Correction for Unreliability of Multifactor Measures: Comparison of Alpha and Parallel Forms Approaches

William M. Rogers
Grand Valley State University

Neal Schmitt
Michigan State University

Morell E. Mullins
Bowling Green State University - Main Campus

METHODOLOGICAL RESOURCES

Correction for Unreliability of Multifactor Measures: Comparison of Alpha and Parallel Forms Approaches

WILLIAM M. ROGERS
Grand Valley State University

NEAL SCHMITT
Michigan State University

MORELL E. MULLINS
Bowling Green State University

Coefficient alpha is a widely used and useful index of reliability, but it provides an inappropriate estimate of reliability when used with multidimensional composites. The problematic nature of this inappropriate usage varies primarily as a function of the prominence of distinct factors in a composite and to a lesser extent as a function of the length of the measure. Implications for use of coefficient alpha estimates of reliability as the basis for correcting observed correlations for unreliability are described for a variety of levels of multidimensionality and measure length. Researchers should correct for unreliability to estimate construct interrelationships but must exercise care as to how they conceptualize errors of measurement and how they estimate reliability.

Coefficient alpha (Cronbach, 1947, 1951) is used as an index of the internal consistency reliability of measures in virtually all research domains in psychological and organizational science. Various authors have noted that alpha assumes that the measure being assessed consists of a unidimensional set of items and that it is not an index of item homogeneity (Cortina, 1993; Green, Lissitz, & Mulaik, 1977; Hattie, 1985; Miller, 1995). This point, however, must be weighed against the difficulties in developing pure, unidimensional measures of relevant constructs in any domain of behavioral science. In the majority of cases where a measure possesses some degree of multidimensionality, the use of coefficient alpha is inappropriate and its application to correcting for attenuation in observed correlations between constructs (Schmidt & Hunter, 1996) compounds the problem by overestimating true construct relationships.

The purpose of this article is primarily pedagogical and will focus on illustrating the circumstances in which overestimating true construct relationships is likely, how to assess (at least empirically) when the inappropriate use of alpha is likely to produce an underestimate of reliability, and the magnitude of such underestimation.

Coefficient Alpha and Parallel Forms Reliability

Cronbach (1951) realized the effects of multidimensionality when he asserted that alpha was a lower bound index of internal consistency reliability. He also compared the computation of alpha and parallel forms reliability when a measure comprised items varying in factor structure. Equation (1) shows Cronbach's coefficient alpha:

$$r_{xx} = \frac{N^2 \cdot \overline{\sigma}_{i,j}}{V_t},$$ (1)

where N is the number of items in the measure, $\overline{\sigma}_{i,j}$ is the mean off-diagonal covariance, and V_t is the variance of the item composite. Equation (2) shows the parallel forms reliability formula described by Cronbach (1951):

$$r_{xx(pf)} = \frac{\sum_i \sum_j \sigma_i \sigma_j\, r_{ij}}{V_t},$$ (2)

where i and j are subcomponent indices, V_t is the total composite variance as above, and r_{ij} is the correlation between subcomponents. Note that when $i = j$, r_{ij} is simply replaced by the alpha reliability estimate of the subcomponent.

Similar problems with the use of alpha have been noted more recently (Komaroff, 1997; Murphy & DeShon, 2000; Raykov, 1998, 2001). For a prespecified set of items (that do not necessarily assess the same underlying dimension) with uncorrelated errors, alpha underestimates reliability. Only when these items are tau equivalent (i.e., they measure an underlying dimension in the same units of measurement, but perhaps with different precision) does alpha equal reliability (Lord & Novick, 1968).

Error Sources and Reliability Estimates

In considering these alternate versions of reliability, it is useful to clarify what we assume is true and error variance. When treating the content of items, Cronbach (1947) divided composite variance into general factor variance common to all items, group factor variance that was common to some items, and specific factor variance unique to each item in a multi-item measure, as well as random error. Parallel forms reliability (Equation (2)) is based on the notion that each form of a measure comprises equal portions of general and group factor variance. The use of coefficient alpha, however, assumes a single dimension accounts for true variance in the measure. All variance shared by subsets of items and not part of the general factor is considered error. Alpha, then, will provide a lower estimate of internal consistency reliability than will parallel forms reliability when items are a function of group factor variance as well as a general

factor. Similar ideas about error and reliability are discussed by Lord and Novick (1968) and are the basis for the subsequent development of generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972).

For the situations examined by Cronbach (1947), alpha was a lower-bound estimate of reliability relative to a strictly parallel estimate of reliability when the item pool was composed of items representing several strong group factors and no general factor. When a set of items is multidimensional, an appropriate estimate of reliability can only be obtained when the various factors are represented equally in the various forms of the measure. Correlations between these strictly parallel forms are reliability coefficients. An ideal set of items that allows for unambiguous interpretation of the measurement of a single construct was identified as one with a high coefficient alpha and one that is not divisible into smaller blocks of items whose intercorrelation within the block exceeds their intercorrelation with the remainder of the scale.

