In Search of Underlying Dimensions: The Use (and Abuse) of Factor Analysis in *Personality and Social Psychology Bulletin*

Daniel W. Russell
Iowa State University

An examination of the use of exploratory and confirmatory factor analysis by researchers publishing in Personality and Social Psychology Bulletin *over the previous 5 years is presented, along with a review of recommended methods based on the recent statistical literature. In the case of exploratory factor analysis, an examination and recommendations concerning factor extraction procedures, sample size, number of measured variables, determining the number of factors to extract, factor rotation, and the creation of factor scores are presented. These issues are illustrated via an exploratory factor analysis of data from the University of California, Los Angeles, Loneliness Scale. In the case of confirmatory factor analysis, an examination and recommendations concerning model estimation, evaluating model fit, sample size, the effects of non-normality of the data, and missing data are presented. These issues are illustrated via a confirmatory factor analysis of data from the Revised Causal Dimension Scale.*

Factor analysis is a commonly used statistical procedure in the areas of personality and social psychology. A recent article by Fabrigar, Wegener, MacCallum, and Strahan (1999) reported that 159 of the 883 articles (18%) appearing in the *Journal of Personality and Social Psychology* from 1991 to 1995 reported an exploratory factor analysis. I conducted a similar review of articles appearing in *Personality and Social Psychology Bulletin* (*PSPB*) during the years 1996, 1998, and 2000.[1] Eighty-five of the 320 empirical articles (27%) appearing in *PSPB* over these 3 years included one or more factor analyses, either an exploratory factor analysis, principal components analysis, or a confirmatory factor analysis. Although the change is not statistically significant, the frequency of factor analyses in these articles appears to be increasing slightly over time, from 26% in 1996 to 29% in 2000.

Why do investigators conduct factor analyses? My review of articles published in *PSPB* indicated that 80 of the 156 factor analyses (51%) that were reported over this 3-year period were performed for data reduction (i.e., reducing a set of items to a smaller set of more reliable measures). Another 60 factor analyses (39%) were conducted to test a hypothesized factor structure for a set of measures. Eleven (7%) of the remaining factor analyses were conducted to test a measurement model associated with a structural equation modeling analysis, using confirmatory factor analysis software. The remaining factor analyses involved an evaluation of the redundancy among a set of measures ($n = 4$, 3%) or a replication of results from a prior factor analysis ($n = 1$, 1%).

Despite the common use of these procedures, there appear to be a number of problems associated with both the use of these statistical procedures and the reporting of results of the analyses in these articles. The purpose of this article is to review the use of factor analysis in *PSPB* and to discuss both how investigators are using these techniques and ideally how such analyses should be conducted. My purpose is not to criticize any specific publications. Instead, I will focus on how researchers in general are using these procedures and ways in which this usage can be improved. In presenting this analysis, I will rely on recent work, especially Monte Carlo studies rele-

Author's Note: The author would like to thank Peter Bentler, Carolyn Cutrona, and K. A. S. Wickrama for their comments on an earlier version of this article. Preparation of this work was partially supported by Grant MH62666 from the National Institute of Mental Health to Carolyn Cutrona and Daniel W. Russell. Reprint requests should be sent to Daniel W. Russell, Institute for Social and Behavioral Research, Iowa State University, 2625 N. Loop Drive, Suite 500, Ames, IA 50010-8296; e-mail: drussell@iastate.edu.

PSPB, Vol. 28 No. 12, December 2002 1629-1646
DOI: 10.1177/014616702237645

vant to exploratory and confirmatory factor analysis and the article by Fabrigar et al. (1999) in *Psychological Methods* cited earlier. I also will express some of my own views concerning the use of these methods, as will be noted at various points in the discussion. The discussion will first focus on exploratory factor analysis and then will turn to confirmatory factor analysis.

EXPLORATORY FACTOR ANALYSIS

Factor Extraction Procedures

The first issue faced by any investigator planning an exploratory factor analysis concerns how to extract factors from the data. The classic factor analysis equation specifies that a measure being factored can be represented by the following equation:

$$x_1 = w_{11}F_1 + w_{21}F_2 + \ldots w_{n1}F_n + w_1U_1 + e_1,$$

where the *F*s represent the common factors that underlie the measures being analyzed and the *U*s represent factors that are unique to each measure. The *w*s represent loadings of each item (or measure) on the respective factors, whereas the *e*s reflect random measurement error in each item. Note that each measured variable has its own unique factor, reflecting systematic variance in the item or measure that is not shared with the other measures being analyzed.

On the basis of this equation, the variance in the measure being factored (i.e., σ^2_x) can be separated into three parts. First, there is a part of the variance in the measure that reflects the influence of the common factors, termed the communality of the variable. Second, there is a part of the variance that reflects the influence of the factor unique to that measure. Third, there is random error variance.

Two commonly used methods for extracting factors in the context of an exploratory factor analysis are principal components analysis and principal axis factoring.[2] Indeed, of the 137 exploratory factor analyses reported in *PSPB* during the years of 1996, 1998, and 2000, 85 (62%) employed principal components analysis to extract the factors, whereas 15 (11%) used principal axis factoring. (The remaining analyses either did not specify the method of extraction [36, or 26%] or used generalized least square extraction [1 analysis].) Principal components analysis and principal axis factoring involve the same procedure for extracting factors from the correlation matrix. Where they differ involves the estimation of communalities for the measured variables, or the variance that each measured variable shares with the other measured variables. In principal components analysis, the communalities for the measures are set at 1.0. In essence, this procedure assumes that all of the variance in a measure is potentially explicable by the factors (components) that are derived. It should be noted that because the communalities of the measures are set at 1.0, a principal components analysis extracts the factors based on the correlations among the measures (i.e., a correlation matrix is analyzed).

By contrast, in principal axis factoring, some estimate of the communality for each measure is employed, typically the squared multiple correlation between that measure and the other measures used in the analysis. Theoretically, this estimate of the communality reflects the variance in each measure due to the influence of the factors; one minus the communality reflects variance in each measure due to the unique factor and random error. Rather than extracting the factors using the correlations among the items, principal axis factoring extracts the factors using a reduced correlation matrix, where the 1.0 values on the diagonal of the correlation matrix are replaced by these initial communality estimates. In essence, one is analyzing a covariance matrix, where the variance of each measure reflects its association with the other measures included in the factor analysis.

In conducting a principal axis factor analysis, statistical programs such as SPSS or SAS will initially use the squared multiple correlation between each measure and the other measures being factored as the estimate of the communality for each measure.[3] After the number of factors is determined, a new estimate of the communality of each measure can be derived by squaring the loading of the measure on each factor and summing these squared loadings together. A second extraction of the factors is then conducted using this new estimate of the communality of each measure. This procedure continues until the estimates of the communalities converge (i.e., change minimally from one factor extraction to another).

Due to space limitations, I will not review how principal factoring is conducted; the interested reader should see the description of the procedure in books on factor analysis, such as Gorsuch (1983) or Comrey and Lee (1992). Some comments on the results of such an extraction procedure are in order, however, and apply to both principal components analysis and principal axis factoring. In extracting the factors, principal factoring derives loadings of the measures on the factors that maximize the variance explained by each factor. So, for example, loadings on the first factor that is extracted are designed to account for the maximum variance in the data matrix being analyzed (i.e., the correlation matrix in the case of a principal components analysis or the reduced correlation matrix in the case of principal axis factoring). Once this first factor has been extracted, a residual matrix is computed by calculating what the correlations between the measures should be based on the first factor[4] (i.e., the predicted correlation) and subtracting these values from the actual correlation between the measures. So, for example, if the actual correlation between the measures is .30 and the predicted correlation is .35, then the residual correlation is –.05. Factor 2 is then extracted based on this residual correlation matrix and, as a consequence of this extraction procedure, is orthogonal or uncorrelated with Factor 1. This procedure continues until all the variance in the initial matrix being analyzed is accounted for, which typically requires as many factors as there are measures being analyzed.

Table 1 provides the results from both methods of factor extraction for the 489 college students who completed the University of California, Los Angeles (UCLA), Loneliness Scale (Version 3) in Russell (1996). For this analysis, the items worded in a negative or non-lonely direction were reversed or "reflected" prior to the analysis; as a result, the 20 items from the scale were all positively correlated with one another. Examination of the loadings in Table 1 indicates one difference in the results of a principal components analysis and a principal axis factor analysis: Due to the higher communality estimates of the former analysis, the loadings are typically higher for the principal components analysis. Notice, however, that otherwise the loadings of items on the factors are very similar. Indeed, the correlation between the 20 loadings of the items on Component 1 and Factor 1 is 1.0, as is the correlation between the 20 loadings of the items on Component 2 and Factor 2.

An examination of the loadings on Component 1 and Factor 1 indicates that all of the items have strong positive loadings on this factor. Loadings of the items on Component 2 and Factor 2 indicate that these dimensions are bipolar, with some items having strong negative loadings and some items having strong positive loadings. Before one begins to develop elaborate theories of loneliness to account for this factor structure, it should be noted that this is an artifact of the factor extraction procedure. Given that all of the items on the loneliness scale are positively correlated with one another, the loadings on Component 1 and Factor 1 are of necessity positive because that maximizes the variance in the positive correlations explained by the first component or factor. After the predicted correlations among the 20 items due to Component 1 or Factor 1 are removed from the matrix being analyzed, the residual matrix will consist of some positive values (where the correlation between two items was underestimated) and some negative values (where the correlation between two items was overestimated). As a consequence, Component 2 or Factor 2 will tend to be bipolar, with a mixture of positive and negative loadings.

