DOCUMENT RESUME

ED 333 030 TM 016 530

AUTHOR Nandakumar, Ratna
TITLE Assessing Dimensionality of a Set of Items--Comparison of Different Approaches.
PUB DATE Apr 91
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Ability Identification; Comparative Analysis; Correlation; *Evaluation Methods; *Factor Analysis; High Schools; High School Students; Simulation; *Test Items
IDENTIFIERS Armed Service Vocational Aptitude Battery; National Assessment of Educational Progress; *Unidimensionality (Tests)

ABSTRACT

Performance in assessing the unidimensionality of tests was examined for four methods: (1) W. F. Stout's procedure (1987); (2) the approach of P. W. Holland and P. R. Rosenbaum (1986); (3) linear factor analysis; and (4) non-linear factor analysis. Each method was examined and compared with the others using simulated and real test data. Seven data sets were simulated, three unidimensional and four two-dimensional, all with 2 000 examinees. Two levels of correlation between abilities were considered. Eight different real test data sets were used: four were unidimensional, and four were two-dimensional. Real data came from the National Assessment of Educational Progress tests for U.S. history and literature for grade 11 and from the Armed Services Vocational Aptitude Battery for grade 10 for arithmetic reasoning and general science. Findings suggest that, while linear factor analysis overestimated the number of underlying dimensions, the other three methods correctly confirmed unidimensionality but differed in their ability to detect a lack of dimensionality. Stout's procedure showed excellent power in detecting a lack of unidimensionality. Holland and Rosenbaum's procedure and the non-linear factor analysis approach showed good power provided the correlation between abilities was low. A 40-item list of references is included. Four tables present study data. (SLD)

# Assessing Dimensionality of a Set of Items—Comparison of Different Approaches

Ratna Nandakumar
University of Delaware



Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, April, 1991.

The author is thankful to Brian Junker, William Stout, and Larry Kortering for helpful comments and suggestions on this article.

## Assessing Dimensionality of a Set of Items—Comparison of Different Approaches


### Abstract

This study examines the performance of the following four methodologies for assessing unidimensionality: Stout's procedure, Holland and Rosenbaum's approach, linear factor analysis, and nonlinear factor analysis. Each method is examined and compared with other methods on simulated test data and real test data. Seven data sets were simulated: three unidimensional test data, and four two–dimensional test data, all with 2000 examinees. Two levels of correlation between abilities were considered ($\rho$=.3 and $\rho$=.7). Eight different real test data were used: four of them are known unidimensional test data, and the other four were two–dimensional test data created from unidimensional tests. Findings suggest that, while the linear factor analysis overestimated the number of underlying dimensions, the other three methods correctly confirmed unidimensionality but differed in their ability to detect lack of unidimensionality. Stout's procedure showed excellent power in detecting lack of unidimensionality; Holland and Rosenbaum's and nonlinear factor analysis approaches showed good power provided the correlation between abilities is low.

It is well known that most item response theory models require the assumption of unidimensionality. According to Lord and Novick (1968), dimensionality is defined as the total number of abilities required to satisfy the assumption of local independence. If there is only one ability affecting the responses of a set of items to meet the assumption of local independence then that set is referred to as a unidimensional set. It is also been long argued that test items are multiply determined (Humphreys, 1981, 1985, 1986; Hambleton & Swaminathan, 1985; Reckase, 1979, 1985; Stout, 1987; Traub, 1983; Yen, 1985) and several abilities unique to items or common to relatively few items are inevitable. The ability which the test is intended to measure (i.e., the ability common to all items) will be referred to as the dominant ability and abilities unique to or influencing few items will be referred to as minor abilities. Given that tests are multiply determined, it is intuitively clear that in order to satisfy the assumption of unidimensionality it is required that a given test measures a single dominant ability. A number of simulation studies have demonstrated that dominant ability can be recovered well, using computer programs such as LOGIST, in tests with one dominant factor in the presence of several minor factors (Reckase, 1979; Drasgow & Parsons, 1983; Harrison, 1986). Although counting only dominant dimensions violates Lord and Novick's (1968) definition of dimensionality, it is commonly accepted that in order to apply unidimensional item response theory models it is sufficient to show that there is one dominant ability underlying the responses to a set of items.