All this is well known among psychometricians (e.g., Green et al., 1977); in fact, our references span more than half a century and include reviews of the literature. In practice, however, researchers often work with multidimensional measures because of the need for parsimony, inadequate examination of measurement models (Anderson & Gerbing, 1988), or the allure of alpha's well-known relationship with the length of the measure, which may promote the combination of independent measures into a large composite. Cortina (1993) demonstrated the general tendency of alpha to decrease as a function of multidimensionality and increase as a function of item intercorrelation, but noted that the effect of measure length can create high values of alpha even in the case of three orthogonal subscales. Similar observations were made by Green et al. (1977). In describing these effects, Cortina cautioned that alpha is a useful statistic when its assumptions are met; that is, one is interested in assessing reliability when error consists of item-specific variance. The present article will extend the work of Cortina by examining the effects of violations of these assumptions in the context of one prominent application—the correction for attenuation.

An Empirical Examination of Alpha and Parallel Forms of Reliability and the Correction for Attenuation

Reliability has many practical implications for research in the organizational sciences, foremost among them being the attenuating effects of low reliability on observed correlations between theoretical constructs. The degree of this attenuation was first explored by Spearman (1904), who derived a formula relating the observed correlation to the correlation between true scores on the constructs measured by two sets of items or groups of items. This correction for attenuation is given as follows:

$$r_{t1t2} = \frac{r_{12}}{\sqrt{r_{11}r_{22}}} ,$$

(3)

where r_{t1t2} refers to the correlation between true scores represented by measures 1 and 2, r_{12} is the observed correlation between the two measures, and r_{11} and r_{22} are the reliabilities of the two measures. Recently, Schmidt and Hunter (1996) illustrated the importance of reliability in estimating the relationships among constructs of interest and described 26 different research scenarios in which appropriate corrections for un-

reliability in one or both measures provided a more accurate reflection of evidence for hypothesized relationships.

In situations where measures 1 and/or 2 in Equation (3) are multidimensional, the use of coefficient alpha to represent reliabilities r_{11} or r_{22} underestimates reliability, thereby overestimating the true relationship between the measures. Applied psychologists assessing validity coefficients typically correct only for unreliability in the criterion, thus making the potential multidimensionality of job performance (e.g., Campbell, McCloy, Oppler, & Sager, 1993) a concern. In research designed to estimate the relationship between constructs, the correction for attenuation is applied to both measures 1 and 2, potentially causing a correspondingly greater inflation of construct relationship. Schmitt (1996) recently provided an example in which this overestimation of the true correlation is illustrated. The two six-item measures in his illustration consisted of two factors whose cross-factor item intercorrelations were .30 and within-factor item intercorrelations were .80. The observed correlation of these two measures was .94; correcting for unreliability using coefficient alpha of the two measures (i.e., .86) resulted in 1.09 as the estimate of the intercorrelation. Computing the corrected coefficient using Equation (2) presented above to estimate the reliability of "lumpy" measures (Cronbach, 1951) yielded a value of 1.00, which reflects the fact that these two measures were exact replicates in terms of factor structure. In this case, using the lower-bound estimate of parallel forms reliability (i.e., coefficient alpha) produced an obvious overestimate of the true correlation between these measures.

An example of the impact of the use of Equation (1) versus (2) to correct for attenuation as well as a demonstration of the use of the two equations is provided in Table 1. In this table, the relationship between an organizational commitment measure investigated by Bar-Hayim and Berman (1992) and an organizational citizenship behavior (OCB) index presented by Smith, Organ, and Near (1983) is estimated. Although neither work provides all the information required to calculate reliability using both Equations (1) and (2), we believe the calculations in Table 1 are realistic. Standard deviation information was unavailable for items in the Bar-Hayim and Berman study and for subscales in the Smith et al. study, so we used 1.00 as the standard deviation of items and component measures.

In the case of both OCB and organizational commitment, researchers have used these measures as composites as well as multidimensional representations of the two constructs. The alpha coefficient for the organizational commitment composite (.73 by Equation (1)) would be considered acceptable by many researchers (Nunnally & Bernstein, 1994), although the alpha for the composite OCB measure (.62) would often be considered marginal. Corresponding estimates of the reliability of these composites using Equation (2) are .78 and .90, respectively. The difference between the two estimates for organizational commitment (.73 vs. .78) is relatively modest, but the difference for OCB (.62 vs. .90) is large. The use of these alternative estimates of reliability to estimate the true correlation between OCB and commitment is also illustrated in Table 1. The use of Equation (1) estimates as opposed to Equation (2) estimates would result in an inflation of this estimate of 25% (see the bottom of Table 1 for estimates of the corrected relationship for observed correlations of .3, .4, and .5). The percentage inflation is a constant across levels of observed correlation. Most of the inflation in this case came from the difference in estimates of the reliability of the OCB composite as opposed to the difference in estimates of the reliability of organizational commitment (21.5% vs. 3.5%).

Table 1
Estimate of Reliability Using Equations (1) and (2) and
Corresponding Corrections to the Correlation Between Organizational
Commitment and Organizational Citizenship Behavior

Organizational Commitment			*Organizational Citizenship Behavior* [b]		
Component	*SD* [a]	*Alpha*	*Component*	*SD*	*Alpha*
Dimension 1	2.98	.73	Altruism	1.00	.91
Dimension 2	2.22	.59	Compliance	1.00	.81
Dimension 3	1.64	.52	Composite	1.70	.62
Composite	5.03	.73			

Equation (2) Estimates

Commitment = $((2.98 \times 2.98 \times .73) + (2.22 \times 2.22 \times .59) + (1.64 \times 1.64 \times .52)$
$+ (2(2.98 \times 2.22 \times .30)) + (2(2.98 \times 1.64 \times .30)) + (2(2.22 \times 1.64 \times .27)))/25.3$
$= .78$

