

An Applied Comparison of Methods for Least-Squares Factor Analysis of Dichotomous Variables

Charles D. H. Parry
University of Pittsburgh

J. J. McArdle
University of Virginia

A statistical simulation was performed to compare four least-squares methods of factor analysis on datasets comprising dichotomous variables. Input matrices were: (1) phi correlation coefficients between the observed variables, (2) tetrachoric correlations estimated from bivariate tables of the observed variables, (3) tetrachoric correlations estimated on the basis of the latent continuous normal response variables underlying the observed variables (using LISCOMP with a weighted least-squares factor extraction), or (4) correlations between the latent response variables underlying the observed variables based on a variant of latent trait theory (using NOHARM). The simulations were studied under varying sample sizes, threshold values, and population loadings of a factor model. Factor extraction was performed, and a measure of deviation between the population and estimated factor loadings was used as an index of fit. The more sophisticated and less readily available third and fourth methods were not found to be markedly superior to the first two methods, even for highly skewed data with small sample sizes. Further simulations were performed to demonstrate the stability of the results. *Index terms: binary factor analysis, LISCOMP, NOHARM, simulation.*

Since the development of factor analysis in the early 1900s, there has been extensive debate over its applicability to categorical data. This issue has been of special relevance to researchers in the social and behavioral sciences, where observed variables frequently are dichotomous or have a small number of categories with unequal intervals.

APPLIED PSYCHOLOGICAL MEASUREMENT

Vol. 15, No. 1, March 1991, pp. 35-46

© Copyright 1991 Applied Psychological Measurement Inc.
0146-6216/91/010035-12\$1.85

An excellent review of this literature is given in Mislevy (1986).

In the case of dichotomous data, in particular, there has been considerable controversy surrounding the appropriateness of using factor analytic techniques. Some believe that assigning numbers to a dichotomy does not really require measurement assumptions, and that it is therefore legitimate to apply the usual multiple common factor analysis procedures to matrices of either phi coefficients (the Pearson formula applied to dichotomous variables) or dichotomous variables scored 0 or 1 (Burt, 1950). Others have rejected this position, and have argued that the measurement problem of dichotomous variables requires that other approaches be employed (Alexander, Alliger, Carson, & Barrett, 1985; Carroll, 1961; Kim & Mueller, 1978).

A major concern about traditional factor analysis using phi coefficients has been raised when responses are highly skewed. In binary variables, the mean is identical to the proportion correct and it defines the variance. Thus, a distorted analysis can result when variables are highly skewed (McDonald, 1969; Muthén, 1983). In the limiting case of two dichotomous variables having a perfect Guttman ordering, the phi correlation coefficient will depend solely on the means of the two variables, and will reach a value of 1 only when both variables have equal means (Mislevy, 1986). Traditional factor loadings calculated from these summary statistics will also be limited.

The argument was made in the 1940s that factor analysis of variables having a limited number of categories would yield misleading artifacts—termed “difficulty factors”—and that common factor analysis was not applicable to dichotomous data (McDonald, 1985). This stimulated a search for the selection of “the correct” coefficient of association between binary variables, and has ultimately led to the common practice of estimating tetrachoric correlations on the basis of bivariate tables in place of phi coefficients when performing factor analysis on datasets comprising dichotomous variables (Carroll, 1961; Christoffersson, 1975).

The notion that “difficulty factors” confound factor analysis of dichotomous data has recently been refuted (McDonald, 1985; McDonald & Ahlwat, 1974), and this has been accompanied by a recognition that there are in general no necessary distinctions between *qualitative* and *quantitative* variables. This has resulted in the acceptance that matrices of phi coefficients or the covariances of dichotomous items scored 0 or 1 may be factored (in the usual sense of multiple common factor analysis), provided that the assumption of linearity of regression between the observed dichotomous variables and latent factors/traits is met (McDonald, 1969).

In practice, however, the assumption of linearity may not be met (McDonald, 1969). One reason for this is that the value of a dichotomous variable is bounded by 0 and 1, even though the linear function is not so bounded (McDonald, 1985; Mislevy, 1986). In order to deal with this situation, a number of approaches have been developed based on the assumption that dichotomous variables are indicators of underlying continuous variables for which factor analysis is appropriate (Christoffersson, 1975; McDonald, 1969; McDonald, 1981; Muthén, 1979; Muthén & Christoffersson, 1981). These newer methods seek to determine the factor structure after first estimating the correlations among the latent continuous multivariate normal response variables underlying the observed dichotomous variables.