To illustrate this point, consider the correlation between Items 1 and 2 ($r = .13$). On the basis of the loadings of these two items on Component 1 in Table 1, the correlation between these two items is estimated to be .35 (i.e., .55 * .63). The residual correlation between Items 1 and 2 is therefore .13 – .35, or –.22. Given that this residual correlation is negative, it is not surprising that the loading of Item 1 on Component 2 is –.42 and the loading of Item 2 on Component 2 is .36. On the basis of these latter two loadings, the predicted correlation between the two items is estimated to be –.15, leading to a residual correlation of –.07 (i.e., –.22 – [–.15]).

TABLE 1: Unrotated Factor Loadings From the Principal Components and Principal Axis Factoring Analyses

Item	*Component 1*	*Component 2*	*Factor 1*	*Factor 2*
1	.55	–.42	.53	–.32
2	.63	.36	.61	.30
3	.70	.07	.68	.06
4	.66	.28	.64	.24
5	.61	–.35	.59	–.28
6	.61	–.37	.59	–.30
7	.67	.23	.65	.19
8	.56	.25	.53	.19
9	.51	–.37	.49	–.27
10	.66	–.33	.64	–.28
11	.60	.33	.58	.27
12	.63	.35	.61	.29
13	.71	.16	.69	.15
14	.73	.21	.72	.19
15	.61	–.15	.58	–.12
16	.68	–.15	.66	–.13
17	.36	.27	.33	.17
18	.60	.29	.57	.22
19	.66	–.35	.64	–.31
20	.68	–.32	.66	–.28

Is it important which factor extraction procedure you use? Given that both principal components analysis and principal axis factoring employ the same method of extracting factors, the issue revolves around the estimation of the communalities for the measures. Some writers, such as Velicer and Jackson (1990), have argued that one gets very similar results using the two factor extraction procedures, particularly if one is analyzing a large number of measures with relatively high communalities (i.e., squared multiple correlations). A number of simulation studies (Bentler & Kano, 1990; Schneeweiss, 1997; Schneeweiss & Mathes, 1995) have demonstrated the equivalence of derived components and factors in the context of an infinite sample of measured variables. Finally, a recent study by Ogasawara (2000) has generated the asymptotic correlations between components and factors for both actual and simulated data sets, demonstrating when the correlation between factors and components should equal 1.0.

However, a Monte Carlo analysis by Widaman (1993) indicated that principal axis factoring using squared multiple correlations as the initial estimate of the communalities and then iterating this estimate to a final communality estimate provided more accurate results in terms of the population factor loadings than principal components analysis. His findings suggest that in certain situations the extraction procedure could lead to substantive differences in the factor loadings. Therefore, researchers would be wise to use principal axis factoring rather than principal components analysis.

As illustrated in Table 1, in my experience, the main difference between extractions based on principal components analysis and principal axis factoring involves the magnitude of the factor loadings. Why do researchers publishing in *PSPB* tend to use principal components analysis rather than principal axis factoring? One reason may be that the default factor extraction procedure in both SPSS and SAS is principal components analysis. If you choose to conduct a principal axis factor analysis, both of these statistical packages employ the squared multiple correlations as the initial communality estimate and iterate to a final communality estimate for each measure, as recommended by Widaman (1993).

Sample Size

The traditional rule concerning the number of participants to include in a factor analysis has focused on the number of cases relative to the number of measures that are being factored. Minimums of 5 or 10 cases per measure have typically been recommended (Comrey & Lee, 1992; Gorsuch, 1983). Based on my review of exploratory factor analyses appearing in *PSPB* during the years 1996, 1998, and 2000, 19 of the exploratory factor analyses (14%) involved fewer than five cases per measure and 58 of the exploratory factor analyses (42%) involved fewer than 10 cases per measure. Clearly, a large number of the exploratory factor analyses in *PSPB* involve less than the recommended number of cases relative to the number of items.

Recently a Monte Carlo study by MacCallum, Widaman, Zhang, and Hong (1999) examined the issue of sample size in exploratory factor analysis. These investigators tested the ability of an exploratory factor analysis to reproduce the population factor loadings for different sample sizes and variation in the communalities of the variables. They found that results were very consistent with the population loadings even with sample sizes as low as 60 cases if the communalities of the items were high (e.g., .60 or greater). With lower communality levels (e.g., around .50), samples of 100 to 200 cases were required to accurately reproduce the population loadings.

My review of exploratory factor analyses reported in *PSPB* indicated that 54 of the factor analyses (39%) involved samples of 100 or fewer cases and another 31 analyses (23%) involved from 100 to 200 cases. Unfortunately, very few of these researchers provided any information on the communalities associated with the measures, and none of the reports included information on the communalities for all of the measures that were analyzed. Therefore, it is impossible to judge how appropriate the sample sizes were for the exploratory factor analyses that were conducted.

Number of Measures Per Factor

A related issue involves the number of measures that are employed in an exploratory factor analysis relative to the number of factors that are extracted. This is the issue of identification, or having a sufficient number of measures that load on each factor to be able to adequately operationalize the factor. At least three items per factor are required for a factor model to be identified; more items per factor results in overidentification of the model. A number of writers recommend that four or more items per factor be included in the factor analysis to ensure an adequate identification of the factors (Comrey & Lee, 1992; Fabrigar et al., 1999; Gorsuch, 1988).

MacCallum et al. (1999) found that in addition to the communality of the measures, the results were more accurate for given sample sizes if there were more measures per factor included in the analysis. Therefore, it appears wise to test overidentified factor models where the investigator includes four or more measures per factor in the analysis.

On the basis of their review of exploratory factor analyses reported in the *Journal of Personality and Social Psychology* (*JPSP*) during the years 1991 to 1995, Fabrigar et al. (1999) reported that 18% of the analyses involved three or fewer items per factor. I found that 33 of the exploratory factor analyses in *PSPB* (25%) during the years 1996, 1998, and 2000 included three or fewer measures per factor. Clearly, many investigators need to include a larger number of measured variables per factor. One can argue, of course, that a researcher does not necessarily know how many factors will emerge from an exploratory factor analysis. As discussed further below, however, I would maintain that many investigators have a prediction regarding the likely factor structure of the measures they are analyzing prior to the factor analysis.

Determining the Number of Factors

Once factors have been extracted from a correlation matrix, one has to decide how many factors to retain as being meaningful or important. By default, SPSS will retain factors that have eigenvalues ≤ 1.0; this is sometimes referred to as the Kaiser criterion.[5] The eigenvalues refer to the amount of variance explained by a factor and are computed by squaring the loadings on a factor and summing them together. Although SPSS uses this criterion no matter what technique is used to extract the factors, in fact this criterion should only be used when principal components analysis (with communalities fixed at 1.0) is used as the extraction procedure (see discussion by Gorsuch, 1983).

The eigenvalue ≥ 1.0 criterion is also used by SAS to determine the number of factors to extract when conducting a principal components analysis. However, a different procedure is used by SAS in determining the number of factors when conducting a principal axis factor analysis. Specifically, the amount of common factor variance across the items is determined based on the initial communality estimates for the items. So, for example, in the factor analysis of the items from the UCLA Loneliness Scale, the initial communality of the items (based on the squared multiple correlations) ranged from .21 to .56 ($M = .44$). The sum of these values was 8.82, which represents an initial estimate of the total common factor variance of the 20 items. Factors were then extracted by SAS from the reduced correlation matrix, with the number of factors determined by how many were required to account for this common factor variance. In the analysis of data from the loneliness scale, three factors were required to account for the common factor variance and were therefore extracted by the SAS program.

As noted by Fabrigar et al. (1999), the eigenvalue ≥ 1 criterion often leads to extracting too many factors or overfactoring. To understand why this would be true, imagine a situation where you are factoring a set of 5 measures versus a set of 20 measures. In the case of 5 measures, a factor that had an eigenvalue of 1.0 would be accounting for 20% of the total variance in the measures (i.e., 1/5, given that the total variance of the measures equals 5 in the context of a principal components analysis). By contrast, with a set of 20 items, a factor that had an eigenvalue of 1.0 would be explaining 5% of the variance (i.e., 1/20). Clearly, when one is factoring a large set of items it is more likely that using this criterion will lead to extracting factors that account for only a small amount of the total variance.

My review of factor analyses reported in *PSPB* indicated that a large number of the analyses (74 of 134, or 55%) did not indicate the criteria used to determine the number of factors. Of the remaining analyses, 31 of 60 (52%) used the eigenvalue ≥ 1.0 criterion. However, only 18 of these 31 analyses (58%) used this criterion in the context of a principal components analysis. Therefore, more than 40% of these factor analyses employed this criterion when it was inconsistent with the factor extraction procedure.

A third criterion that is often used to determine the number of factors is the scree test attributed to Cattell (1966). This procedure was used by 18 of the 60 exploratory factor analyses (30%) reported in *PSPB* where the criterion for determining the number of factors was indicated. To use this method, one plots the eigenvalues of the factors extracted either from the correlation matrix (i.e., with 1.0 in the diagonal) or the reduced correlation matrix (with communalities in the diagonal) in descending order. Computer programs such as SPSS or SAS will plot these values. One then looks for a break in the values, where there is the last substantial drop in the eigenvalues. The number of factors prior to this drop represents the number of factors to be extracted. Figure 1 shows the plot of eigenvalues from the principal axis factor analysis of the 20 items from the UCLA Loneliness Scale. Clearly, there is a substantial drop in the eigenvalues after Factor 2, suggesting that two factors should be extracted.