Organizational
citizenship
behavior = $((1 \times 1 \times .91) + (1 \times 1 \times .81) + (2(.45 \times 1 \times 1)))/2.9$
$= .90$

Correlations Corrected for Attenuation Using Alpha
(Equation (1)) and Parallel Forms (Equation (2)) Reliability Estimates
at Three Levels (.3, .4, and .5) of Observed Correlation [c]

Observed Correlation	*Corrected by Alpha*	*Corrected by Parallel Forms Reliability*
.30	$.30/(\sqrt{.73} \times \sqrt{.62}) = .45$	$.30/(\sqrt{.78} \times \sqrt{.90}) = .36$
.40	$.40/(\sqrt{.73} \times \sqrt{.62}) = .60$	$.40/(\sqrt{.78} \times \sqrt{.90}) = .48$
.50	$.50/(\sqrt{.73} \times \sqrt{.62}) = .74$	$.50/(\sqrt{.78} \times \sqrt{.90}) = .60$

a. Bar-Hayim and Berman (1992) did not present the standard deviations of items, so all computations involved correlations between items that were presented in the appendix of their article.
b. Smith, Organ, and Near (1983) did not provide the standard deviations of the altruism or compliance dimensions or the composite organizational citizenship behavior measure, so the standard deviations of altruism and compliance were set to 1.00 to provide computational values. If the standard deviations of the two dimensions were unequal, our calculations would be different.
c. Estimates of corrected correlations indicate that the degree of inflation resulting from use of alpha estimates of reliability as opposed to parallel forms estimates is a constant across values of the observed correlation between the two composite measures, in this case, approximately 25% (e.g., (.45 − .36)/.36).

   Reliability has many practical implications for research in the organizational sciences. Foremost among these implications is the attenuating effect of low reliability on observed correlations between theoretical constructs. In the next sections of this article, we systematically address the importance of the issues raised in Cortina (1993) and Schmitt (1996) by presenting three computational analyses. First, we present an analysis of the effect of using alpha as an estimate of internal consistency reliability when the measure is multidimensional. Second, we assess the effects of the use of alpha in these instances to estimate true construct relationships. Third, we link this overestimate to estimates of the degree of fit offered by structural equation models

(Miller, 1995). Finally, we provide the concerned researcher with a flowchart representing the manner in which we believe the information presented in this article can be used to assess when an alpha estimate of reliability should be replaced with the estimate provided by Equation (2).

Computational Analysis of Alpha and Parallel Forms of Reliability and the Correction for Attenuation

To estimate the degree to which the use of alpha might result in overcorrections of true correlations, we constructed measures that represented a range of realistic values of four factors we considered important determinants of the size of alpha and the relative unidimensionality of a measure. These four factors are as follows: factor structure of the measure (i.e., number of group factors), items per factor, within-factor interitem correlation, and between-factor interitem correlation. The factor structure of the measures was set at two, four, or six factors. The number of items per factor was set to 3, 5, or 7. This resulted in a total measure length ranging from 6 to 42 items (the product of the number of items per factor and number of factors). Values for within-factor interitem correlations and between-factor interitem correlations were .6, .5, .4, and .3 and .3, .2, .1, and .0, respectively. These levels were chosen to be representative of item intercorrelation structures present in typical attitude scales and job performance rating measures. Between- and within-factor item correlations were uniform at the levels defined above. The various combinations of within- and between-factor item correlations also provide a range of "distinctiveness" of group factors, which varies in proportion to the difference between within- and between- factor item correlations. The complete crossing of the above design parameters resulted in a $3 \times 3 \times 4 \times 4$ design, or 144-cell matrix. For the purposes of assessing the effects of these factors on the two estimates of reliability and true construct correlations, both Cronbach's alpha (Equation (1)) and parallel forms reliability (Equation (2)) indices were computed for the 144 cases produced by the design described above. These values were then used to calculate the raw differences between the reliability estimates, as well as the percentage inflation of true construct correlation when alpha is used instead of the parallel forms formulation.

In the second phase of this study, correlation matrices that represented the 144 cases were analyzed using LISREL 8.0 (Jöreskog & Sörbom, 1994). The purpose of the structural equation analysis was to determine the departure from fit of the generated data from a unidimensional latent structure and to correlate this departure with the difference between the two versions of reliability referred to above. Sample size was fixed at 500 for all cases. Four popular indices available in LISREL were chosen to assess departure from fit: the root mean square error of approximation (RMSEA), the comparative fit index (CFI), the nonnormed fit index (NNFI), and the relative fit index (RFI). RMSEA is based on the population error of approximation and is a measure of discrepancy per degree of freedom. Values of RMSEA less than .05 indicate good model fit (Browne & Cudeck, 1993). The CFI is an indicator of relative model fit compared to the independence model and is adjusted for total degrees of freedom in the model. Values of .90 or above on the CFI, RFI, and NNFI are generally indicative of good model fit. A more detailed discussion of these statistics can be found in Jöreskog and Sörbom (1994) and Bollen and Long (1993).

Finally, correlation matrices generated by the 144 different item composites were submitted to a principal components analysis using SAS PROC FACTOR. The percentage of variance accounted for by the first principal component (FPC) was compared to both the raw differences between alpha and parallel forms reliability indices and the percentage true correlation inflation values for each cell.