McDonald’s least-squares item factor analy-

sis procedure (NOHARM; Fraser & McDonald, 1988; McDonald, 1967) and Muthén’s latent variable structural equations modeling procedure (LISCOMP; Christoffersson, 1975; Muthén, 1987) are two prominent approaches to implementing latent variable factor analysis. McDonald (1982) outlined the equivalence between the general case of latent trait theory and item response theory, and developed software that would allow researchers to use this theory in solving problems requiring the factor analysis of categorical data (Fraser & McDonald, 1988). For a detailed description of the theory behind the NOHARM program, see McDonald (1981, 1982, 1985). Muthén’s (1987) LISCOMP program permits the factor analysis of polychotomous data, based on theory outlined earlier (Muthén, 1983, 1984).

Choices in Factor Analyzing Dichotomous Variables

An applied researcher must make two choices in implementing a factor analysis of dichotomous variables: the coefficient of association, and the method of factor extraction. These choices result in at least four potentially different methods of least-squares factor analysis:

1. ULS-PHI: defined as (1) calculation of a matrix of phi coefficients, followed by (2) fitting of a factor model using a minimum unweighted least-squares (ULS) procedure.
2. ULS-TC: defined as (1) calculation of a matrix of tetrachoric correlations, followed by (2) fitting of a factor model using a minimum ULS procedure.
3. LISCOMP: defined as (1) calculation of a matrix of tetrachoric correlations, followed by (2) fitting of a factor model based on a *weighted* least-squares (WLS) factor extraction. The estimates are obtained using the estimated covariance matrix of the tetrachorics as the weight matrix (see Muthén, 1987, p. 6-4).
4. NOHARM: defined as (1) calculation of a cross-products matrix using phi coefficients plus means, followed by (2) fitting of a factor model using an approximation to a ULS

procedure for binary variables.

These methods above are outlined schematically in Figure 1. NOHARM and ULS-PHI both use a matrix of product-moment correlations as input to the factor extraction, whereas LISCOMP—in the case of dichotomous variables—and ULS-TC use a matrix of tetrachoric correlations. However, ULS-PHI and ULS-TC derive the input matrices directly from the observed variables, whereas NOHARM and LISCOMP use a more indirect approach.

Analysis of a Small Data Matrix

Little has appeared in the literature about the benefits of using the more sophisticated procedures (e.g., LISCOMP, NOHARM) over the more traditional methods of factor analysis for dichotomous data (e.g., ULS-PHI, ULS-TC). A rudimentary comparison of these four methods of factor analysis on a binary dataset appears in Table 1. The data were taken from Bock and Lieberman (1970); these data were also analyzed by Christoffersson (1975). The data consisted of items 11 through 15 of Section 6 of the Law School Admission Test. A single common factor

model with a least-squares fitting function was fit using four different techniques for estimation. Least-squares factor analysis solutions were sought for all methods to ensure uniformity among methods, because NOHARM is based on a least-squares factor extraction. In addition, least-squares is the simplest of the standard fitting functions. Table 1 shows the different correlation matrix used by each method, and the different factor matrices that resulted. The solutions are similar but not identical in value or rank order, and these differences needed to be explored in further detail.

Purpose

The objective of this research was to provide a more systematic comparison of these four selected methods of factor analysis of binary data. Additional factors that might influence the efficacy of the various methods were taken into account in comparing these methods in a small set of applied statistical simulations designed to compare the methods. These factors included sample size, population loadings, and threshold values. An overview study examined a wide vari-

Figure 1
 An Applied Decision Chart for Dichotomous Variable Factor Analysis

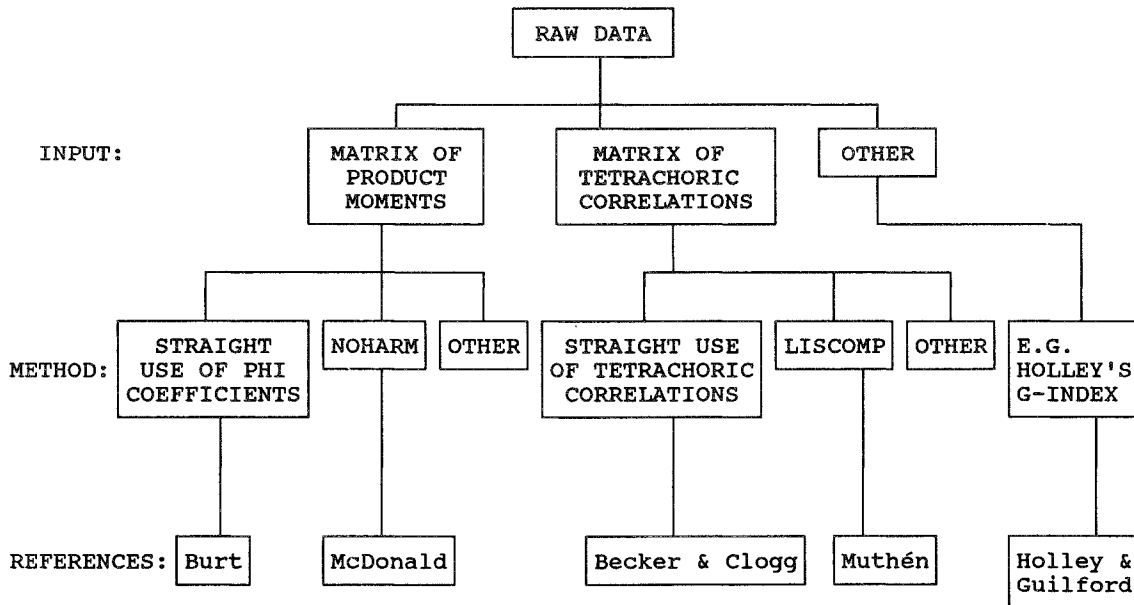


Table 1
 Five-Variable Problem From Bock and Lieberman (1970)