Use of this procedure has been criticized by a number of writers due to its subjectivity (e.g., Kaiser, 1970). In some cases, there may be no clear break in the eigenvalues, with a linear decline in the values occurring from the largest to the smallest factor. This can occur, for example, where several factors are extracted on which pairs of measures load highly. However, as discussed by Fabrigar et al. (1999), studies have indicated that examining the scree plot for breaks provides a reasonably accurate indication of the number of factors.

A fourth procedure that appears to perform better than the scree test in determining the number of factors is termed a parallel analysis (see discussion by Reise, Waller, & Comrey, 2000). This is a variant on the scree test, where one also plots the eigenvalues derived from factoring a completely random set of data involving the same number of items and research participants. The number of factors to extract is indicated by the point at which the eigenvalues for the actual data drop below the eigenvalues for the random data. Unfortunately, neither SPSS nor SAS provide this information in the context of plotting the eigenvalues from a factor analysis. Reise et al. (2000, pp. 290-291) describe how one can obtain a computer program that will generate these random values and also include the commands to generate the average eigenvalues.[6] Figure 1 presents a plot of these values for the factor analysis of the loneliness scale items. As can be seen, the eigenvalues drop below the line defined by the average eigenvalues from the random data for three or more factors. Therefore, these results also suggest that two factors should be extracted for the loneliness scale.[7]

Factor Rotation

Rotation of the factors involves reorienting them or altering the location of the factors in the dimensional space to improve the interpretability of the results. Two types of rotations can be conducted. An orthogonal rota-

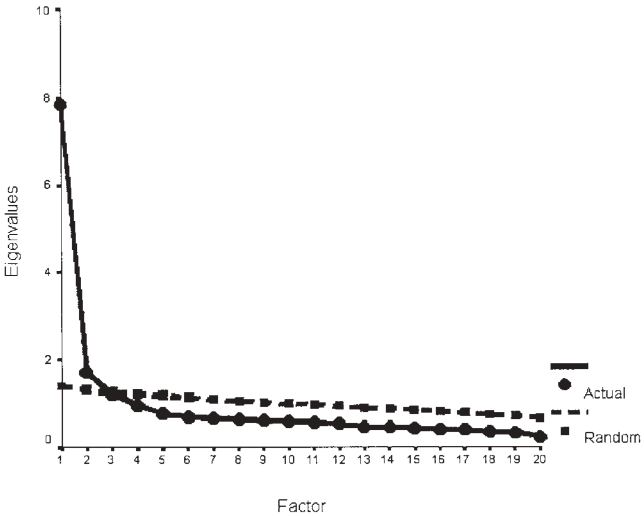

Figure 1 **Plot of the eigenvalues from the factor analysis of the loneliness scale items.**

TABLE 2: Varimax Rotated Factor Loadings From the Principal Components and Principal Axis Factoring Analyses

Item	*Component 1*	*Component 2*	*Factor 1*	*Factor 2*
1	.11	.68	.16	.60
2	.71	.17	.65	.20
3	.55	.43	.53	.42
4	.68	.25	.64	.26
5	.21	.67	.24	.61
6	.19	.69	.23	.62
7	.64	.30	.60	.31
8	.58	.20	.51	.22
9	.12	.62	.17	.53
10	.25	.69	.27	.64
11	.67	.17	.61	.20
12	.70	.18	.65	.20
13	.63	.37	.60	.37
14	.68	.35	.65	.35
15	.35	.53	.35	.48
16	.39	.58	.39	.55
17	.44	.05	.35	.10
18	.63	.21	.57	.23
19	.24	.70	.26	.66
20	.28	.70	.29	.66

tion involves rotating the factors that have been extracted, with the constraint that the factors continue to be uncorrelated with one another (i.e., kept at right angles to one another when the results are plotted). Of the 88 exploratory factor analyses that extracted two or more factors published in *PSPB* during the years 1996, 1998, and 2000, 46 (52%) conducted an orthogonal rotation of the extracted factors. A second type of rotation is an oblique rotation, wherein the constraint that the factors be uncorrelated with one another is relaxed. Twenty of the 88 factor analyses (23%) conducted an oblique rotation of the extracted factors. Eight of the analyses (9%) did not involve a rotation of the extracted factors, and the remaining analyses (14, or 16%) failed to indicate which type of rotation was used.

Each type of rotation will be described below. The rotation procedures will be illustrated with data from the factor analysis of the UCLA Loneliness Scale described above.

Orthogonal rotation. It is important to realize that the original location of the factors prior to rotation is completely arbitrary. As was noted previously, factors are extracted so as to maximize the variance in the measures that is accounted for by each succeeding factor. There are an infinite number of orientations of the factors that account for the data (i.e., the association among the variables) equally well. To illustrate this point, the components and factors shown in Table 1 were rotated using the Varimax procedure. This is the orthogonal rotation procedure that is recommended by Fabrigar et al. (1999) and that was used by all the investigators who conducted orthogonal rotations published in *PSPB*. The Varimax procedure attempts to achieve "simple structure," wherein each of the measures tends to load highly on some of the factors and have low loadings on the other factors. As the "Varimax" name implies, such a pattern of loadings will tend to maximize the variance of the squared loadings on any given factor.

The results of the Varimax rotation of the principal components and principal factors are presented in Table 2. These loadings along with the original unrotated loadings are plotted in Figure 2; each dot in the figure indicates the location of one of the items in the two-dimensional space. The dashed axes show the orientation of the original unrotated factors, with the factor number indicated in parentheses. Solid axes indicate location of the Varimax rotated factors. From this plotting of the two sets of factors one can see how the location of the original factors is rotated to arrive at the final Varimax orientation of the factors.

A common error made by investigators involves reporting the variance explained by the factors before and after rotation. Table 3 presents information on the variance in the 20 items explained by both the unrotated components and factors presented in Table 1 and the Varimax rotated components and factors presented in Table 2. In calculating the variance explained by each component or factor, it is important to recognize that the loadings represent correlations between each measure (item) and the respective component or factor. Therefore, the variance explained by a factor is computed by squaring the loadings on the factor (component) and summing them together and then dividing that value by 20 (the total variance in the measures) to convert that value to a proportion. For the unrotated fac-

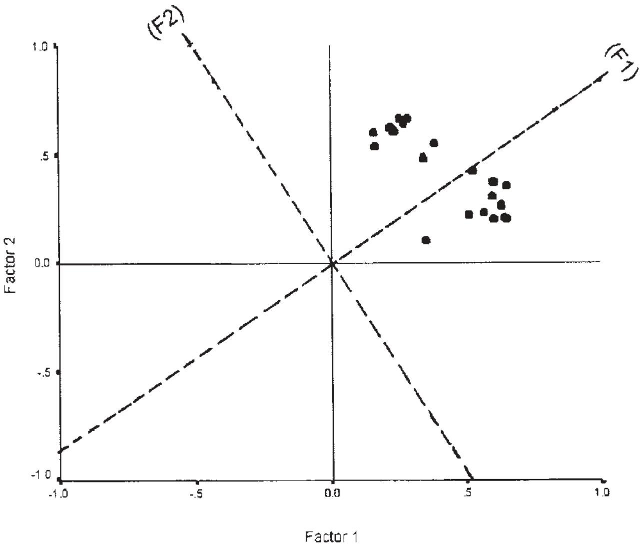

Figure 2 **Plot of the factor loadings following a Varimax rotation, with the original factor prior to rotation indicated.**

TABLE 3: Variance Explained by the Unrotated and Varimax Rotated Factors

	Principal Components		*Principal Axis Factors*	
Factor	*Unrotated*	*Rotated*	*Unrotated*	*Rotated*
1	39.34	24.82	36.56	22.09
2	8.61	23.13	5.74	20.21
Total	47.95	47.95	42.30	42.30

tors, these values are identical to the eigenvalues for the two factors presented in Figure 1.

Two important aspects of the results presented in Table 3 should be noted. First, both the unrotated and rotated factors are uncorrelated with one another, given that an orthogonal rotation was conducted. Therefore, the factors do not overlap in the variance they explain in the items that were analyzed. Second, the total variance in the items that is explained by the two unrotated and the two rotated factors is identical. So, for example, a total of 42% of the variance in the loneliness scale items is accounted for by the two Principal Axis factors that were extracted both before and after rotation. Thus, with rotation there is no change in the variance explained by the factors. However, the variance explained by each factor is changed dramatically. For the unrotated solution, Factor 1 accounted for 86% of the explained variance, whereas Factor 2 accounted for 14% of the explained variance. Following the Varimax rotation, Factor 1 accounted for 52% of the explained variance, whereas Factor 2 accounted for 48% of the explained variance. A consequence of rotation is to spread the explained variance from the first factor, which always appears to be the most important given the method of factor extraction that is used, to later factors. Therefore, the variance explained by the factors prior to rotation indicates nothing about their importance following rotation.

Researchers who conduct an exploratory factor analysis and subsequently conduct an orthogonal rotation of the factors, such as Varimax, need to report the variance explained by the factors both before and after rotation. Of the 46 exploratory factor analyses reported in *PSPB* during the years 1996, 1998, and 2000 that used a Varimax rotation, none of the reports indicated the variance explained by the factors after rotation. Because the interpretation of the factors is based on the rotated solution, the variance explained by the factors after rotation should be reported. It should be noted that both SPSS and SAS report the variance explained by the extracted factors both before and after an orthogonal rotation.