Table 2 contains the values of coefficient alpha and parallel forms reliability across levels of the four design factors. It is clear from Table 2A through 2C that the primary factor influencing differences between the reliability indices is the correlation between items in different group factors. As this correlation increases, alpha and parallel forms reliability formulas converge. To further assess the impact of the different factors on reliability estimates, we computed the correlations between the different factors in the design and the two estimates of reliability, the difference between these estimates, and the inflation in estimates of true correlation that results when alpha is used to correct for attenuation. These correlations are presented in Table 3. In Table 3, the impact of the between-factor correlations is evidenced by the correlation (−.78) between the between-factor correlation and the difference between alpha and parallel forms reliability. This is due to the group factors becoming less distinct as the correlations between items from different factors approach the correlation between items within the same factor. There was also a decrease in convergence of the two formulas as within-factor correlations increase. This tendency has a similar explanation as the previously noted effect for between-factor correlations; as within-factor correlations increase, the group factors become more distinct from one another. The effects of both within- and between-factor item correlations define the multidimensionality of the matrix. The more similar the two types of correlations are, the closer the matrix will approximate one obtained from a unidimensional measure.

It is notable that the number of group factors had little main effect on differences between the two indices (see Tables 2A through 2C). The correlation between the number of factors and differences between the two forms of reliability was only .06 (see Table 3). It is the aforementioned magnitude of differences in intra- and interfactor item correlations that drives differences, not the number of group factors. The effect of the number of group factors is predicated on the intra- and intercorrelation differences. Table 4 contains a summary of the results for those parts of our design relevant to these two factors. As can be seen in Table 4, additional group factors only increase divergence in alpha and parallel forms indices when differences between within- and between-factor item correlations are large. Table 4 also shows an opposite pattern when these differences are small (i.e., more group factors reducing discrepancy between alpha and parallel forms indices), but this is likely due to the overriding effect of the number of total items, which in this design was directly tied to the number of group factors.

The number of items per factor also affected the differences between the reliability indices ($r = -.37$). The analysis in Table 3 reveals the known effects of measure length on both alpha ($r_{alpha,\ length} = .42$) and parallel forms ($r_{pf,\ length} = .48$) indices. The analysis suggests that length of the total measure had the effect of moderately decreasing the difference between alpha and parallel forms indices ($r_{diff,\ length} = -.18$). Increasing the number of items per factor has a direct effect on coefficient alpha, and a somewhat indirect effect on the parallel forms index, by increasing the reliability of the component subscales. Recall from Equation (2) that variance components of the subscales are

Table 2
Alpha and Parallel Forms Reliability as a Function of Within- and
Between-Factor Item Correlations and Number of Items per Factor

	Within-Factor r	Between-Factor r .00 r_{xx}	.00 $r_{xx(pf)}$	.10 r_{xx}	.10 $r_{xx(pf)}$	.20 r_{xx}	.20 $r_{xx(pf)}$	.30 r_{xx}	.30 $r_{xx(pf)}$
A. Two-group factors									
Three items per factor	.30	.45	.56	.57	.63	.65	.68	.72	.72
	.40	.53	.67	.63	.71	.70	.75	.76	.78
	.50	.60	.75	.68	.78	.74	.81	.79	.83
	.60	.65	.82	.72	.84	.77	.86	.81	.87
Five items per factor	.30	.61	.68	.70	.74	.76	.78	.81	.81
	.40	.68	.77	.75	.81	.80	.83	.84	.85
	.50	.74	.83	.79	.86	.83	.88	.86	.89
	.60	.78	.88	.83	.90	.86	.91	.88	.92
Seven items per factor	.30	.69	.75	.77	.80	.82	.85	.86	.86
	.40	.76	.82	.81	.85	.85	.88	.88	.89
	.50	.81	.88	.85	.89	.88	.91	.90	.92
	.60	.84	.91	.87	.92	.90	.93	.92	.94
B. Four-group factors									
Three items per factor	.30	.41	.56	.65	.72	.77	.79	.84	.84
	.40	.48	.67	.69	.78	.79	.83	.85	.87
	.50	.55	.75	.71	.83	.80	.87	.86	.89
	.60	.60	.82	.74	.87	.82	.90	.87	.91
Five items per factor	.30	.57	.68	.77	.81	.85	.87	.90	.90
	.40	.65	.77	.80	.85	.86	.89	.90	.92
	.50	.70	.83	.82	.89	.88	.92	.91	.93
	.60	.74	.88	.84	.92	.89	.94	.92	.95
Seven items per factor	.30	.67	.75	.83	.86	.89	.90	.92	.92
	.40	.73	.82	.85	.89	.90	.92	.93	.94
	.50	.78	.88	.87	.92	.91	.94	.94	.95
	.60	.81	.91	.88	.94	.92	.95	.94	.96
C. Six-group factors									
Three items per factor	.30	.40	.56	.72	.77	.83	.85	.89	.89
	.40	.47	.67	.74	.82	.84	.88	.89	.90
	.50	.53	.75	.76	.86	.85	.90	.90	.92
	.60	.58	.82	.77	.89	.86	.92	.90	.94
Five items per factor	.30	.56	.68	.81	.85	.89	.90	.93	.93
	.40	.64	.77	.83	.88	.90	.92	.93	.94
	.50	.69	.83	.85	.91	.91	.94	.94	.95
	.60	.73	.88	.86	.93	.91	.95	.90	.96
Seven items per factor	.30	.66	.75	.86	.89	.92	.93	.95	.95
	.40	.72	.82	.88	.91	.96	.94	.95	.96
	.50	.77	.88	.89	.93	.96	.95	.95	.97
	.60	.80	.91	.90	.95	.94	.97	.96	.97

weighed by their associated coefficient alpha reliability. These results are not novel, and are consistent with the well-documented relationship between measure length and coefficient alpha. Less obvious, however, is the fact that coefficient alpha will be at conventionally acceptable levels, regardless of factor structure, as long as factors and items are positively correlated.