Matrix Analyzed					Factor Loadings	Rank
ULS-PHI (phi coefficients)						
1.0000					.215	5
.0738	1.0000				.306	2
.0989	.1148	1.0000			.369	1
.0443	.0623	.1091	1.0000		.279	3
.0238	.0862	.0532	.0992	1.0000	.223	4
ULS-TC (unweighted tetrachorics)						
1.0000					.374	3
.1700	1.0000				.409	2
.2280	.1890	1.0000			.482	1
.1070	.1110	.1870	1.0000		.373	4
.0670	.1720	.1050	.2010	1.0000	.326	5
LISCOMP (weighted tetrachorics)						
1.0000					.547	1
.1700	1.0000				.518	2
.2280	.1890	1.0000			.492	3
.1070	.1110	.1870	1.0000		.456	4
.0670	.1720	.1050	.2010	1.0000	.434	5
NOHARM (phi coefficients plus means)						
.9240					.404	2
.6640	.7090				.387	3
.5240	.4180	.5530			.475	1
.7100	.5530	.4450	.7630		.373	4
.8060	.6300	.4900	.6780	.8700	.344	5

ety of parameters without replication. An intensive replication then examined several model conditions of most interest.

Method

Overview Study

Datasets. Each dataset consisted of eight dichotomous variables, and was constructed to have one common factor and a predetermined factor structure. Software developed by SAS (SAS, 1985) was used to simulate single datasets of known sample size ($N = 50, 100, 200$), population factor loadings (.45, .70, .90), and threshold values yielding 50/50, 60/40, 70/30, 80/20, 90/10 splits for the dichotomous variables. These values were selected to be representative of real applied situations. Forty-five datasets were created, one for each of the various combinations of sample size, population loadings, and threshold values.

In order to create the desired dichotomies in the variables, z values were used as threshold values to create the splits at the 50th, 60th, 70th, 80th and 90th percentiles.

Factor analysis methods. Each of the 45 datasets was analyzed using the following four least-squares methods for factor extraction:

1. ULS-PHI. Phis between the eight observed dichotomous variables were estimated using the CORR procedure in SAS, with the product-moment calculation producing phi coefficients, and ULS factor extraction was performed on the resulting correlation matrix using the FACTOR procedure (SAS, 1985).
2. ULS-TC. Tetrachoric correlation coefficients were estimated using Muthén's (1987) LISCOMP program. These coefficients are calculated using the computational procedures outlined by Divgi (1979) and Kirk (1973).

ULS factor extraction was performed on the resulting correlation matrix using the FACTOR procedure (SAS, 1985). The "HEYWOOD" option was used in the SAS FACTOR procedure to set to 1 any communality that was greater than 1, in order to allow the iterations to proceed.

3. LISCOMP. Muthén's (1987) LISCOMP procedure for structural equations modeling of polychotomous variables uses tetrachoric correlations, which are estimated on the basis of the latent continuous normal response variables in cases of dichotomous observed variables. LISCOMP uses a relatively complex statistical approximation to account for the sampling characteristics of the data. When WLS factor extraction is performed, the matrix of pairwise tetrachoric correlations is weighted by the matrix of standard errors of the tetrachorics. The WLS option was selected because it was felt that, at least in the case of "larger" samples, the results should be less biased and much more efficient than the other methods. Using the ULS option in LISCOMP yields the same results as ULS-TC; therefore, the findings reported for ULS-TC hold for LISCOMP when ULS estimation is performed.
4. NOHARM. McDonald's NOHARM procedure (Fraser & McDonald, 1988) assesses the correlations between the latent response variables underlying the observed variables, and it performs factor extraction using an ULS approach. The input to this program was a lower triangular matrix of average cross-products. These elements may be interpreted as mean products plus phi coefficients derived from the dichotomous variables. Thus, NOHARM accounts for the statistical distribution of the data within the error theory of the model. NOHARM uses a nonlinear calculation of the matrix of residuals that is based on a simple or harmonic approximation to the normal ogive function. Thus, both first- and second-order statistical information is extracted from the data and used in the

model calculations (see Mislevy, 1986).

