DOCUMENT RESUME

ED 411 269 TM 027 230

AUTHOR	De Champlain, Andre F.; Gessaroli, Marc E.
TITLE	An Empirical Comparison of Two LISREL Chi-Square Goodness-of-Fit Statistics and the Implications for Dimensionality Assessment of Item Response Data.
PUB DATE	1997-03-00
NOTE	19p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Chi Square; Comparative Analysis; *Goodness of Fit; *Item Response Theory; Sample Size; *Statistical Analysis; *Test Items; Test Length
IDENTIFIERS	*Dimensionality (Tests); *LISREL Computer Program; Type I Errors

ABSTRACT

A study was conducted to compare, with simulated unidimensional and two-dimensional sets, the Type I error probabilities and rejection rates obtained with two versions of the LISREL computer program, the earlier version PRELIS/LISREL 7 and the later version PRELIS2/LISREL8, a version that corrects the asymptotic covariance matrix. Unidimensional data sets were generated according to sample sizes of 2,500 and 5,000 and test lengths of 10 and 20 items. Two-dimensional item response vectors were generated for the same sample sizes and test lengths. Findings with unidimensional data sets suggest that the correction in PRELIS2 resulted in higher Type I error rates with the chi-square goodness-of-fit statistic. Rejection rates obtained using the LISREL8 chi-square fit statistic were high across all simulated two dimensional conditions. Results suggest that the LISREL7 chi-square fit statistic should be recommended to the researcher interested in determining whether the assumption of unidimensionality has been violated. (Contains 2 tables and 29 references.) (SLD)

An Empirical Comparison of Two LISREL Chi-Square Goodness-of-Fit Statistics and the Implications for Dimensionality Assessment of Item Response Data

André F. De Champlain and Marc E. Gessaroli

National Board of Medical Examiners

Paper presented at the meeting of the American Educational Research Association

Chicago, Illinois

Friday, March 28, 1997

RUNNING HEAD: A COMPARISON OF TWO LISREL CHI-SQUARE STATISTICS

An Empirical Comparison of Two LISREL Chi-Square Goodness-of-Fit Statistics and the Implications for Dimensionality Assessment of Item Response Data

Item response theory (IRT) models have been utilized extensively by researchers and practitioners alike over the past three decades to address a myriad of measurement-related issues. These issues include score equating (Lord, 1977; 1980) and, more recently, the development of computerized adaptive test forms (Hambleton, Zaal, & Pieters, 1993; Wainer, Dorans, Flaugher, Green, Mislevy, Steiberg, & Thissen, 1990). IRT models have been heavily relied upon in the assembly of nationally administered admissions (Stocking, 1988) as well as licensure examinations (Luecht, De Champlain, & Nungester, 1996).

The legitimate use of common IRT models, such as the family of logistic models implemented in popular software packages (e.g. BIGSTEPS; Linacre & Wright, 1993; BILOG; Mislevy & Bock, 1990), nonetheless requires that several strict assumptions be met. One of these assumptions is unidimensionality of the latent proficiency space. For example, the three-parameter logistic IRT model (Lord & Novick, 1968) given by,

$$P_i(x_i=1\,|\,a_i,b_i,c_i,\theta_j)=c_i+(1-c_i)\;\frac{e^{Da_i(\theta_j-b_i)}}{1+e^{Da_i(\theta_j-b_i)}}, \qquad \textbf{(1)}$$

assumes that the probability of correctly answering item $\underline{i}$ (denoted by $\underline{x}_i$=1) is dependent upon item discrimination ($\underline{a}$), difficulty ($\underline{b}$) and lower asymptote ($\underline{c}$) parameters as well as a latent trait or proficiency ($\underline{\theta}$) postulated to underlie the item responses. Clearly, the assumption of unidimensionality is often compromised with actual achievement data sets where the response to an item is usually dependent on not only the hypothesized proficiency but also on several other ancillary abilities. This is illustrated quite clearly by the dependencies that frequently exist among reading comprehension items referring to a common stem (passage). In that instance, the dimensional structure of the item response set matrix is (potentially) augmented by the presence of content-related factors that are unrelated to the proficiency underlying the item response matrix (e.g., reading proficiency). This has led to the elaboration of a multitude of descriptive and inferential statistics to assess dimensionality, or more commonly, departure from the assumption of unidimensionality. De Champlain & Gessaroli (in press) have provided an outline of most of the procedures proposed thus far in this area along with their respective contributors.