Oblique rotation. As noted above, an oblique rotation allows the rotated factors to be correlated with one another. One consequence of such a rotation is that the resultant factors overlap to some degree in the variance they explain in the measures that are being analyzed. In essence, the resultant factors reflect predictor variables that are not independent of one another.

Writers such as Fabrigar et al. (1999) recommend that investigators use procedures such as Promax when conducting an oblique rotation. The reason is that this procedure initially conducts a Varimax rotation and then relaxes the constraint that the factors are uncorrelated with one another to improve the fit to simple structure. If it is the case that factors that are uncorrelated (or nearly so) fit the data well, then this rotation will result in factors that are close to orthogonal to one another. Another oblique rotation procedure recommended by Fabrigar et al. (1999) is Direct Quartimin, which is equivalent to a Direct Oblimin rotation with the delta parameter (which controls the level of factor correlation that is permitted) set at 0. Of the 20 exploratory factor analyses published in *PSPB* that conducted an oblique rotation of the extracted factors, two (10%) used the Promax procedure. Eight of the remaining analyses (40%) used the Oblimin procedure. Unfortunately, none of these investigators reported the value of the delta parameter that they used. Finally, 10 of the factor analyses (50%) did not indicate which oblique rotation procedure was used.

To illustrate the use of an oblique rotation, consider the factor analysis of the UCLA Loneliness Scale presented above. The results of the Varimax rotation of the two factors derived from the loneliness scale shown in Figure 2 were not consistent with simple structure. That is, the factors are not oriented so that the items tend to load highly on one factor and near zero on the other factor. Instead, most items appear to have positive loadings on both factors. This pattern of loadings suggests that an improvement in the orientation of the factors relative to the items could be achieved if the factors were not forced to be at right (90°) angles to one another. Therefore, the loadings on the two factors for the loneliness scale were rotated using the Promax procedure.

The results of an oblique rotation using the Promax procedure resulted in a correlation of .63 between the two principal components and .70 between the two factors derived from the Principal Axis Factoring analysis. Due to the correlation between the factors following an oblique rotation, both the SPSS and SAS programs report two matrices. The Factor Structure matrix provides the correlation between each of the measures and the factors that have been extracted and rotated; this is, of course, what we typically think of as factor loadings. However, given that the two factors are correlated with one another, there may be overlap in these loadings. Therefore, the second matrix, termed the Factor Pattern matrix, is designed to indicate the independent relationship between each measure and the factors. One can think of the values reported here as being equivalent to standardized regression coefficients, where the two factors are used as predictors of each measure.

Table 4 reports the Factor Structure matrix for the Principal Components and the Principal Axis Factoring analyses, whereas Table 5 reports the Factor Pattern matrix for these two analyses of the loneliness scale. As can be seen in Table 4, there were strong positive correlations between each of the items and the factors. The Factor Pattern matrix shown in Table 5 is more useful in interpreting the meaning of the factors. So, for example, items 1, 5, 6, 9, 10, 15, 16, 19, and 20 appear to load more highly on Factor 2, whereas the remaining items appear to load more highly on Factor 1. Factor 1 represents items worded in a negative or lonely direction, whereas Factor 2 represents items loaded in a positive or non-lonely direction. Thus, it appears that these two factors reflect a method factor corresponding to the direction of item wording. (See Russell [1996] for a further discussion of the factor structure underlying the loneliness scale.)

Figure 3 shows a plot of the factor loadings based on the Varimax rotation, which is identical to that shown in Figure 2. Now, however, I have also plotted the location of the two factors derived from the Promax rotation, shown by the dashed lines in Figure 3; the factor numbers are indicated in parentheses. As can be seen, these factors are no longer at right angles (90°) to one another, reflecting the correlation between the factors.

TABLE 4: Factor Structure Matrices Following the Promax Rotation From the Principal Components and Principal Axis Factoring Analyses

Item	*Component 1*	*Component 2*	*Factor 1*	*Factor 2*
1	.34	.68	.38	.61
2	.73	.39	.68	.43
3	.67	.59	.65	.59
4	.72	.46	.69	.48
5	.42	.70	.45	.65
6	.41	.72	.45	.66
7	.71	.49	.67	.51
8	.61	.38	.56	.40
9	.32	.62	.35	.56
10	.47	.74	.50	.70
11	.69	.38	.64	.41
12	.72	.40	.67	.43
13	.72	.56	.70	.57
14	.76	.56	.74	.57
15	.50	.61	.50	.58
16	.56	.68	.57	.65
17	.43	.19	.37	.23
18	.66	.40	.61	.43
19	.46	.74	.49	.71
20	.50	.75	.52	.72

TABLE 5: Factor Pattern Matrices Following the Promax Rotation From the Principal Components and Principal Axis Factoring Analyses

Item	*Component 1*	*Component 2*	*Factor 1*	*Factor 2*
1	–.15	.78	–.11	.69
2	.79	–.10	.74	–.09
3	.49	.28	.46	.27
4	.71	.01	.68	.00
5	–.03	.72	–.01	.66
6	–.05	.75	–.03	.69
7	.66	.08	.62	.08
8	.61	–.01	.55	.02
9	–.11	.70	–.07	.60
10	.02	.73	.02	.69
11	.74	–.08	.68	–.06
12	.77	–.08	.73	–.08
13	.61	.18	.59	.16
14	.67	.13	.66	.11
15	.20	.49	.20	.44
16	.23	.53	.22	.50
17	.51	–.13	.40	–.05
18	.68	–.02	.62	.00
19	–.01	.75	–.01	.72
20	.04	.72	.03	.70

Factor Scores

It is often the case that an exploratory factor analysis is conducted as a data reduction procedure. As noted previously, 51% of the factor analyses that appeared in *PSPB*

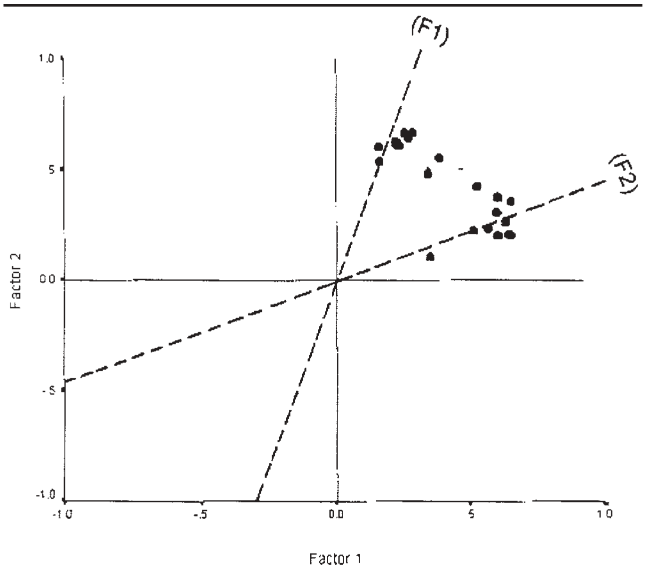

Figure 3 Plot of the factor loadings following a Promax rotation, with the location of the factors following a Varimax rotation also indicated.

during the 3 years being examined were conducted for this reason. So, for example, an investigator may have included a number of measures to assess a certain type of variable, such as affect, and he or she may want to conduct an exploratory factor analysis to identify the dimension(s) underlying these measures. Once these dimensions are identified, the investigator will typically create scores to represent the factor or factors, which can in turn be used in subsequent analyses of the data. Such scores should be more reliable than the original measures and also should yield scores that are less highly correlated with one another than the original measures.

Although such a factor score may be more reliable, it should be noted that it is still a measured variable. That is, this computed score is not identical with the factor or latent variable. As a consequence, the results of any subsequent analyses with this new measured variable will be attenuated by random error in the factor score. An alternative approach is to conduct subsequent analyses involving the set of variables using latent variable modeling procedures, which remove the biasing effects of random measurement error from the analyses.

If an investigator insists on computing factor scores, how should these scores be created? The factor analysis procedures included in the SPSS and SAS programs provide a number of different methods for computing factor scores. I recommend, however, that researchers choose a simpler strategy of identifying the measures or items that load highly on a given factor and summing together scores on those measures to create a score reflecting that factor. (One would, of course, reverse scores on a measure that loaded negatively on the factor prior to creating such a score.) In using this procedure the investigator is losing the information provided by the loadings of the measures on the factors. However, the score created in the manner I am suggesting will correlate highly with the various "weighted" scores provided by the factor analysis programs. Given that the factor score weights that are derived from a factor analysis are likely to be sample-specific and therefore not replicable, a more reasonable strategy involves simply identifying the measures or items that load highly on a factor and summing them together.[8]

Recommendations

On the basis of this review and analysis of exploratory factor analyses that have been reported in *PSPB* during the past 5 years, a number of recommendations can be formulated. These are detailed in the following sections.

Study design. There appears to be a general need to increase the size of the samples. Unless a researcher can demonstrate a high level of communality (.60 or greater) for the measures used in a factor analysis, a sample of at least 100 cases should be used. Investigators also need to ensure that they have a sufficient number of measures relative to the number of factors that are derived from the analysis; at least three measures per factor is required, and preferably four or more measures per factor should be included in the analysis.