ULS-PHI and ULS-TC differed only in the matrices input into the factor extraction. LISCOMP used the identical matrix as ULS-TC, but differed from ULS-TC (and the other methods) in the method of factor extraction. NOHARM differed from the other methods in both the method of estimation of the input matrix and in the method of factor extraction.

Discrepancy index. Because ULS factor extraction procedures do not provide standard errors of estimation or statistical tests of fit, an index of fit (deviation) between the population and the observed loadings was created as:

$$\text{Dev}_j = \sum_{m=1}^M \frac{(\hat{L}_m^2 - L_m^2)^2}{L_m^2}, \quad (1)$$

and

$$D_j = \ln(\text{Dev}_j), \quad (2)$$

where for the j th model fitted,

\hat{L}_m = the estimated (or observed) factor loading on the m th variable ($m = 1, \dots, M$),

L_m = the population (or expected) factor loading on the m th variable, and

\ln = the natural logarithm.

This index is similar to the χ^2 test statistic and to a measure of the accuracy of estimated factor matrices outlined by Browne (1968, pp. 288-289).

Low deviation scores represent a close fit between the observed and true population loadings. Equation 1 is scaled by the expected value, so it reflects a proportional rather than an absolute difference. Another measure of fit was considered—the root mean square error of off-diagonal elements of the residual matrix—which provides a measure of whether the model fits the data or whether it would be better to try a higher-order factor model. In most cases, this is different from assessing whether the model accurately reproduces population loadings. For LISCOMP, ULS-TC and ULS-PHI, the input to the factor extraction matrix was the matrix of correlations between the variables, whereas for NOHARM it was a matrix of cross-products. For this

reason the root mean square error of off-diagonal residuals is not comparable across methods.

Intensive Replication

The first study employed a single replication per cell because there was only one randomly selected dataset for each combination of method, sample size, loading, and threshold. This allowed for the possibility of the findings being biased or inefficient. In order to test the stability of the results reported in the overview study, analyses in selected cells were repeated 30 times to generate a sampling distribution for these cells.

The selected cells reflect all combinations of methods (ULS-PHI, ULS-TC, NOHARM), excluding LISCOMP (see explanation below) for low and high population loadings (.45 versus .9), and skewed and non-skewed data (50/50 versus 90/10 splits). The sample size was 100. The matrix calculations and factor extractions used were identical to those of the overview study.

Results

Overview Study

Of the 180 simulations (45 datasets × 4 methods), 21 of the 45 for LISCOMP (using the WLS method of factor extraction) could not be

completed because the weight matrix was not positive definite. This computational problem decreased as sample size increased. Therefore, the results are reported only for the comparison between ULS-PHI, NOHARM, and ULS-TC. Because the ULS option in LISCOMP yields the same results as ULS-TC, the findings reported for ULS-TC hold for LISCOMP when ULS estimation is performed.

A summary of the findings from the remaining 135 simulations is presented in Table 2. The simulations were compared in order to assess how the methods functioned under situations of varying sample size, population loadings, and threshold values. A four-way analysis of variance (Method × Sample × Loading × Threshold) was performed on the natural logarithm of the *D* statistic. The log transformation was used in order to obtain a more symmetric distribution of these scores prior to performing the ANOVA. Also, all effects were treated as fixed, because no attempt was made to randomly select methods, population loadings, sample sizes, or threshold values.

The strength of association, η^2 , [computed by $(SS_{effect}/SS_{total}) \times 100$] was calculated for each effect. Only two of the higher-order interactions

Table 2
 Sums of Squares for the 13 Effects (SS) and Strength of Association for the 13 Effects (η^2) From the Four-Way ANOVA With *D* as the Dependent Variable

Source	df	SS	η^2	η^2/df	<i>F</i>
Method	2	2.698	1.82	.91	37.6*
Loading	2	54.347	36.70	18.35	756.8*
Sample	2	19.397	13.10	6.55	270.1*
Threshold	4	25.801	17.42	4.36	179.6*
Method × Loading	4	26.946	18.19	4.55	187.6*
Method × Sample	4	1.656	1.12	.28	11.5*
Method × Threshold	8	1.045	.71	.09	3.6*
Loading × Sample	4	.809	.55	.14	5.6*
Loading × Threshold	8	2.850	1.92	.24	9.9*
Sample × Threshold	8	3.383	2.28	.29	11.8*
Method × Loading × Sample	8	.669	.45	.05	2.3
Method × Loading × Threshold	16	1.923	1.30	.08	3.4*
Method × Sample × Threshold	16	4.711	3.18	.20	8.2*
Loading × Sample × Threshold	16	.721	.49	.03	1.3
Residual	32	1.149			
Corrected Total	134	148.105	100.00		

involving Methods (the factor of central interest) were particularly noteworthy (η^2 significant at $p < .01$): Method \times Loading \times Threshold, and Method \times Sample (the Method \times Loading and Method \times Threshold interactions are included in the significant three-way Method \times Loading \times Threshold interaction).