The use of indices and statistics based on nonlinear factor analysis (NLFA) to assess the dimensionality of item response data has proven to be popular due primarily to the efforts of Bartholomew (1983), McDonald (1967) and Takane and de Leeuw (1987). These researchers have shown that common IRT models and NLFA functions are mathematically equivalent. Chi-square goodness-of-fit statistics have been derived for use with both limited and full information factor analytic models as well as a variety of estimation procedures (Bock, Gibbons, & Muraki, 1988; Gessaroli & De Champlain, 1996; 1997; Gibbons & Hedeker, 1992; Jöreskog & Sörbom, 1993a; 1993b; Muthén, 1978).

Among the factor analytic packages that are commercially available, PRELIS2/LISREL8 (Jöreskog & Sörbom, 1993a; 1993b) is particularly appealing in that it allows the user to fit confirmatory factor analytic models to a dichotomous item response matrix via several estimation procedures. For example, it is possible to determine the extent to which the assumption of unidimensionality has been violated with a given data set by simply fitting a one-factor model to a data set prior to calibrating the item responses using an IRT model. The parameters of factor analytic models in LISREL are estimated so as to minimize the following fit function:

$$F=(s-\sigma)^{\prime}\ W^{-1}(s-\sigma), \qquad \textbf{(2)}$$

where

$\underline{s}$ = Sample item covariance matrix;

$\underline{\sigma}$ = Population covariance matrix;

$\underline{W}^{-1}$ = A weight matrix referred to as the correct weight matrix.

With dichotomous item responses, $\underline{s}$ usually corresponds to sample estimates of the threshold and tetrachoric correlations; $\underline{\sigma}$ contains the population threshold and tetrachoric correlation values and $\underline{W}^{-1}$ is a consistent estimator of the asymptotic covariance matrix of $\underline{s}$. Note that the latter weight matrix can be estimated solely with the LISREL7 and LISREL8 versions of the program (Jöreskog & Sörbom, 1991a; 1991b; 1993a; 1993b).

A chi-square goodness-of-fit statistic, provided in LISREL7 and LISREL8 to aid in assessing model fit, is given by

$$\chi^2 = (N-1) * Min(F), \quad (3)$$

where $\underline{N}$ corresponds to the number of examinees in the sample and $\underline{Min\ (F)}$ is the minimum value of the fit function given in Equation (2) for a specific model. This statistic is distributed asymptotically as a chi-square distribution with degrees of freedom equal to

$$.5(p)*(p + 1) - t,$$

where $\underline{p}$ is equal to the number of items and $\underline{t}$ is the number of independent parameters estimated in the model.

Researchers have proposed several methods for estimating the asymptotic variances and covariances of polychoric correlations (Gunsjö, 1994; Muthén, 1984). Christoffersson and Gunsjö (1996) suggest using a Taylor expansion for the equations that define the two-step estimator for tetrachoric/polychoric correlations. Jöreskog (1994) suggests that a contingency table approach can be utilized: thresholds can be estimated from the univariate marginals while polychoric correlations can be estimated from the bivariate marginals. Specifically, assuming that

$$\hat{\tau}=(\hat{\tau}_1,\hat{\tau}_2,\ldots,\hat{\tau}_k) \qquad (4)$$

is the vector of estimated threshold values and

$$\hat{\rho}=(\hat{\rho}_{21},\hat{\rho}_{31},\hat{\rho}_{32},\hat{\rho}_{41},\hat{\rho}_{42},\hat{\rho}_{43},\ldots,\hat{\rho}_{k,k-1}) \qquad (5)$$

is a vector containing all estimated polychoric correlation values, then the asymptotic matrix of $\hat{\rho}$ can be obtained as

$$NACov(\hat{\rho}_{gh},\hat{\rho}_{ij})=\sum_{a=1}^{m_g}\sum_{b=1}^{m_h}\sum_{c=1}^{m_i}\sum_{d=1}^{m_j}\gamma_{ab}^{(gh)}NCov(p_{ab}^{(gh)},p_{cd}^{(ij)})\gamma_{cd}^{(ij)}, \qquad (6)$$

where

$$NCov(p_{ab}^{(gh)},p_{cd}^{(ij)})=\pi_{abcd}^{(ghij)}-\pi_{ab}^{(gh)}\pi_{cd}^{(ij)}. \qquad (7)$$