Table 3
Intercorrelations of Four Study Factors, Alpha, Parallel Forms,
and Differences Between the Two Versions of Reliability

	W	B	F	I	L	A	P	D	%
Within factor (W)	1.00								
Between factor (B)	0.00	1.00							
Number of factors (F)	0.00	0.00	1.00						
Number of items (I)	0.00	0.00	0.00	1.00					
Length of measure (L)	0.00	0.00	0.76	0.61	1.00				
Alpha (A)	0.26	0.71	0.17	0.49	0.42	1.00			
Parallel forms (P)	0.51	0.53	0.28	0.47	0.48	0.93	1.00		
Difference (D)	0.27	−0.78	0.06	−0.37	−0.18	−0.79	−0.52	1.00	
Inflation (%)	0.12	−0.74	0.06	−0.40	−0.20	−0.86	−0.63	0.97	1.00

Table 4
Alpha/Parallel Forms Differences by Number of Factors
and Within/Between Correlation Differences

Differences Between Within-Factor and Between-Factor Interitem Correlations	Number of Factors		
	2	4	6
0.00	.000 (3)[a]	.000 (3)	.000 (3)
0.10	.017 (6)	.015 (6)	.012 (6)
0.20	.034 (9)	.034 (9)	.030 (9)
0.30	.057 (12)	.063 (12)	.061 (12)
0.40	.074 (9)	.088 (9)	.087 (9)
0.50	.092 (6)	.117 (6)	.118 (6)
0.60	.110 (3)	.153 (3)	.167 (3)

a. Numbers in parentheses denote the number of design cells contributing to each mean.

To identify situations in which the inappropriate application of alpha to correct observed correlations for unreliability results in the greatest inflation of estimates of true correlations, conditions that resulted in an inflation percentage of 10% or greater based on corrections for only one variable were selected from the 144 conditions analyzed (see Table 5A). Obviously, more situations would be included if we applied this criterion to instances when corrections are made to both variables in an observed relationship. Although these may represent "extreme" situations, they are useful in determining what sort of impact the inappropriate use of Cronbach's alpha may have on estimates of true correlation coefficients. Included in Table 5 is the absolute difference between alpha and parallel forms reliability estimates. The worst situations are clearly scenarios in which several unrelated components, each relatively high in internal consistency, are combined to form a single, supposedly unidimensional measure. In these cases, the inappropriate use of alpha to correct observed correlations results in an inflation of 10% to 19% over the true correlation when only one variable in the pair is multidimensional. If both variables are multidimensional in these relatively extreme ways, the corrected intercorrelations between constructs can be as high as 40% (see the last column of Table 5). The situations illustrated in Table 5 can be explained by recalling

Table 5
Cases Resulting in True Correlation Inflation of 10% or Greater and Cases That
Involve Minimal Differences in Within- and Between-Scale Item Intercorrelations

	I	W	B	A	P	D	%-1 var.	%-2 var.
A. Cases in which inflation of the correction for attenuation involving a single multidimensional composite exceeded 10%								
Two-group factors	3	0.30	0.00	0.45	0.56	0.11	11.80	25.9
	3	0.40	0.00	0.53	0.67	0.13	11.80	25.9
	3	0.50	0.00	0.60	0.75	0.15	11.80	25.9
	3	0.60	0.00	0.65	0.82	0.16	11.80	25.9
Four-group factors	3	0.30	0.00	0.41	0.56	0.15	17.26	36.5
	3	0.40	0.00	0.48	0.67	0.18	17.26	36.5
	3	0.50	0.00	0.55	0.75	0.20	17.26	36.5
	3	0.60	0.00	0.60	0.82	0.22	17.26	36.5
Six-group factors	3	0.30	0.00	0.40	0.56	0.17	19.02	41.4
	3	0.40	0.00	0.47	0.67	0.20	19.02	41.4
	3	0.50	0.00	0.53	0.75	0.22	19.02	41.4
	3	0.60	0.00	0.58	0.82	0.24	19.02	41.4
	5	0.30	0.00	0.56	0.68	0.12	9.92	20.2
	5	0.40	0.00	0.64	0.77	0.13	9.92	20.2
	5	0.50	0.00	0.70	0.83	0.14	9.92	20.2
	5	0.60	0.00	0.73	0.88	0.15	9.92	20.2
B. Cases in which differences of between- and within-scale item intercorrelations are relatively small (.10 or .20)								
Two-group factors	3	0.40	0.20	0.70	0.75	0.05	3.5	7.1
	3	0.30	0.40	0.76	0.78	0.02	1.3	2.5
Four-group factors	3	0.40	0.20	0.79	0.83	0.04	2.5	5.0
	3	0.40	0.30	0.85	0.87	0.02	1.2	2.4
Six-group factors	3	0.40	0.20	0.84	0.88	0.04	2.3	4.6
	3	0.40	0.30	0.89	0.90	0.01	.7	1.2