The Method \times Loading \times Threshold interaction is illustrated in Figure 2, which contains plots of the Method \times Loading interaction for each of the five threshold values. Lower D values indicate a higher degree of recovery of the underlying structure than higher D values. When averaging over the different sample sizes, ULS-PHI consistently yielded lower D s than ULS-TC and NOHARM when the loading was low (.45) for all threshold values studied. There does not appear to be any major difference between the methods with population loadings of .7 for 50/50, 60/40, and 70/30 splits. Mean D was lowest for ULS-PHI and ULS-TC, with 80/20 and 90/10 splits, respectively. When the population loadings were .9, mean D was consistently lower for ULS-TC when compared with the other methods, for all threshold values. The difference in mean D between ULS-TC and LISCOMP (not shown in Figure 2) was not great at any of the threshold values studied. In contrast, the mean deviation scores were consistently higher for ULS-PHI for all threshold values.

As expected, a tendency was observed for datasets that are more highly skewed to generate higher D s than did datasets that were less skewed; there was also a tendency for datasets having a structure with higher loadings to generate lower mean D than datasets with lower population loadings. Figure 2 also reveals that, in contrast to the other methods, ULS-PHI was not as greatly affected by variation in the population loadings.

The Method \times Sample interaction is illustrated in Figure 3. When averaging over the different threshold and loading values, ULS-PHI and ULS-TC yielded lower mean D than NOHARM when the sample size was 50 and 200. For datasets with a sample size of 100, mean D was

substantially lower for ULS-PHI. Higher mean D for all sample sizes studied was noted for NOHARM. As expected, Figure 3 indicates a tendency for datasets with larger sample sizes to generate lower mean D than datasets with smaller sample sizes.

Intensive Replication

Means, standard deviations (SD), skewness, and kurtosis estimates for D for each of 12 selected cells are provided in Table 3. The last column of this table contains the estimated values from the overview study. Support for the stability of the findings reported in the overview study is indicated by the fact that each of these estimated values fell within 1 SD of the mean obtained from the repeated replications.

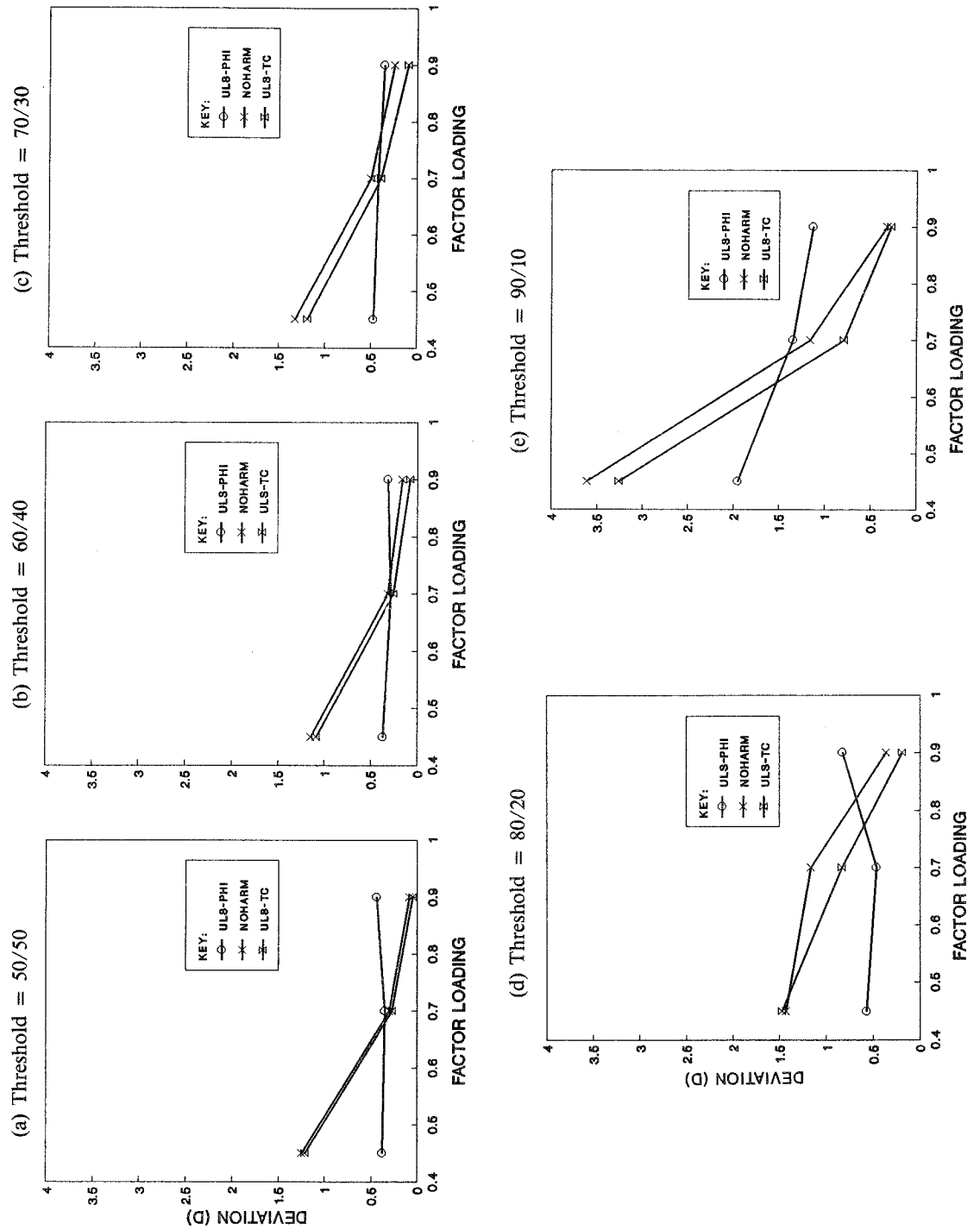
The Method \times Loading interaction is plotted in Figure 4 for the 50/50 and 90/10 threshold values (splits). Only the two extreme values of loading were studied (.45 and .9), with a sample size of 100. Comparison of the data in Figure 4 with Figures 2a and 2e reveals a virtually identical pattern at both threshold values, with the exception of the deviation scores for NOHARM and ULS-TC when the loading was .45 and the threshold value was 90/10.