In equations (6) and (7) $\pi^{(ghij)}_{abcd}$ are the probabilities of variables g, h, i and j in a four-way contingency table, the latter being consistently estimated by their corresponding sample proportions $p^{(ghij)}_{abcd}$. Jöreskog (1994) also offers an alternative method of estimating these probabilities without the use of the four-way contingency table. Readers interested in obtaining more information regarding this approach and other issues should refer to Jöreskog (1994) for more detail.

It is important to point out that the estimation procedure advocated by Jöreskog (1994) has only been implemented in the most recent version of LISREL (i.e., LISREL8). In fact, Jöreskog & Sörbom (1993a) clearly state that in earlier versions of the software (e.g., PRELIS/LISREL7; Jöreskog & Sörbom, 1991a; 1991b) the asymptotic covariance matrix is incorrect as it is based on the sometimes erroneous premise that two different polychoric correlations are asymptotically uncorrelated for given thresholds. The authors state that the correction implemented in the more recent version of LISREL enables the user to obtain a consistent estimate of the asymptotic covariance matrix of polychoric correlations without having to accept the simple assumption inherent in the earlier version of the program. This improvement should also, according to Jöreskog & Sörbom (1993a), improve the chi-square goodness-of-fit statistic (i.e., yield better control over Type I error probabilities as well as increase power). In spite of this claim, it is important to point out that little empirical work has been undertaken to compare Type I error rates and rejection rates of the chi-square goodness-of-fit statistic prior to and after implementation of the correction brought to the estimation of the asymptotic covariance matrix.

The purpose of the present investigation was to compare, with simulated unidimensional and two-dimensional data sets, Type I error probabilities and rejection rates obtained with the PRELIS/LISREL7 (prior to correction) and PRELIS2/LISREL8 (after correction) chi-square goodness-of-fit statistics.

Methods

<u>Unidimensional conditions</u>

In the first part of this investigation, the empirical Type I error rates of both statistics were computed under various conditions. Unidimensional item response vectors were simulated using a two-parameter logistic IRT function. The latter function is equivalent to the model outlined in Equation (1) with a zero lower asymptote parameter value. In addition, the unidimensional data sets were generated according to two sample sizes (2500 and 5000 examinees) as well as two test lengths (10 and 20 items).

The IRT item parameters used to simulate the item response vectors were selected from a nationally administered admissions examination. Note that the simulated 20 item data sets were composed of two 10 item tests. The item parameters utilized to simulate responses to items 11-20 were therefore identical to those selected to generate responses to items 1-10. Proficiencies were randomly generated from a N(0,1) distribution. Each cell of this 2 x 2 design was replicated 100 times for a total of 400 unidimensional data sets.

Two-dimensional conditions

In the second part of this investigation, two-dimensional item response vectors were simulated using the following multidimensional two-parameter compensatory logistic IRT model (Reckase, 1985)

$$P_i(x_i=1\,|\,\underline{a}_i, d_i, \underline{\theta}_j) = \frac{e^{(\underline{a}_i\underline{\theta}_j + d_i)}}{1 + e^{(\underline{a}_i\underline{\theta}_j + d_i)}}, \qquad \textbf{(8)}$$

where

$\underline{\mathbf{a}}_i$ = a vector of discrimination parameters for item i;

$\underline{d}_i$ = a scalar parameter related to the difficulty of item i;

$\underline{\theta}_j$ = a latent trait vector.

These two-dimensional item response vectors were generated according to the same two sample sizes (2500 and 5000 examinees) and two test lengths (10 and 20 items) outlined in the previous section of the proposal. These vectors were also generated according to the following dimension dominance and latent trait correlation conditions:

- Dimension dominance: 50% of the items on $\underline{\theta}_1$ and 50% of the items on $\underline{\theta}_2$.