NOTE: I = items per factor, W = within-interitem *r*, B = between-interitem *r*, A = alpha, P = parallel forms, D = P-A, % = percentage inflation. 1 var. and 2 var. refer to cases in which one or two measures of the same structure are involved in the estimated true correlation and correction for attenuation. Actual calculations of percentage inflation may differ slightly from values above due to rounding errors.

the interactive effects of the number of group factors and correlational structure. As the uniformity of the item-intercorrelation matrix decreases, the number of factors will have a greater impact on differences between the two reliability indices. In situations where the correlation has greater uniformity, a larger number of factors has little effect on index differences.

It might be argued that no knowledgeable researcher would combine items with patterns shown in Table 5A into a single measure. We are more likely to combine measures when the evidence for multidimensionality is relatively weak. Table 5B includes conditions in which the within- and between-scale item intercorrelations are less

extreme (i.e., .10 or .20). As can be seen, in these instances, the inflation of the estimates of true intercorrelation is minimal. All of these cases involve three-item group factors; results for five- and seven-item group factors show slightly less inflation. So, in most realistic cases and with reasonable attention to the dimensionality of one's measures, the use of alpha versus the parallel forms estimate of reliability will produce no significant overestimate of construct intercorrelations.

Especially with long measures, it is often difficult to ascertain the multiple dimensions in a new measure so as to compute a more realistic index of reliability such as that provided by Equation (2). In these instances, it would be useful to know when multidimensionality is likely to make alpha a potentially misleading index of reliability. To provide some guidance whereby researchers can determine when the use of alpha is likely to result in practically meaningful overestimates of corrected coefficients, we conducted the analyses described in the next section.

Factor Analysis and Alpha Versus Parallel Forms of Reliability and the Correction for Attenuation

In this phase of our study, we sought to use structural equation modeling methods to assess the dimensional structure of the correlation matrices generated in the analyses described above and to assess the relationship of design factors with the percentage of variance accounted for by the FPC. Tables 6A through 6C show values of RMSEA, CFI, and FPC for different combinations of the design factors in our simulation. CFI and RMSEA values result from a model that specifies that a single factor account for the item variance-covariance matrix. To demonstrate these relationships at reasonable levels of the SEM indicators, only within-factor correlation levels of .30 and .40 and between-factor correlation values of .00, .10, and .20 were included in the tabled results. Many of the nontabled scenarios result in LISREL output indicating perfect fit (which would be the case for those situations in which there is no difference between the within- and between-factor correlations) or indices of fit well outside the range that would indicate fit to a unidimensional model.

Examination of Tables 6A through 6C reveals that although RMSEA and FPC indices are sensitive to both within- and between-factor item correlations, they are not as reflective of the interactive relationship between these design factors in determining percentage inflation. The CFI, on the other hand, displays such a pattern. Further correlation analysis involving these indices confirms this observation, with the correlation between CFI and percentage inflation ($r_{CFI, \%} = -.85$) being substantially greater than the respective correlations for RMSEA ($r_{RMSEA, \%} = .40$) or FPC ($r_{FPC, \%} = -.50$). Although not presented here, the CFI also was more highly correlated with percentage inflation than any of the other fit indices normally part of the LISREL 8.0 output and the output of other structural equation modeling software packages. Values of the NNFI and the RFI were comparable to those of the CFI.

Practical Guidelines for the Researcher

Treatment of a multidimensional index as representative of a single construct is always theoretically and conceptually incorrect. However, when it makes little difference empirically how the measures are treated, arguments of parsimony might dictate the combination of item sets in which there is some evidence of multidimensionality.

Table 6
RMSEA, CFI, and FPC Variance as a Function of Within- and Between-Factor Item
Correlations and Number of Items Per Factor

| | | Between-Factor r | | | | | | | | |
| | | .00 | | | .10 | | | .20 | | |
Within-Factor r	RMSEA	CFI	FPC	RMSEA	CFI	FPC	RMSEA	CFI	FPC
A. Two-group factors										
Three items per factor										
	.30	.16	.17	.27	.13	.45	.32	.06	.86	.37
	.40	.21	.17	.30	.20	.30	.35	.15	.63	.40
Five items per factor										
	.30	.13	.36	.22	.10	.58	.27	.05	.87	.32
	.40	.17	.36	.26	.16	.46	.31	.12	.70	.36
Seven items per factor										
	.30	.11	.41	.20	.09	.61	.25	.04	.87	.30
	.40	.15	.41	.24	.14	.50	.29	.10	.71	.34
B. Four-group factors										
Three items per factor										
	.30	.11	.08	.13	.08	.52	.21	.03	.89	.28
	.40	.15	.08	.15	.13	.33	.23	.09	.69	.30
Five items per factor										
	.30	.10	.16	.11	.07	.52	.19	.02	.87	.26
	.40	.13	.16	.13	.11	.37	.21	.08	.67	.28
Seven items per factor										
	.30	.09	.19	.10	.06	.52	.18	.01	.85	.25
	.40	.12	.19	.12	.10	.38	.20	.07	.65	.27
C. Six-group factors										
Three items per factor										
	.30	.08	.06	.09	.06	.55	.17	.00	.90	.26
	.40	.12	.06	.10	.10	.36	.18	.07	.72	.27
Five items per factor										
	.30	.08	.10	.07	.05	.52	.16	.00	.87	.24
	.40	.11	.10	.09	.09	.36	.17	.06	.67	.25
Seven items per factor										
	.30	.07	.12	.07	.05	.50	.15	P	P	.23
	.40	.10	.12	.12	.08	.35	.25	.06	.65	.37

NOTE: RMSEA = root mean square error of approximation, CFI = comparative fit index, FPC = percentage of variance accounted for by the first principal component. CFI and RMSEA are based on tests of fit of a unidimensional model to the data.