Factor extraction. Although many researchers believe there is little difference between Principal Components analysis and Principal Axis Factoring in the results that are obtained, the findings reported by Widaman (1993) indicate that results based on Principal Axis Factoring are more accurate in reproducing the population loadings. Therefore, this appears to be the preferable method of factor extraction.

Determining the number of factors. There is no simple answer to the problem of determining the number of factors to extract, given that available computer programs do not provide results for random data sets as described by Reise et al. (2000). The default method of extracting factors with eigenvalues ≥ 1.0 is clearly not accurate. Although the scree test is more accurate, the subjectivity of this criterion is problematic. Conducting a parallel analysis using random data sets is recommended, although that will require using the program developed by Reise et al. (2000) to generate the required data.

Factor rotation. Although orthogonal rotations simplify the presentation and interpretation of factor analysis results, they often do not lead to simple structure due to underlying correlations between the factors. I therefore agree with the recommendation of Fabrigar et al. (1999)

that investigators conduct an oblique rotation such as Promax. This procedure first conducts an orthogonal Varimax rotation and then allows correlations between the factors in an attempt to improve the fit to simple structure. Therefore, if the factors are in fact uncorrelated with one another, that will be revealed by a Promax rotation.

Factor scores. Finally, if an investigator insists on computing scores to represent the factors, I recommend that a simple unweighted procedure be used to derive scores to represent factors that are discovered to underlie a set of measures. In that way one is not relying on the replicability of the weights that are derived from a factor analysis on a single sample. Alternatively, latent variable modeling procedures may be used in subsequent analyses to provide results involving the underlying dimensions or factors that are not attenuated by random measurement error.

CONFIRMATORY FACTOR ANALYSIS

Confirmatory factory analysis (CFA) is designed to assess how well a hypothesized factor structure "fits" the observed data. Unlike exploratory factor analysis, the researcher has an explicit prediction concerning both the number of factors that underlie a set of measures and which measures load on the hypothesized factor(s). These factor loadings are equivalent to the factor pattern coefficients described above in the context of exploratory factor analysis, reflecting unique relationships between each measured variable and the underlying factor(s). Once loadings of the measures on the factors are derived and (if an oblique factor structure is hypothesized) correlations among the factors are estimated, a chi-square goodness-of-fit test is conducted. If the hypothesized factor model fits the data, then the goodness-of-fit test will be nonsignificant. Therefore, in a reversal of the statistical logic that is typically employed, one's hope is to find a nonsignificant result.

In a confirmatory factor analysis, measures typically have estimated non-zero loadings on their respective factors and zero loadings on the other factors. To illustrate this point, Table 6 presents the factor loading matrix for items from the locus of causality and stability subscales from the Revised Causal Dimension Scale (McAuley, Duncan, & Russell, 1992). In conducting a confirmatory factor analysis for that measure, loadings of the items on the factors would be estimated where there are "Xs" in Table 6. So, for example, the three items designed to measure locus of causality would be allowed to load on the Locus factor. Their loadings on the Stability factor would be fixed at 0. This contrasts with the case of an exploratory factor analysis, where the items would have non-zero loadings on all the factors that were found to underlie the measures. So, for example, if two factors were extracted via an exploratory factor analysis procedure for these six items, each item would have a loading on both of the factors.

TABLE 6: Factor Loading Matrix for the Causal Dimension Scale

Item	*Locus*	*Stability*
L1	X	O
L2	X	O
L3	X	O
S1	O	X
S2	O	X
S3	O	X

NOTE: L1, L2, and L3 represent the three items from the Locus scale; S1, S2, and S3 represent the three items from the Stability scale. Loadings of the items on the factors were estimated where an "X" is included in the factor-loading matrix, whereas no loading was estimated where an "O" is indicated in the factor-loading matrix.

Maximum likelihood estimation is the method typically used to estimate a confirmatory factor model.[9] Using this method, programs such as LISREL (Jöreskog & Sörbom, 1996) or EQS (Bentler, 1995) will generate a set of initial factor loadings or start values for the hypothesized model. The fit of these values to the data is then evaluated based on the following equation:

$$F = \{\log(\det[\Sigma])\} + \{\text{trace}(S * \Sigma^{-1})\} - \{\log(\det[S])\} - p,$$

where Σ represents the reproduced covariance matrix (or what the covariances among the measures should be based on the factor model), S represents the observed covariance matrix based on the data, and p represents the number of measured variables. Once the fit of the initial factor model to the data is evaluated, factor loadings are adjusted based on lack of fit of the model to the data and a new value of F is computed. This continues until F converges or changes a minimal amount from one iteration to another. At that point, the chi-square statistic for the final factor model is computed, based on the following formula:

$$\chi^2 = (N - 1)F,$$

where *N* represents the number of cases included in the analysis. The degrees of freedom associated with the chi-square statistic are computed from the following formula:

$$df = 1/2(p)(p+1) - t,$$

where t represents the number of free parameters estimated in testing the hypothesized factor model. These

TABLE 7: Covariances and Correlations Among the Items

	L1	*L2*	*L3*	*S1*	*S2*	*S3*
L1	5.49	.44	.36	.13	.04	–.02
L2	2.23	4.71	.45	.05	–.01	–.07
L3	1.55	1.80	3.40	–.04	–.06	–.05
S1	.70	.28	–.17	5.67	.47	.39
S2	.21	–.03	–.24	2.41	4.58	.36
S3	–.09	–.26	–.16	1.69	1.39	3.36

NOTE: Variances of the items are shown on the diagonal, with covariances among the items shown below the diagonal and correlations among the items shown above the diagonal. L1, L2, and L3 represent the three items from the Locus scale; S1, S2, and S3 represent the three items from the Stability scale.

TABLE 8: Start Values and Final Estimates of the Factor Loadings for the Causal Dimension Scale

	Start Values		*Final Estimates*	
Items	*Locus*	*Stability*	*Locus*	*Stability*
L1	.57	.00	.59	.00
L2	.84	.00	.74	.00
L3	.58	.00	.61	.00
S1	.00	.76	.00	.72
S2	.00	.65	.00	.66
S3	.00	.53	.00	.54

NOTE: L1, L2, and L3 represent the three items from the Locus scale; S1, S2, and S3 represent the three items from the Stability scale. The standardized factor loadings (i.e., correlations between the items and the factors) are presented.

free parameters include factor loadings, random error terms, and correlations or covariances among the factors.

An Example of a CFA

To illustrate the use of a confirmatory factor analysis, the model for the Revised Causal Dimension Scale shown in Table 6 was tested based on data from 380 college students analyzed by McAuley et al. (1992). The covariance matrix that was used in the analyses is presented in Table 7, with the correlations among the items shown above the diagonal of the matrix. The initial factor model specified loadings of the items on the two factors as shown in Table 6, with the factors specified as uncorrelated (orthogonal). This hypothesized model was fit to the variances and covariances among the variables shown in Table 7 using the maximum likelihood procedure in LISREL 8.3 (Jöreskog & Sörbom, 1996). Table 8 presents the start values for the factor loadings along with the final estimates. On the basis of the start values, the value of the fit function (F) was .10468, resulting in $\chi^2(9, N = 380) = 39.68$, $p < .001$. After iteration to the final estimates of the factor loadings, the value of F was .03810, resulting in $\chi^2(9, N = 380) = 14.44$, $p = .11$. This latter result indicates that the model fits the data, given that the chi-square statistic is nonsignificant.

As noted above, this first model specified that the factors were orthogonal, or that the correlation between the two factors was zero. An obvious modification of this model that may improve the fit to the data would involve permitting these two factors to be correlated. Therefore, a second model that allowed the two factors to be correlated (i.e., an oblique factor structure) was fit to these data. In this model, I allowed the program to generate a value for the correlation between the factors that maximizes the fit of the model to the data. The results for this second model indicated that it also fit the data well, $\chi^2(8, N = 380) = 14.39$, $p = .07$. However, the correlation between the two factors was nonsignificant, $r = .02$.

Because the second model involves estimating an additional parameter (i.e., the correlation between the factors), there are eight degrees of freedom. Bentler and Bonett (1980) noted that in the case of two nested models (i.e., where one model involves freeing up one or more parameters in the second model), the difference in the chi-square statistics for the two models is itself distributed as a chi-square. Therefore, we can evaluate the effect of adding this parameter (correlation) between the two factors to the fit of the model by computing a chi-square difference test. The result is $\chi^2(1, N = 380) = .05$, $p = .82$, indicating that adding the correlation between these two factors did not lead to a significant improvement in the fit of the model to the data.

Use of CFA in PSPB

Nineteen of the 156 factor analyses (12%) reported in *PSPB* during the years 1996, 1998, and 2000 included a confirmatory factor analysis. There was clear evidence that the use of this statistical method is increasing over time. Of the 85 articles that reported a factor analysis, none of the 28 articles published during 1996 included a confirmatory factor analysis, whereas 2 of the 23 articles published in 1998 (9%) and 10 of the 34 articles published in 2000 (29%) used this statistical method.

As noted above, to conduct a confirmatory factor analysis one must have a clear prediction as to the factor structure underlying a set of measures. In the case of 10 of these 19 CFAs (53%), the investigators were testing the fit of a hypothesized factor structure to a set of measures. The remaining 9 CFAs involved the testing of a measurement model prior to conducting an analysis of a structural equation model. Nearly half of these analyses (9 of 19, or 47%) were conducted using the LISREL program, with 7 of the 19 analyses (37%) using the EQS program; the remaining 3 analyses did not indicate the program that was used. Maximum likelihood estimation was used for 8 of the 19 analyses (42%). The remaining 11 analyses (58%) did not indicate the estimation method that was used. Finally, despite the fact that the APA Publication Manual indicates that the correlation or covariance matrix that was analyzed should be reported, only 11 of the 19 analyses (58%) included this matrix.