  80% of the items requiring $\underline{\theta}_1$ and 20% of the items on $\underline{\theta}_2$.

This corresponds to a weak two-dimensional structure.

- Latent trait correlation: 0.0 and 0.7.

The parameters used to generate the two-dimensional item response vectors were identical to those selected for the unidimensional simulations. As was previously outlined, the parameters selected to simulate responses to items 11-20 were the same as those employed to generate responses to items 1-10.

Finally, proficiencies were randomly generated from a $\underline{N}(0,1)$ distribution. Each cell of this 2 x 2 x 2 x 2 design was replicated 100 times for a total of 1600 two-dimensional data sets.

Analyses

Initially, the asymptotic covariance matrix of the tetrachoric correlations was estimated for all data sets using PRELIS and PRELIS2. The parameters for a one-factor model were then estimated using weighted least-squares which enabled the computation of the chi-square goodness-of-fit statistics with both LISREL7 and LISREL8. A nominal Type I error rate of .05 was selected for all analyses. Regarding unidimensional data sets, a logit-linear analysis was undertaken to model the effects of test length, sample size, asymptotic covariance matrix estimation procedure (i.e. PRELIS vs PRELIS2) and the interaction of the latter variables with respect to decision accuracy , i.e., the number of times the assumption of unidimensionality was accepted and rejected (Type I error). For two-dimensional data sets, the effects of test length, sample size, dimension dominance, latent trait correlation, asymptotic covariance matrix estimation procedure and the various interaction terms of the latter factors with respect to decision accuracy were also estimated via a logit-linear analysis. Note that the logit-linear analyses were undertaken in a forward hierarchical fashion, that is, starting with the simplest main effect and progressing towards incrementally more complex models while heeding to the principle that higher-order effects are included in the model solely if the corresponding lower-order effects are also included. A model was deemed acceptable if its corresponding p-value exceeded 0.15. Also, effects with z-values greater than 2.00 were treated as statistically significant. For the sake of consistency, significant associations in the logit-linear analyses will be discussed only in light of the independent variable(s). For example, should the decision accuracy by asymptotic covariance matrix estimation procedure association be statistically significant, it will be referred to as the effect of asymptotic covariance matrix estimation procedure.

Results

Unidimensional data sets

The number of rejections of the assumption of unidimensionality for each simulated condition is shown in Table 1.

Insert Table 1 about here

Empirical Type I error rates ranged from 0.00 (for LISREL7 chi-square values based on data sets generated to contain 10 items and 2500 examinees as well as 20 items and 2500/5000 examinees) to .30 (for LISREL8 chi-square values associated with data sets simulated to contain 20 items and 2500 examinees). The results from the logit-linear analysis indicate that a model containing the main effects of test length, sample size and asymptotic covariance matrix estimation procedure is sufficient to adequately account for the empirical Type I error rates, $L^2(4) = 5.105$, $p=.277$. Regarding the test length effect, the empirical Type I error rate computed for the 10 item data sets (0.05 or 21/400 incorrect rejections of unidimensionality) was significantly lower than that obtained with the 20 item data sets (0.12 or 49/400 incorrect rejections of unidimensionality). Similarly, with respect to the sample size effect, the empirical Type I error probability estimated for data sets simulated to contain 2500 examinees (0.11 or 43/400 incorrect rejections of the assumption of unidimensionality) was significantly greater than that associated with the 5000 examinee data sets (0.07 or 27/400 incorrect rejections of the assumption of unidimensionality). Finally, and more importantly given the primary aim of this investigation, the asymptotic covariance matrix estimation procedure effect indicates that the empirical Type I error rate computed for the LISREL7 chi-square (0.005 or 2/400 incorrect rejections of the assumption of unidimensionality) was significantly lower than that obtained with the LISREL8 chi-square statistic (0.17 or 68/400 incorrect rejections of the assumption of unidimensionality).

Two-dimensional data sets

The number of rejections of the assumption of unidimensionality for each simulated condition is shown in Table 2.