More to the practical issue of when multidimensionality actually makes a difference empirically, we determined that a CFI value above .80 is always associated with differences in reliability indices less than .03. Hence, this CFI value may be used as an indication that alpha would adequately represent the reliability of a measure. In terms of the correction for attenuation (i.e., Equation (3)), such differences in reliability would result in overcorrections of correlation approximately equal to .01 over a wide range of probable reliability and correlation coefficients. Whether this value is too large to tolerate depends on the research situation and researcher or user values. CFI values of .80 are substantially lower than those recommended by structural equation modeling researchers (Hu & Bentler, 1999) as indicative of good fit. In all the instances exam-

ined in Table 6, a multidimensional model actually was appropriate in the sense that within-group item intercorrelations exceeded between-group item intercorrelations. CFIs always dropped below .80 when that difference exceeded .10. This degree of multidimensionality does not make a great deal of practical difference in what reliability estimate is used (Equations (1) or (2)) and in terms of estimating true correlations between different constructs.

It is interesting to note that a similar issue is being investigated by structural equation modeling researchers (e.g., Bagozzi & Edwards, 1998; Hall, Snell, & Foust, 1999). These investigators are interested in how to parcel items to form indicators of an underlying construct in the presence of a second construct. In the context of structural equation modeling, this decision also has an impact on estimates of structural (i.e., between construct) relationships, but general consensus and guidelines on this issue are still being developed.

Our advice to researchers who wish to use a composite that is influenced to some degree by secondary constructs is summarized in Figure 1. The good news is that the degree of multidimensionality that will actually make a practical impact on research results is clearly recognizable if a researcher evaluates the dimensionality of measures. Alpha is not a serious underestimate unless a measure comprises clearly identifiable multiple dimensions.

In any event, our results underscore Anderson and Gerbing's (1988) caveat that a satisfactory measurement model (i.e., clear notions of what is true and error variance with empirical support) always be developed prior to the estimation of structural coefficients in structural equation models—in the context of the discussion in this article, before the computation of corrected correlation coefficients.

Discussion

The analyses described in this article illustrate the degree to which the inappropriate use of alpha as an index of reliability will overcorrect an observed correlation when a measure is actually multidimensional. We also demonstrated that this effect is primarily due to the distinctiveness of the factors present in the measure and only to a small degree to the number of factors and number of items in the measure. The degree to which the interrelationships between constructs is inflated by the use of alpha in Equation (3) depends on the degree to which measures of one or both constructs are multidimensional, or lumpy, to use Cronbach's (1951) term. We also provide evidence that a popular measure of goodness of fit in structural equation modeling (CFI) can be used as a rough estimate of the appropriateness of using alpha as a reliability estimate. In the presence of multidimensionality, it is clearly inappropriate to use alpha. It is a lower-bound estimate of parallel forms reliability, as Cronbach emphasized a half century ago, or tau equivalence as more recently noted (Cortina, 1993; Green et al., 1977; Lord & Novick, 1968). However, the article also illustrates that in most realistic situations, the use of alpha in correcting for attenuation will not provide overestimates of construct-construct relationships. A researcher who conducts relatively routine examinations (e.g., observes item intercorrelations for evidence of group factors or conducts exploratory factor analyses of the items) is unlikely to ignore the level of multidimensionality (within- versus between-group item intercorrelations that exceed .10) that results in significant misuse of alpha.
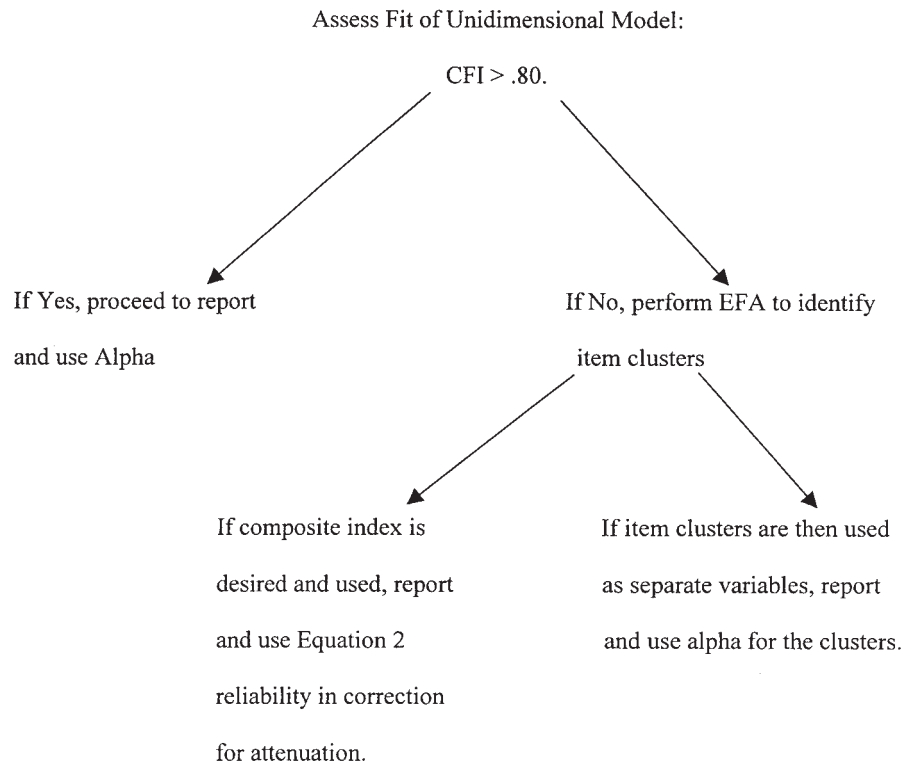
Assess Fit of Unidimensional Model:

