	.01 (.04)	.01 (.04)	.01 (.05)	.01 (.05)
v=12 (10+2)	3.1	.12 (.08)	.11 (.09)	.12 (.08)	.12 (.09)	.13 (.09)	.14 (.09)
	10.1	.04 (.05)	.03 (.05)	.03 (.05)	.04 (.06)	.04 (.06)	.05 (.06)
	25.1	.01 (.04)	.01 (.03)	.01 (.03)	.02 (.04)	.02 (.04)	.02 (.04)
	40.1	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)	.01 (.03)

Note. Correlation matrix **G** was created with r_w=.5 and r_b=.5. k denotes univariate kurtosis. n=5,000 per cell. Bolded entries represent values that exceeded the ±.30 R_c^2 criterion for extreme bias. [a] The bias of the R_c^2 alues is denoted as the mean difference between the known population value of R_c^2 and the average sample value of R_c^2 across 5,000 sample replications. [b] The precision of the R_c^2 values is denoted as the standard deviation of the accuracy values.

139