Insert Table 2 about here

As shown in Table 2, the assumption of unidimensionality was correctly rejected for all simulated data sets using the LISREL8 chi-square statistic. With the exception of two conditions, the assumption of unidimensionality was also correctly rejected for all simulated data sets using the LISREL7 chi-square statistic. More precisely, the use of the LISREL7 chi-square statistic incorrectly led to the acceptance of the null hypothesis of unidimensionality for 50/100 data sets simulated to contain 10 items, 2500 examinees, according to a weak two-dimensional structure and with a correlation of .70 between the two underlying proficiencies. Also, there were four incorrect acceptances of the assumption of unidimensionality with two dimensional data sets generated to contain 10 items, 5000 examinees, according to a weak two-dimensional structure and with a specified correlation of .70 between both underlying proficiencies.

Due to the large number of empty cells in the design attributable to the zero Type II error rates across nearly all conditions, it was impossible to undertake the logit-linear analysis as anticipated.

Nonetheless, it is quite clear from the findings presented in Table 2 that results differed noticeably in only one condition and solely for the LISREL7 chi-square statistic.

Discussion

The assessment of dimensionality is central to both classical and modern test theories. At the most basic level, the validity of a scored-based inference (what Messick, 1989 refers to as the structural aspect of construct validity) rests upon our knowledge of the underlying dimensional structure of an item response matrix. The need to better understand the structure of our data is therefore of the utmost importance. Past research has shown that NLFA models and accompanying fit statistics are very useful for assessing the dimensionality of an item response matrix given their relationship to common IRT functions. The PRELIS2/LISREL8 package in particular offers the practitioner a great deal of flexibility in assessing the fit of NLFA models to item response data. Nonetheless, significant changes have been implemented in the most recent version of the software which could have an impact on the behavior of the chi-square goodness-of-fit statistic and hence the decision made with respect to the underlying structure of a data set.

The findings obtained in this study, with respect to unidimensional data sets, seem to suggest that the correction implemented in PRELIS2, regarding the estimation of the asymptotic covariance item response matrix, resulted in higher Type I error rates with the chi-square goodness-of-fit statistic. With the exception of one condition in which data sets were simulated to contain 10 items and 5000 examinees, empirical Type I error rates obtained with the LISREL8 chi-square statistic were well beyond two standard errors of the nominal alpha value (.05). It is important, however, to reiterate that the fit statistic provided in both LISREL packages is chi-square distributed asymptotically. It is possible that the empirical Type I error rates estimated with the LISREL8 chi-square fit statistic would adhere more closely to the nominal alpha level with sample sizes exceeding those that were simulated in the present investigation. Nonetheless, it is also quite possible that these larger sample sizes would represent unrealistic testing situations. Based on the logit-linear analysis results, the LISREL7 chi-square fit statistic seems to be preferable to the LISREL8 chi-square statistic given its greater control over Type I error probability (at least with data sets that resemble those that were simulated in this study). It should be pointed out that the latter fit statistic even appeared to be quite conservative in that it led to very few rejections of the assumption of unidimensionality.

Not surprisingly given its inflated Type I error probabilities, rejection rates obtained using the LISREL8 chi-square fit statistic were high across all simulated two-dimensional conditions. With the exception of one condition in which item response matrices were generated to contain 10 items, 2500 examinees, reflect a weak two-dimensional structure and a correlation of .70 between underlying proficiencies, high rejection rates were also noted with the LISREL7 chi-square fit statistic. Examining both empirical Type I error probabilities and rejection rates, it appears as though the LISREL7 chi-square fit statistic should be recommended to the researcher interested in determining whether the assumption of unidimensionality has been violated or not with a data set reflecting the conditions simulated in the current investigation.

Having said this, it is important to state that these findings should be interpreted in light of several caveats. First, the reported findings are highly dependent upon the conditions that were simulated and generalizations to other configurations should be undertaken cautiously, if at all. For example, the two test lengths that were examined obviously did not reflect most testing situations. However, due to memory restrictions associated with PRELIS/LISREL7 it was not possible to analyze data sets that contained more than 20 items. Second, it is important to re-emphasize that the purpose of this study was to examine the behavior of both chi-square statistics in several conditions that would hopefully allow us to gather practical information regarding both procedures. Obviously, numerous additional simulations should be undertaken before making any definitive statements about the Type I and Type II error rates of both statistics.