CFI > .80.

If Yes, proceed to report

and use Alpha

If No, perform EFA to identify

item clusters

If composite index is

desired and used, report

and use Equation 2

reliability in correction

for attenuation.

If item clusters are then used

as separate variables, report

and use alpha for the clusters.

Figure 1:   Flowchart Depicting the Recommended Use of Different Estimates of Reliability (CFI = Comparative Fit Index, EFA = Exploratory Factor Analysis)

Exact estimates of the degree to which the use of alpha has served to provide overestimates of construct interrelationships are difficult to document because sufficient data to make appropriate estimates are rarely available. The results of the analyses presented here suggest that the magnitude of overestimation due to erroneous use of alpha is minimal in most scenarios assessed, yet it has the possibility of being substantial in specific cases. The correction for attenuation due to unreliability in the criterion is routinely used in the testing of theoretical models, criterion-related validation research, and estimation of population validity coefficients derived from meta-analyses (Hunter & Schmidt, 1990). Like Anderson and Gerbing (1988), we believe that conceptual and empirical issues related to the researcher's measurement model should be resolved prior to consideration of the relationship between constructs. Furthermore, determinations about the measurement model should always include concern for theory (what is considered true and error variance) and the content of the items in a measure (Murphy & DeShon, 2000). Like many psychometricians (e.g., Nunnally, 1967, 1978; Nunnally & Bernstein, 1994), we believe that researchers ought to consider each set of unidimensional, substantively interpretable items separately (rather than in any composite fashion) when relating those measures to important outcomes or other constructs. Such combination serves to obscure important factor or scale information and theoretically interesting relationships.

It should be noted that our treatment of multidimensionality in this article is relatively narrow in that we consider only the relative size of the correlations between and within groups of items. There are instances in which ability items produce difficulty factors that are a function of the behavior of correlations based on dichotomies (right-wrong scoring) rather than the underlying construct or ability being measured. Bollen and Lennox (1991) pointed to those instances in which items may be seen as causing a construct rather than as indicators of a construct. When a causal indicator model is substantively meaningful, item intercorrelations and indices such as alpha are meaningless.

Finally, we should emphasize that we support the use of the correction for attenuation due to unreliability when reliability is appropriately estimated from a measurement and conceptual point of view (Murphy & DeShon, 2000). Schmidt and Hunter (1996) amply demonstrated the degree to which lack of reliability has clouded the results of empirical research in many domains of social and organizational science research. However, this article demonstrates the crucial importance of using one popular index of reliability appropriately. If one does not, inappropriate estimates of reliability and overestimates of construct interrelationships are the likely result.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*, 411-423.

Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods*, *1*, 45-87.

Bar-Hayim, A., & Berman, G. S. (1992). The dimensions of organizational commitment. *Journal of Organizational Behavior*, *13*, 379-387.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*, 305-314.

Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*, 98-104.

Cronbach, L. J. (1947). Test "reliability": Its meaning and determination. *Psychometrika*, *12*, 1-16.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurement*. New York: John Wiley.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, *37*, 827-838.

Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, *2*, 233-256.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, *9*, 139-164.

Hu, L., & Bentler, P. M. (1999). Cut off criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage.

Jöreskog, K. G., & Sörbom, D. (1994). *LISREL 8 user's reference guide*. Chicago: Scientific Software International.

Komaroff, E. (1997). Effect of simultaneous violations of essential of tau-equivalence and uncorrelated errors on coefficient alpha. *Applied Psychological Measurement*, *21*, 337-348.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspective of classical test theory and structural equation modeling. *Structural Equation Modeling*, *2*, 255-273.

Murphy, K. R., & DeShon, R. P. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology*, *53*, 873-900.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

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

Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement*, *22*, 369-377.

Raykov, T. (2001). Bias of coefficient alpha for fixed congeneric measure with correlated errors. *Applied Psychological Measurement*, *25*, 69-76.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, *1*, 199-223.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*, 350-353.

Smith, C. A., Organ, D. W., & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology*, *68*, 653-663.

Spearman, C. (1904). General intelligence: Objectively determined and measured. *American Journal of Psychology*, *15*, 201-293.

William M. Rogers is an assistant professor in the Department of Psychology at Grand Valley State University.

Neal Schmitt is a professor in the Department of Psychology at Michigan State University.

Morell E. Mullins is an assistant professor in the Department of Psychology at Bowling Green State University.