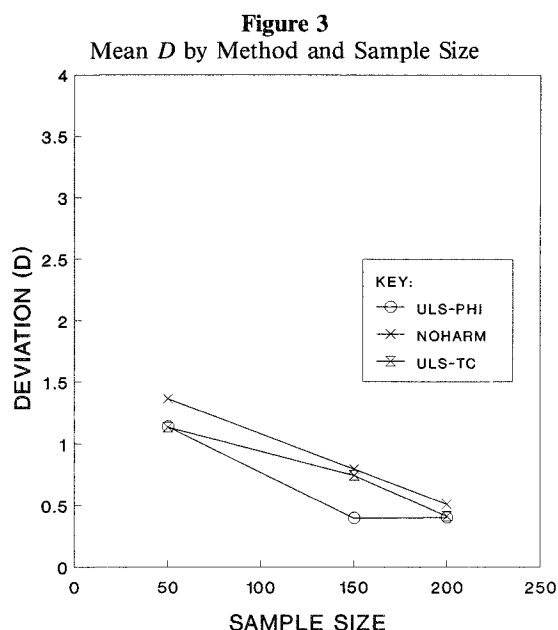
The unexpected finding in the overview study of a negligible loading effect for the ULS-PHI method was also studied further. Mean D s of .91 (SD = .90) and .70 (SD = .46) were obtained for loadings of .45 and .9, respectively, by averaging the 60 replications for the 50/50 and 90/10 splits. This yielded a 95% confidence interval around the mean of 0.0 to 2.67 for the .45 loading, and 0.0 to 1.60 for the .9 loading. Although mean D was lower for the higher loading, there was still considerable overlap between the confidence bands. This supports the finding of a minimal loading effect under the ULS-PHI method reported in the overview study.

Discussion and Conclusions

In comparing the four different methods for performing factor analysis on dichotomous data, it appears that the advantage of one method over

Figure 2
 Mena D by Method, Population Loading, and Threshold Values





another depends on the sample size, as well as on the combination of magnitude of the loading and the skewness of the data (threshold). There does not appear to be any single best approach to use in all circumstances.

Using the WLS method of factor extraction

with LISCOMP does not appear to work well for datasets having a small sample size, because the weight matrix is not always positive definite and this method often fails to converge. This approach cannot be used in common instances of small sample size (e.g., $N < 200$). Differences among the three remaining methods appear to be smallest when the data are not highly skewed (50/50 to 70/30 split) and when loadings are of moderate size (.7).