Evaluating Goodness-of-Fit

An important issue in the literature on confirmatory factor analysis concerns the evaluation of model fit to the data. One problem involves the impact of sample size. As can be seen from the formula for computing the chi-square statistic, the number of cases employed in the analysis can affect one's conclusion concerning model fit. For example, the model for the Revised Causal Dimension Scale shown above was found to fit the data based on a nonsignificant chi-square statistic. That result was obtained for a sample of 380 cases. What would the result have been if I had used data from a sample of 1,000 cases? Assuming that the same value of the fit function (F) would have resulted (i.e., .03810), then the chi-square statistic would have increased to 38.06, which is highly significant with 9 degrees of freedom, $p < .001$. Therefore, I would have concluded that the hypothesized factor model did not fit the data.

Due to the influence of sample size and other characteristics of the data (such as non-normality) on the chi-square goodness-of-fit test, a variety of measures of model fit have been developed. These efforts have been focused on developing indicators of model fit that are relatively unaffected by variations in the size of the sample that is used in testing a model and non-normality of the data. As a consequence, programs for conducting confirmatory factor analysis now report a large number of indicators of model fit in addition to the chi-square test. For example, LISREL 8.3 reports 20 measures of model fit, whereas EQS 5.7 reports 10 measures of model fit. Hu and Bentler (1998) provide a description of these different measures of model fit, along with the formulas used in computing each statistic. Which of these indicators of fit should be used in evaluating your factor model?

A number of studies have appeared over the past two decades evaluating the extent to which these various measures of model fit are affected by variations in sample size and normality of the data (for a review, see Hu & Bentler, 1998). Recently, Hu and Bentler (1998, 1999) have examined the sensitivity of these measures of fit to miss-specification of the factor model. That is, if an incorrect factor model is being tested, how likely is it that a measure of model fit will reject that model? On the basis of an extensive Monte Carlo analysis, Hu and Bentler (1999) recommend a "two criteria" strategy in evaluating model fit. First, they advise using the standardized root mean square residual (SRMSR) in evaluating the model. In computing this statistic, S (the matrix being analyzed) and Σ (the relationships among the variables based on the model) are first standardized or converted to a correlation matrix and then the residual matrix (i.e., the difference between S and Σ) is computed. The average squared residual is calculated as an indication of how well the model fits the data, with a value of .08 or less indicating a good fit to the data. Second, they recommend the use of one of several fit statistics, such as the Tucker-Lewis Index (TLI) (Tucker & Lewis, 1973), Bollen's (1989) Index (IFI), the Comparative Fit Index (CFI) (Bentler, 1990), the Relative Noncentrality Index (RNI) (McDonald & Marsh, 1990), Gamma Hat (Steiger, 1989), McDonald's (1989) Centrality Index (MFI), or the Root Mean Square Error of Approximation (RMSEA) developed by Steiger and Lind (1980). Of importance, Hu and Bentler (1999) also indicate that the criteria used in evaluating model fit for many of these latter statistics should be increased. For example, for the TLI, IFI, CFI, RNI, and Gamma Hat, the widely used criterion of .90 or greater should be increased to .95 or greater. For MFI, the criterion should be .90 or greater, and for RMSEA, the criterion should be .06 or lower.[10]

To illustrate these measures of fit, Table 9 presents the results for the factor model of the Revised Causal Dimension Scale that was tested above. As can be seen, the values for the various fit indices that range from 0 to 1 (e.g., TLI, IFI, CFI, MFI) were all greater than .95 and were nearly identical in value. The SRMSR (average squared residual correlation) for the model was .04. Finally, the value of RMSEA was also .04, which is below the criterion of .06 recommended by Hu and Bentler (1999).

It should be noted that most of the studies (79%) reporting confirmatory factor analyses in *PSPB* during the years 1998 and 2000 reported indicators of model fit beyond the chi-square test of significance. As can be seen in Table 10, these researchers used some of the fit indices recommended by Hu and Bentler (1998, 1999). However, investigators also report other indices (such as the normed fit index and the goodness-of-fit index) that were not recommended by Hu and Bentler. Furthermore, none of the researchers reported the SRMSR or used .95 as the criterion for model fit. Clearly, the results reported by Hu and Bentler (1999) indicate the need to use more stringent criteria in evaluating model fit.

Sample Size

Another issue that arises in conducting a confirmatory factor analysis involves sample size. The statistical theory underlying these methods is based on the assumption that the data are drawn from large samples of the population; how large is large enough? Confirmatory factor analyses reported in *PSPB* during the years 1998 and 2000 used from 51 to 547 cases in the analyses

TABLE 9: Measures of Fit for the Two-Factor Model of the Causal Dimension Scale

Measure of Fit	*Value*
Standardized Root Mean Square Residual (SRMSR)	.04
Tucker-Lewis Index (TLI)[a]	.98
Bollen 89 Index (IFI)	.99
Comparative Fit Index (CFI)	.99
McDonald (MFI)	.99
Root Mean Square Error of Approximation (RMSEA)	.04

NOTE: The other two measures of fit recommended by Hu and Bentler (1998) are not reported by either LISREL 8.3 or EQS 5.7.
a. This measure of fit is referred to as the Non-Normed Fit Index (NNFI) in LISREL 8.3 or the Bentler-Bonett Nonnormed Fit Index in EQS 5.7.

TABLE 10: Goodness-of-Fit Indices Reported in *Personality and Social Psychology Bulletin* Articles

Index	*No.*	*%*
Tucker-Lewis Index	2	10.5
Normed Fit Index	3	15.8
Goodness-of-Fit Index	5	26.3
Adjusted Goodness-of-Fit Index	4	21.1
Comparative Fit Index	14	73.7
χ^2/df	6	31.6
Root Mean Square Error of Approximation (RMSEA)	6	31.6

(M = 241 participants). Three of the analyses (16%) involved samples of fewer than 100 cases, and 5 analyses (26%) involved from 100 to 199 cases.

Early research on the effect of sample size indicated that a minimum of 100 cases was required for accurate results (Boomsma, 1982), with samples of 200 or more cases being preferable. In the context of testing a "true" model (i.e., where the sample was drawn from a population in which the factor model was correct), the chi-square statistic was found to be inflated for samples of fewer than 100 cases. Subsequently, it was argued that the sample size should be evaluated in the context of the number of parameters (i.e., factor loadings, error terms) being estimated in the model, much like the N:k ratio in regression. For example, Bentler (1990) recommended that a sample of at least 5 cases per parameter be used in testing the model.

A recent study by Jackson (2001) directly examined the impact of sample size and number of parameters being estimated on the results of a confirmatory factor analysis. He conducted analyses for samples that ranged in size from 50 to 800 cases. The results of his Monte Carlo analysis clearly indicated that it was sample size and not the number of parameters being estimated that was important. Furthermore, the effect of sample size on different indicators of model fit varied. So, for example, the CFI was less affected by sample size than other measures of model fit. Also, average loading of the variables on the factors, termed "indicator reliability" or "saturation," moderated the effects of sample size on the results. That is, sample size was less important for model fit when the average factor loading was .80 versus .60.[11]

In conclusion, as was true in the case of exploratory factor analysis (see MacCallum et al., 1999), it is possible to derive accurate results from a confirmatory factor analysis using a sample of fewer than 100 cases if the average loading of the measures on the factors is high. Furthermore, with small samples it appears wise to examine an indicator of model fit such as the CFI that does not appear to be as strongly influenced by sample size as other indicators of model fit.

Non-Normality

An assumption that underlies maximum likelihood estimation of confirmatory factor models is that the distribution of the data is multivariate normal. As noted by Micceri (1989), very few actual data sets meet this assumption. The typical impact of violations of this assumption is to increase the value of the chi-square statistic and the standard errors associated with the parameter estimates. As noted by Bentler (personal communication, August 5, 2001), non-normality may result in a decrease in the chi-square statistic if the tails of the distribution are too small relative to a normal distribution. The parameter values themselves are typically not affected. To address this problem, Browne (1984) developed the asymptotically distribution free (ADF) estimation method, which is available in programs such as LISREL or EQS. Unfortunately, this method requires relatively large samples (more than 1,000 cases) to derive accurate estimations of model fit (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992; Muthén & Kaplan, 1992). Recently, Yuan and Bentler (1999) have developed an F statistic that is a modification of the ADF estimation developed by Browne (1984). As is true of the chi-square statistic for evaluating model fit, a nonsignificant F value indicates that the model fits the data. Analyses by Bentler and Yuan (1999) indicate that this F statistic performs well in the context of non-normality for samples as small as 90 cases. This new F statistic will be implemented in Version 6 of the EQS program. Using a preliminary or beta version of that program, the model for the Revised Causal Dimension Scale that was analyzed above was tested. As expected, the result of the Yuan-Bentler test was nonsignificant, $F(9, 371) = 1.39$, indicating that the model fit the data.