It is hoped nonetheless that the results obtained in this study will provide valuable preliminary information regarding the behavior of the LISREL7 and LISREL8 chi-square goodness-of-fit statistics. It is also hoped that these initial findings will foster future research that will bridge the areas of IRT and structural equation modeling with respect not only to goodness-of-fit but also to other issues of common interest that would benefit from a greater collaboration between both fields.

References

Bartholomew, D.J. (1983). Latent variable models for ordered categorical data. Journal of Econometrics, 22, 229-243.

Bock, D.R., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. Applied Psychological Measurement, 12, 261-280.

Christoffersson, A., & Gunsjö, A. (1996). A short note on the estimation of the asymptotic covariance matrix for polychoric correlations. Psychometrika, 61, 173-175.

De Champlain, A.F., & Gessaroli, M.E. (in press). Assessing the dimensionality of item response matrices with small sample sizes and short test lengths. Applied Measurement in Education.

Fraser, C., & McDonald, R.P. (1988). NOHARM: Least-squares item factor analysis. Multivariate Behavioral Research, 23, 267-269.

Gessaroli, M.E., & De Champlain, A.F. (1997, March). Assessing dimensionality using a likelihood-ratio chi-square test based on a nonlinear factor analysis of item response data. Paper presented at the meeting of the National Council on Measurement in Education, Chicago, IL.

Gessaroli, M.E., & De Champlain, A. F. (1996). Using an approximate chi-square statistic to test for the number of dimensions underlying the responses to a set of items. Journal of Educational Measurement, 33, 157-179.

Gibbons, R.D., & Hedeker, D.R. (1992). Full-information item bi-factor analysis. Psychometrika, 57, 423-436.

Gunsjö, A. (1994). Factor analysis of ordinal variables. Unpublished doctoral dissertation, Uppsala University, Uppsala.

Hambleton, R.K., Zaal, J.N., & Pieters, J.P.M. (1991). Computerized adaptive testing: Theory, applications and standards. In R.K. Hamleton & J.N. Zaal (Eds.), Advances in educational and psychological testing (pp. 341-366). Boston: Kluwer.

Jöreskog, K.G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. Psychometrika, 59, 381-389.

Jöreskog, K.G., & Sörbom, D. (1991a). PRELIS user's reference guide. Chicago: Scientific Software International.

Jöreskog, K.G., & Sörbom, D. (1991b). LISREL7 user's reference guide. Chicago: Scientific Software International.

Jöreskog, K.G., & Sörbom, D. (1993a). PRELIS2 user's reference guide. Chicago: Scientific Software International.

Jöreskog, K.G., & Sörbom, D. (1993b). LISREL8 user's reference guide. Chicago: Scientific Software International.

Linacre, J.M., & Wright, B.D. (1993). A user's guide to BIGSTEPS: Rasch model computer program. Chicago, IL: Mesa Press.

Lord, F.M. (1977). Practical applications of characteristic curve theory. Journal of Educational Measurement, 14, 117-138.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale: Lawrence Erlbaum.

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

Luecht, R.M., De Champlain, A.F., & Nungester, R.J. (1996, June). Maintaining content validity in computer adaptive testing (CAT). Paper presented at the Ottawa International Conference on Medical Education and Assessment, Maastricht, NE.

McDonald, R.P. (1967). Nonlinear factor analysis. Psychometrika Monograph No. 15, 32(4,Pt.2).

Messick, S. (1989). Validity. In R.L. Linn (Ed.), Educational Measurement (pp. 13-103). New York: American Council on Education.

Mislevy, R.J., & Bock, R.D. (1990). BILOG 3: Item analysis and test scoring with binary logistic models. Mooresville, IN: Scientific Software, Inc.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. Psychometrika, 43(4), 551-560.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49, 115-132.

Reckase, M.D. (1985). The difficulty of test items that measure more than one ability. Applied Psychological Measurement, 9, 401-412.