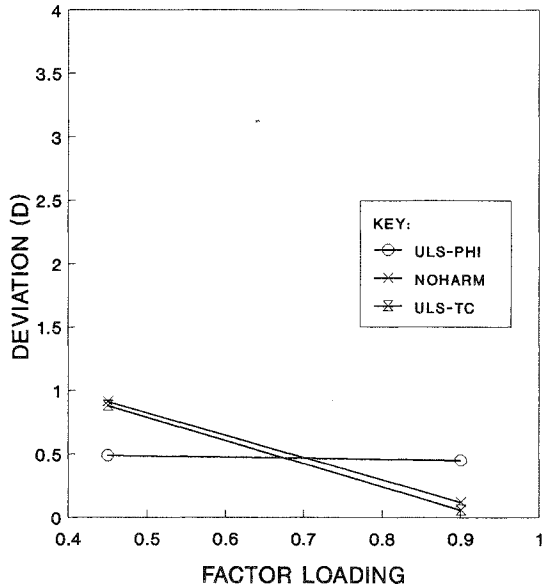
The estimates of population loadings using McDonald's NOHARM procedure (Fraser & McDonald, 1988) were not markedly superior (in terms of lower mean *D*) than those obtained from using ULS-PHI, except when population loadings were high (.9); usually, however, population loadings are unknown. NOHARM did not perform better than ULS-TC, even when the data were more highly skewed. This finding supports the commonly used strategy of estimating tetrachoric correlations on the basis of bivariate tables when performing factor analysis on datasets comprising dichotomous variables (Christofferson, 1975; Kim & Mueller, 1978; a variety of other techniques based on 2×2 tables are available, e.g., Holley & Guilford, 1964; Alexander et al., 1985).

The results from these simulations also sug-

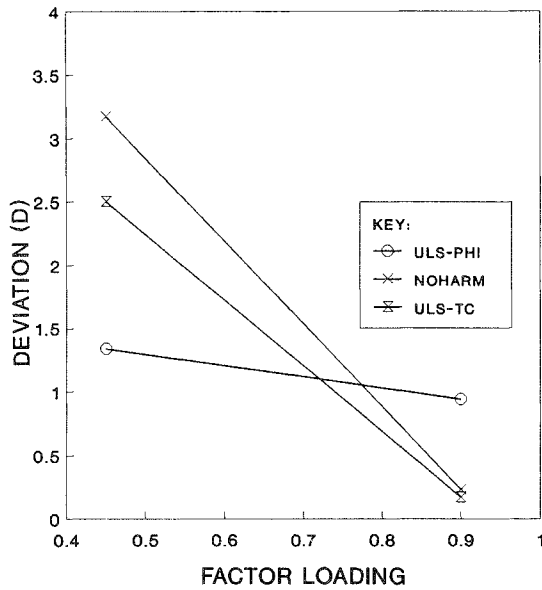
Table 3
 Sampling Distributions for 16 Selected Cells Based on 30 Replications
 Per Cell, and the Estimate from the Overview Study

Method and Loading	Split	Mean	SD	Skew	Kurtosis	Estimate
ULS-PHI						
.45	50/50	.49	.27	2.30	7.26	.32
.45	90/10	1.34	1.09	2.02	3.32	.91
.90	50/50	.45	.17	1.05	.47	.38
.90	90/10	.94	.53	1.22	1.01	.45
ULS-TC						
.45	50/50	.88	.69	1.99	4.19	1.41
.45	90/10	2.50	1.50	1.70	4.70	2.69
.90	50/50	.06	.04	2.46	9.17	.04
.90	90/10	.17	.12	2.01	3.62	.13
NOHARM						
.45	50/50	.91	.72	1.89	3.48	1.45
.45	90/10	3.17	1.14	.75	2.42	2.36
.90	50/50	.12	.05	.78	-.40	.09
.90	90/10	.23	.07	.52	-.02	.21

Figure 4
 Mean *D* by Method and Selected Population Loading
 From Replication Study (*N* = 100)
 (a) Threshold = 50/50



(b) Threshold = 90/10



gest that for datasets containing dichotomous data, factor extraction techniques are more likely to recover the underlying structure when the sample sizes are large, the data are not highly

skewed, and population loadings are high. These findings would be expected from basic factor analysis theory. However, the finding of a negligible loading effect for ULS-PHI was not expected. The replication investigated the notion that this could have occurred as a result of sampling error, but it was not found to be related to large differences in either the variance of the distribution of deviations around the mean or in the skewness of the distributions. In many common situations, it appears that ULS-PHI will suffice.

The research in this study was limited to least-squares methods of factor analyzing categorical data. This restriction to least-squares methods allowed comparability, because McDonald's NOHARM method is based on least-squares factor extraction. Also, the methods denoted ULS-TC and LISCOMP involve the use of tetrachoric correlations as input matrices, which precludes in many cases the use of maximum likelihood factor extraction techniques because the factor extraction matrices may not be positive semi-definite (Christofferson, 1975).

The stability of selected results of the initial study was supported by the replication study. The replication study reported above could be expanded to cover all the different combinations of sample size, population loadings, and threshold values studied. This would permit the use of statistical tests to evaluate the significance of the differences between the four methods studied. Further simulations are needed, however, that will study the effect of changing the number of variables in the model, having more than one underlying factor, having a broader range of sample sizes, threshold values, and population loadings, and investigating generalized least-squares or maximum likelihood factor analysis procedures. Conducting simulations with datasets with large samples would permit a fairer evaluation of the weighted approach to least-squares factor extraction using LISCOMP.