Recent work by Bentler and his colleagues also has focused on developing methods to adjust estimates derived under maximum likelihood estimation, termed "robust" estimation. Satorra and Bentler (1994) developed an adjustment for the extent of non-normality that can be applied to the chi-square statistics and standard errors derived under maximum likelihood estimation. These robust statistics are available in EQS 5.7. To illustrate these methods, they were applied to the data from the analysis of the Causal Dimension Scale presented earlier. The EQS program first provides a test of multivariate normality for the data. As expected, these data were found not to be multivariate normal, Mardia's coefficient = 10.67, $Z = 10.62$, $p < .001$. Due to the non-normality of the data, the Satorra-Bentler rescaled chi-square was 12.98, which is a reduction from the normal-theory-based value of 14.43. Finally, the standard errors of the parameters also were adjusted for non-normality. For example, the standard error for the loading of the first stability item was reduced from .155 to .144, which in turn increased the statistical significance of the loading.

Subsequent research has indicated that this adjustment to the chi-square statistic and the standard errors of the parameters performs well for moderately large samples (i.e., 250 or more cases) (Bentler & Yuan, 1999; Fouladi, 2000). However, for smaller sample sizes the adjustment appears to be too liberal. Therefore, in the case of smaller sample sizes investigators should employ the methods developed by Bentler and Yuan (1999) that will soon be available in EQS 6 to deal with non-normality.

Missing Data

An important area of recent developments related to confirmatory factor analysis (and structural equation modeling more generally) has involved the problem of missing data. Despite the fact that nearly all research projects involve missing data on one or more of the measured variables, only one of the 85 *PSPB* articles reviewed for this report made any reference at all to the treatment of missing data. It seems likely that list-wise deletion of cases with missing data was used in these factor analyses, in which the analysis was confined to individuals with complete data on the measures being factored. Except for the case where missingness is completely at random, such an approach to missing data can lead to serious biases in the results (see discussion by Allison, 2002; Schafer, 1997; Sinharay, Stern, & Russell, 2001).

Current versions of programs such as Amos (Arbuckle, 1997) or Mplus (Muthén & Muthén, 1998) include options for estimating confirmatory factor models in the context of missing data building on the initial work by Allison (1987) and Muthén, Kaplan, and Hollis (1987). Using a method that has been termed full-information maximum likelihood (FIML) estimation, these procedures evaluate the fit of the factor model while including data for cases with complete and partial data on the measured variables (for a review, see Allison, 2002; Enders, 2001). Monte Carlo studies have indicated that this procedure provides a more accurate estimate of model parameters than methods such as the exclusion of cases with missing data when missingness is related to one or more of the variables included in the analysis (see, e.g., Enders & Bandalos, 2001).

Jamshidian and Bentler (1999) have developed an alternative approach to the estimation of confirmatory factor models in the context of missing data. They compute estimates of model parameters (e.g., factor loadings) that optimize the fit of the model by making use of the EM (expectation-maximization) algorithm that is employed in the context of multiple imputation of missing data (see Allison, 2002; Schafer, 1997; Sinharay et al., in press). A recent Monte Carlo analysis by Gold and Bentler (2000) has indicated that this approach provides a more accurate estimate of model parameters than restricting the analysis to cases with complete data. Furthermore, Yuan and Bentler (2000) have demonstrated how this method can be used to deal with non-normal missing data. These methods for dealing with missing data also will be included in Version 6 of the EQS program (Bentler, 1995).[12]

Recommendations

A number of recommendations concerning the use of confirmatory factor analysis can be formulated based on both my review of studies in *PSPB* and recent methodological work. These recommendations are detailed below.

Study design. As was true of exploratory factor analysis, current literature on sample size and confirmatory factor analysis indicates that samples of 100 cases or more should be used. Use of samples smaller than 100 cases should be justified on the basis of the reliability or saturation of the measures, as indicated by loadings of the measures on the factors. The number of parameters being estimated relative to the sample size does not appear to be important, although investigators should strive to have at least three measures per factor or latent variable.

Estimation method. Use of maximum likelihood estimation of factor models appears to be justified. However, investigators should examine their data for violations of the assumption of multivariate normality. Given that it appears likely that most data sets will prove not to be multivariate normal, using some of the techniques to adjust for non-normality developed by Bentler and his colleagues appears wise, given that the apparent fit of the model to the data may be negatively affected by non-normality. The new methods developed by Yuan and Bentler (1999) are currently only available in the version of the EQS program that will soon be released. However, past experience has indicated that innovations introduced in one program for conducting confirmatory factor analyses or structural equation modeling are soon added to other programs. Therefore, we can anticipate that these new methods will likely be added to other programs in the near future.

Evaluating model fit. Recent studies by Hu and Bentler (1998, 1999) suggest an important new strategy for evaluating model fit, involving the use of two fit criteria (i.e., the average standardized residual and another criterion such as the CFI or the RMSEA). Although these results are contradicted to some degree by the findings reported by Fan, Thompson, and Wang (1999; see Note 9), their results are already having an impact on the criteria used in evaluating model fit. Their recommendation that the various fit criteria, such as the CFI, be greater than .95 has clearly raised the bar, making it more difficult to obtain results that appear to fit the data.

One result of these more stringent criteria will be that factors or latent variables will need to be more clearly defined by the measured variables, with smaller residual correlations between measures that load on different factors or latent variables. A recent article by Kenny and McCoach (2000) is relevant in this regard. They demonstrate mathematically that the number of variables being analyzed negatively affects many of the goodness-of-fit indices. As a consequence, one may need to reduce the number of items on a scale that is being factored to obtain results that indicate that the proposed model fits the data well. Therefore, we may see the evolution of scales that involve a small set of items that are very tightly intercorrelated to obtain fit indices greater than .95.

Missing data. Finally, the recent development of methods to estimate confirmatory factor models in the context of missing data using programs such as Amos, Mplus, and EQS 6 argues against excluding cases with missing data in estimating model fit. Although these methods greatly enhance our ability to estimate models in the context of missing data, it should be noted that these methods cannot be used with certain types of models (e.g., dichotomous measures). Investigators can employ other methods, such as multiple imputation of missing data, to address model estimation in the context of missing data with such nonstandard models (see Allison, 2002; Sinharay et al., 2001).

CONCLUSIONS

My analysis of the use of factor analysis, both exploratory and confirmatory, in *PSPB* over the past 5 years has revealed a number of ways in which such analyses can be improved. There is also clear evidence that the use of confirmatory factor analysis is increasing, a trend that I expect will continue to accelerate. Indeed, in many cases, researchers who have used exploratory factor analysis to analyze their data have clear predictions regarding the factor structure of the measures. For example, of the 137 exploratory factor analyses reported in *PSPB* during the years 1996, 1998, and 2000, 54 (39%) involved analyses where investigators had clear predictions regarding the factor structure of the measures being analyzed. As a consequence, conducting a direct test of whether the proposed factor model fits data via confirmatory factor analysis is more appropriate than conducting an exploratory factor analysis.

The use of confirmatory factor analysis in the context of testing a hypothesized factor structure for a measure or set of scales also will permit investigators to address issues that cannot be easily addressed via exploratory factor analysis. One can directly compare the fit of a hypothesized factor structure for different groups of participants in a study. So, for example, models can be tested that evaluate whether the factor loadings on a measure vary by sex of the respondent. Such multiple group models can be extended to conduct an analysis of "structured means," wherein the average scores of different groups of participants (such as men and women or individuals assigned to different treatment conditions) on the latent variables or factors are compared. Such an analysis is essentially equivalent to conducting a *t* test or an ANOVA on error-free measures of a dependent variable or variables. Descriptions of how to conduct such multiple group confirmatory factor analyses are provided in the various textbooks on structural equation modeling, such as the books by Byrne (1994, 1998, 2001), Kelloway (1998), Kline (1998), Maruyama (1997), Raykov and Marcoulides (2000), or Schumacker and Lomax (1996). Finally, it is also possible to conduct a confirmatory second-order factor analysis, wherein one tests the ability of a higher order factor or factors to account for the correlation between first-order factors. An example of such an analysis is provided by our work with the Social Provisions Scale, a multidimensional measure of social support (Cutrona & Russell, 1987; see Marsh & Hocever, 1985, for a discussion of how to test such models).

Methods for conducting confirmatory factor analysis or structural equation modeling are evolving rapidly. One indication of the interest in these methods is the development of the journal *Structural Equation Modeling*, along with the large number of articles relevant to these methods that are published in *Psychological Methods*. Investigators who use these methods need to keep up to date on their evolution, given the rapid pace of change. Clearly, our ability to deal with problems such as non-normality of the data, small sample sizes, or missing data has been enhanced by recent developments, which in turn should increase the utility of these methods in addressing the empirical issues that are of interest to investigators who publish in *PSPB*. I therefore anticipate that the trend of increasing use of confirmatory factor analysis in articles published in *PSPB* and other personality and social psychology journals will continue, which in turn should enhance the quality of the research in these journals.

NOTES

1. To evaluate my accuracy in detecting articles in *PSPB* over these 3 years that included a factor analysis, an independent review was conducted by a graduate student for the months January through April 1996, May through August 1998, and September through December 2000. For the 115 articles published in *Personality and Social Psychology Bulletin* during those months, agreement between the two raters was 90% and the kappa coefficient was .72 ($p < .001$). I would like to thank Young Kim for helping with the reliability coding.

2. Technically speaking, principal components analysis does not represent a factor analysis, given that the method does not involve the assumption that you are extracting common factors underlying a set of measures (see discussion by Fabrigar, Wegener, MacCallum, & Strahan, 1999; Velicer & Fava, 1998). Instead, the goal of the analysis is the exact mathematical transformation of a set of measures into a smaller set of measures or components.