Stocking, M.L. (1988). Some considerations in maintaining adaptive test item pools (ETS Research Report RR-88-33-ONR). Princeton, NJ: Educational Testing Service.

Takane, Y. & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. Psychometrika, 52, 393-408.

Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J., Steinberg, L., & Thissen, D. (1990). Computerized adaptive testing - A primer. Hillsdale: Lawrence Erlbaum.

Table 1

Number of rejections of the assumption of unidimensionality per 100 data sets: Unidimensional conditions

	10 items		20 items	
	N=2500	N=5000	N=2500	N=5000
LISREL7	0	2	0	0
LISREL8	13	6	30	19

Table 2
Number of rejections of the assumption of unidimensionality per 100 data sets:

Two-dimensional conditions

Test length	Sample size	Dimension dominance	Proficiency correlation	LISREL version	Number of rejections
10	2500	80%:20%	0.00	LISREL7	100
10	2500	80%:20%	0.70	LISREL7	50
10	2500	50%:50%	0.00	LISREL7	100
10	2500	50%:50%	0.70	LISREL7	100
10	2500	80%:20%	0.00	LISREL8	100
10	2500	80%:20%	0.70	LISREL8	100
10	2500	50%:50%	0.00	LISREL8	100
10	2500	50%:50%	0.70	LISREL8	100
10	5000	80%:20%	0.00	LISREL7	100
10	5000	80%:20%	0.70	LISREL7	96
10	5000	50%:50%	0.00	LISREL7	100
10	5000	50%:50%	0.70	LISREL7	100
10	5000	80%:20%	0.00	LISREL8	100
10	5000	80%:20%	0.70	LISREL8	100
10	5000	50%:50%	0.00	LISREL8	100
10	5000	50%:50%	0.70	LISREL8	100
20	2500	80%:20%	0.00	LISREL7	100
20	2500	80%:20%	0.70	LISREL7	100
20	2500	50%:50%	0.00	LISREL7	100
20	2500	50%:50%	0.70	LISREL7	100
20	2500	80%:20%	0.00	LISREL8	100
20	2500	80%:20%	0.70	LISREL8	100
20	2500	50%:50%	0.00	LISREL8	100
20	2500	50%:50%	0.70	LISREL8	100
20	5000	80%:20%	0.00	LISREL7	100
20	5000	80%:20%	0.70	LISREL7	100
20	5000	50%:50%	0.00	LISREL7	100
20	5000	50%:50%	0.70	LISREL7	100
20	5000	80%:20%	0.00	LISREL8	100
20	5000	80%:20%	0.70	LISREL8	100
20	5000	50%:50%	0.00	LISREL8	100
20	5000	50%:50%	0.70	LISREL8	100

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

ERIC

TM027230

I. DOCUMENT IDENTIFICATION:

Title: AN EMPIRICAL COMPARISON OF TWO LISREL CHI-SQUARE GOODNESS-OF-FIT STATISTICS AND THE IMPLICATIONS FOR DIMENSIONALITY ASSESSMENT OF ITEM RESPONSE DATA.

Author(s): ANDRE F. DECHAMPLAIN & MARC E. GESSAROLI

Corporate Source: NBME	Publication Date: 3/28/97

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] **Check here** — Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

Sample sticker to be affixed to document

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

[ ] **or here** — Permitting reproduction in other than paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Andre DeChamplain	Position: PSYCHOMETRICIAN
Printed Name: ANDRE DE CHAMPLAIN	Organization: NATIONAL BOARD OF MEDICAL EXAMINERS
Address: 3750 MARKET ST. PHILADELPHIA, PA 19104	Telephone Number: (215) 590-9565
	Date: 3/31/97

OVER

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064
202 319-5120

February 21, 1997

Dear AERA Presenter,

Congratulations on being a presenter at AERA[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a printed copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at http://ericae2.educ.cua.edu.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (523)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to:

AERA 1997/ERIC Acquisitions
The Catholic University of America
O'Boyle Hall, Room 210
Washington, DC 20064

This year ERIC/AE is making a **Searchable Conference Program** available on the AERA web page (http://aera.net). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1]If you are an AERA chair or discussant, please save this form for future use.