In this study, non-normal measurement of a normally distributed trait was simulated by using categorized *z* scores to create skewed distributions. As a result, the conclusions depend on the

normality of the generating distribution. Results could be expected to be different if a skewed distribution of the latent trait was used to generate scores (McDonald & Ahlawat, 1974). Therefore, further research is needed that will examine the effect of non-normal distributions of the generated values of the latent traits. It is highly probable that LISCOMP and NOHARM would perform better under such conditions.

Conclusions

Burt (1950) argued that "if factor analysis is to be adequately employed in problems of individual psychology, a rigorous statistical treatment is needed for qualitative data no less than quantitative. . . ." (p. 184). Theoretically, more rigorous methods for handling categorical data are more available today than ever before. The question facing applied psychologists is whether these methods are more helpful than traditional procedures.

This study examined the narrower question of whether two newer (and also more complicated) methods for performing least-squares factor analysis on datasets containing dichotomous variables—LISCOMP and NOHARM—are better able to reproduce the underlying structure than two computationally simpler methods. The results suggest no clearcut answer to the practical question of which method should be used, but the method needs to be based on empirical knowledge of the skewness of the variables and theoretical knowledge about the underlying population loadings. Sample size is also important, particularly in determining whether factor extraction can be performed (i.e., whether the input matrix is positive definite).

Based on the parameters studied, the newer least-squares methods for factor analyzing dichotomous data (NOHARM, LISCOMP) did *not* outperform factor analysis using the tetrachoric correlation coefficients estimated from bivariate tables of the observed variables as input to the analysis. These methods also did not outperform factor analysis using phi in every case. There may be instances, however, when more specific tech-

niques may be more precise and should be considered, but they are not clear from the findings of this study.

References

- Alexander, R. A., Alliger, G. M., Carson, K. P., & Barrett, G. (1985). The empirical performance of measures of association in the 2×2 table. *Educational and Psychological Measurement*, 45, 79-87.
- Becker, M. P., & Clogg, C. C. (1986). *A note on approximating correlations from odds ratios*. Unpublished manuscript.
- Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 179-197.
- Browne, M. W. (1968). A comparison of factor analytic techniques. *Psychometrika*, 33, 267-334.
- Burt, C. (1950). The factorial analysis of qualitative data. *British Journal of Psychology (Statistical Section)*, 3, 166-185.
- Carroll, J. B. (1961). The nature of data, or how to choose a correlation coefficient. *Psychometrika*, 26, 347-372.
- Christofferson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40, 5-32.
- Divgi, D. R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika*, 44, 169-172.
- Fraser, C., & McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research*, 23, 267-269.
- Holley, J. W., & Guilford, J. P. (1964). A note on the G-index of agreement. *Educational and Psychological Measurement*, 24, 749-753.
- Kim, J. O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues*. Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-014. Beverly Hills CA: Sage Publications.
- Kirk, D. D. (1973). On the numerical approximation of the bivariate normal (tetrachoric) correlation coefficient. *Psychometrika*, 38, 259-268.
- McDonald, R. P. (1967). *Nonlinear factor analysis*. Psychometric Monographs, 15, 1-167.
- McDonald, R. P. (1969). The common factor analysis of multicategory data. *British Journal of Mathematical and Statistical Psychology*, 22, 165-175.
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34, 100-117.
- McDonald, R. P. (1982). Linear versus nonlinear models in item response theory. *Applied Psychological Measurement*, 4, 379-396.
- McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale NJ: Erlbaum.

- McDonald, R. P., & Ahlwat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, 27, 82-99.
- Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, 11, 3-31.
- Muthén, B. (1979). A structural probit model with latent variables. *Journal of the American Statistical Association*, 74, 807-811.
- Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, 22, 43-65.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered, categorical and continuous latent variable indicators. *Psychometrika*, 49, 115-132.
- Muthén, B. (1987). *LISCOMP: Analysis of linear structural equations with a comprehensive measurement model*. Mooresville IN: Scientific Software, Inc.
- Muthén, B., & Christofferson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika*, 46, 407-419.

SAS Institute, Inc. (1985). *SAS user's guide: Statistics, Version 5 Edition*. Cary NC: Author.

Acknowledgments

The authors thank two anonymous reviewers for their useful suggestions and Fumiaki Hamagami for his assistance with the graphics. The majority of this research was undertaken while the first author was a doctoral candidate at the University of Virginia. It was completed during his postdoctoral fellowship in Clinical Services Research at the University of Pittsburgh, NIMH grant number 5T32MH17184-06. This research was also supported by grant AGO7137 from the National Institute on Aging.

Author's Address

Send requests for reprints of further information to Charles D. H. Parry, CERSA, S.A. Medical Research Council, P.O. Box 70, Tygerberg, 7505 Cape, South Africa, or J. J. McArdle, Jefferson Psychometric Laboratory, Department of Psychology, University of Virginia, Charlottesville VA 22903-2477, U.S.A.