3. You can also input your own communality estimates using Proc Factor in SAS.

4. The expected correlation between two measures based on the loadings of these measures on Factor 1 is calculated by multiplying the two loadings together. So, for example, if measure 1 loads .45 on Factor 1 and measure 2 loads .50 on Factor 1, then the expected correlation based on the influence of that factor on each measure is .45*.50, or .225.

5. Occasionally one will hear the expression that "the reliable factors were extracted." This refers to the relationship between coefficient alpha and the eigenvalue associated with the components extracted as part of a principal components analysis. Specifically, the relationship is as follows: $\alpha = (\varepsilon - 1) / \varepsilon$, where ε represents the eigenvalue for a specific component. Based on this formula, the eigenvalue must be greater than 1 for the reliability of the corresponding factor to be positive. Therefore, another justification for the "eigenvalue ≥ 1" criterion is that components with eigenvalues less than 1 are unreliable.

6. As indicated in the article by Reise, Waller, and Comrey (2000), the program for computing the random eigenvalues can be downloaded from the Web site http://lib.stat.cmu.edu/R/CRAN/#source. The program that I used to generate the average eigenvalues that are plotted in Figure 1 is the following: random.eig <– matrix(0,nrow = 100, ncol = 20) for (i in 1:100) {random.data <– matrix(rnorm(487 * 20), nrow = 487, ncol = 20) random.eig[i,] <– eigen(cor(random.data)) $values} average.eig <– apply(random.eig,2,mean). In this code, "487" represents the number of cases in the analysis and "20" represents the number of measures being factored. Please note that the program is case sensitive; do not use capital letters. Once this program runs, entering "average.eig" will lead to the results being listed.

7. Two other approaches to identifying the number of factors to extract are often discussed (Fabrigar et al., 1999). Velicer (1976), in the context of a principal components analysis, proposed an analysis of the average partial correlations computed after removing the influence of the extracted components from the original correlation matrix; the extraction of components would stop when this average reached a minimum. This statistic appears to perform well in identifying the correct number of components but it is limited to that form of extraction. Another approach is to examine the chi-square statistic computed in the context of an exploratory maximum likelihood analysis, continuing to extract factors until the chi-square statistic is nonsignificant. However, this test is negatively affected by the size of the sample employed in the analysis, such that too many factors are likely to be extracted with large samples.

8. One could, of course, first standardize scores on a measure prior to computing a score to represent a factor. Such a procedure makes it easy to reverse measures or items with negative weights on a factor; you simply multiply their z score by –1. In general, you do not need to go through a standardization procedure unless the measures involved have different variances from one another. So, for example, if you have factored items from a scale, then one can anticipate that the variances of scores on the individual items will be very similar. However, if you have factored scores from scales that vary in their variances, then standardizing the scores prior to combining them together will ensure equal weighting of the measures.

9. It should be noted that the statistical theory underlying these methods assumes that one is testing the fit of the factor model to a covariance matrix; indeed, an early label for these methods was "analysis of covariance structures." Although it is the case that one will often get the same results when analyzing a correlation matrix, these methods should be applied to a covariance matrix. Bentler and Lee (1983) and Cudeck (1989) discuss using these procedures to analyze correlation matrices.

10. In an article appearing in the same issue of *Structural Equation Modeling* as Hu and Bentler (1999), Fan, Thompson, and Wang (1999) also examined the sensitivity of different indicators of model fit to misspecification of the model, coming to different conclusions. Consistent with Hu and Bentler (1999), they found that root mean square error of approximation (RMSEA) and McDonald's Centrality Index (MFI) performed well. However, they also found that the goodness-of-fit index performed well although the results were affected by sample size. Other fit indices, such as the Tucker-Lewis Index and the Comparative Fit Index, did not perform well. Finally, Fan et al. (1999) did not evaluate the Standardized Root Mean Square Residual (SRMSR). These differences in results may, of course, be an artifact of the models that each group of researchers examined. Future studies need to examine the models both groups of researchers used for the same set of fit measures, in an effort to ascertain why their conclusions were different.

11. As noted by Bentler (personal communication, August 5, 2001), one must be cautious in drawing inferences about the effects of variations in number of parameters on model fit from the results of Jackson (2001) because he did not investigate models that varied widely in the number of parameters being estimated.

12. This work on missing data in the context of confirmatory factor analysis raises a question concerning the treatment of missing data in the context of an exploratory factor analysis. Exploratory factor programs available from SPSS or SAS are not designed to permit the estimation of such models using the full-information maximum likelihood (FIML) or expectation-maximization (EM) algorithms. However, one can estimate an exploratory factor model using the FIML algorithm using the Mplus program. Also, one can employ multiple imputation in the context of testing an exploratory factor model. The use of multiple imputation is described by Sinharay et al. (2001) using the NORM program developed by Schafer (1997) or by Allison (2002) using Proc MI that is now available from SAS.

REFERENCES

Allison, P. D. (1987). Estimation of linear models with incomplete data. In C. C. Clogg (Ed.), *Sociological methodology 1987* (pp. 71-103). San Francisco: Jossey-Bass.

Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Arbuckle, J. L. (1997). *Amos user's guide version 3.6*. Chicago: SmallWaters.

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

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.

Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, *25*, 67-74.

Bentler, P. M., & Lee, S. L. (1983). Covariance structures under polynomial constraints: Applications to correlation and alpha-type structural models. *Journal of Educational Statistics*, *8*, 207-222.

Bentler, P. M., & Yuan, K. -H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, *34*, 181-197.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Research and Methods*, *16*, 492-503.

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (Part I, pp. 149-175). Amsterdam: North-Holland.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62-83.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Thousand Oaks, CA: Sage.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. New York: Lawrence Erlbaum.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: Lawrence Erlbaum.

Cattell, R. B. (1966). *Handbook of multivariate experimental psychology*. Chicago: Rand McNally.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, *105*, 317-327.

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.

Cutrona, C. E., & Russell, D. (1987). The provisions of social relationships and adaptation to stress. In W. H. Jones & D. Perlman (Eds.), *Advances in personal relationships* (Vol. 1, pp. 37-67). Greenwich, CT: JAI.

Enders, C. K. (2001). A primer on maximum likelihood algorithm available for use with missing data. *Structural Equation Modeling*, *8*, 128-141.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430-457.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272-299.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, *6*, 56-83.

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.

Gold, M. S., & Bentler, P. M. (2000). Treatments of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling*, *7*, 319-355.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Gorsuch, R. L. (1988). Exploratory factor analysis. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 231-258). New York: Plenum.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424-453.

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

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351-362.

Jackson, D. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A Monte Carlo investigation. *Structural Equation Modeling*, *8*, 205-223.

Jamshidian, M., & Bentler, P. M. (1999). ML estimation of mean and covariance structures with missing data using complete data routines. *Journal of Educational and Behavioral Statistics*, *24*, 21-41.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.

Kaiser, H. F. (1970). A second generation Little Jiffy. *Psychometrika*, *35*, 401-415.

Kelloway, E. K. (1998). *Using LISREL for structural equation modeling: A researcher's guide*. Thousand Oaks, CA: Sage.

Kenny, D. A., & McCoach, D. B. (2000). *Effect of the number of variables on measures of fit that use a null model*. Manuscript submitted for publication.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84-99.

Marsh, H. W., & Hocever, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First and higher order factor models and their invariance across groups. *Psychological Bulletin*, *97*, 562-582.

Maruyama, G. (1997). *Basics of structural equation modeling*. Thousand Oaks, CA: Sage.

McAuley, E., Duncan, T. E., & Russell, D. (1992). Measuring causal attributions: The Revised Causal Dimension Scale (CDSII). *Personality and Social Psychology Bulletin*, *18*, 566-573.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification*, *6*, 97-103.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, *107*, 247-255.

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

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

Muthén, B., Kaplan, K., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *42*, 431-462.

Muthén, L. K., & Muthén, B. (1998). *Mplus user's guide*. Los Angeles: Muthén & Muthén.

Ogasawara, H. (2000). Some relationships between factors and components. *Psychometrika*, *65*, 167-185.

Raykov, T., & Marcoulides, G. A. (2000). *A first course in structural equation modeling*. New York: Lawrence Erlbaum.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12*, 287-297.

Russell, D. (1996). The UCLA Loneliness Scale (Version 3): Reliability, validity, and factor structure. *Journal of Personality Assessment*, *66*, 20-40.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Schneeweiss, H. (1997). Factors and principal components in the near spherical case. *Multivariate Behavioral Research*, *32*, 375-401.

Schneeweiss, H., & Mathes, H. (1995). Factor analysis and principal components. *Journal of Multivariate Analysis*, *55*, 105-124.

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. New York: Lawrence Erlbaum.

Sinharay, S., Stern, H. S., & Russell, D. W. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, *6*, 317-329.

Steiger, J. H. (1989). *EzPATH: A supplementary module for SYSTAT and SYGRAPH*. Evanston, NJ: SYSTAT.
Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1-10.
Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321-327.
Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*, 231-251.
Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, *25*, 1-28.
Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, *28*, 263-311.
Yuan, K. -H., & Bentler, P. M. (1999). *F* tests for mean and covariance structure analysis. *Journal of Educational and Behavioral Statistics*, *24*, 225-243.
Yuan, K. -H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In *Sociological Methodology 2000* (pp. 165-200). Washington, DC: American Sociological Association.

Received October 6, 2001
Revision accepted March 4, 2002