Stout (1987, 1990) provided a mathematically rigorous definition of dominant dimensionality referred to as essential dimensionality, and provided a statistical test to assess essential unidimensionality of a set of items. Essential dimensionality is the total number of abilities required to satisfy the assumption of essential independence. Essential independence and essential dimensionality are the weaker forms of local independence and traditional dimensionality (Lord & Novick, 1968), respectively. Stout's definition of essential dimensionality uses an infinite item pool item response theory framework wherein the item pool is conceptualized as the consequence of continuing the test construction

process in the same manner beyond the construction of the N items of the finite test being studied. Hence essential dimensionality is defined for the item pool.

In assessing essential unidimensionality using Stout's procedure, one is assessing the likelihood that the given set of items comes from an essentially unidimensional item pool. The major focus in assessing essential unidimensionality of a given set of item responses is to determine how minor the influence of minor abilities is and whether the influence of the minor abilities can be ignored in assessing essential unidimensionality.

Historically speaking, linear factor analysis has been used to assess the dimensionality of the latent space underlying a set of items. If the results indicate a one–factor solution then it can be inferred that one dominant ability is influencing item responses. There are, however, a number of technical as well as methodological problems associated with using linear factor analyses to assess dimensionality. For example, difficulty level of items and guessing level of multiple choice items can each play a major role in altering the factor structure of item responses resulting in an overestimation of the number of underlying factors (for details see Carroll, 1945, Hulin, Drasgow, & Parsons, 1983, Zwick, 1987). Consequently, many attempts have been made by researchers in recent years to develop new methods to assess dimensionality. Some of the recently developed methods include nonlinear factor analysis (McDonald & Ahlawat, 1974); Bejar's procedure (Bejar, 1980); order analysis (Wise, 1981); modified parallel analysis (Hulin, Drasgow, & Parsons, 1983); residual analysis (Hambleton & Swaminathan, 1985); Bock's full information factor analysis (Bock, Gibbons, & Murake, 1985); Holland and Rosenbaum's test of unidimensionality, monotonicity, and conditional independence (Rosenbaum, 1984; Holland & Rosenbaum, 1986); Humphreys and Tucker's procedures (Tucker, Humphreys, & Roznowski, 1986); and Stout's unidimensionality procedure (Stout, 1987).

Hattie (1985), Hambleton and Rovinelli (1986), and Berger and Knol (1990) have reviewed several procedures for assessing dimensionality including some of the above mentioned procedures. Their conclusions were that none of the procedures were

satisfactory. The main focus of this paper is to study and compare some of the procedures to assess dimensionality that are most recent, seem promising, and are little studied. Four procedures are considered and compared in this paper: Nonlinear factor analysis, Holland and Rosenbaum's procedure, Stout's procedure, and linear factor analysis. Linear factor analysis is used, because of its historical importance, as a benchmark to compare other procedures. Several unidimensional and multidimensional test data are simulated and used to study the performance of all four procedures for assessing dimensionality. The same procedures are then repeated with real test data.

## Description of Procedures

### Linear Factor Analysis

Linear factor analysis is the most commonly used approach to assess dimensionality. With linear factor analysis, each extracted factor is presumed to represent a dimension or trait and the items that load heavily on a given factor are considered good measures of that dimension. There are a number of fundamental problems associated with applying linear factor analysis to binary data. First, the linear factor analysis assumes that the relationship between the observed variables and the underlying factors is linear and that the variables are continuous in nature. But it can be shown that the relationship between the performance and the underlying latent variable is nonlinear. Hence applying factor analysis to binary responses amounts to approximating the nonlinear relationship to a linear one. As a result, difficulty factors are produced if guessing is allowed, irrespective of whether phi or tetrachoric correlations are used (Hulin, Drasgow, & Parsons, 1983). Secondly, in computing tetrachoric correlations, the cell entries of the fourfold table for a pair of dichotomous items frequently become zero thus making it difficult to determine an appropriate value for the correlation. Thirdly, problems associated with determining the number of significant factors exist.

In this study the statistical package LISCOMP is used to perform exploratory linear factor analysis using tetrachoric correlations. Three different approaches are used to determine the number of significant factors: parallel analysis, the chi–square test of goodness of fit, and goodness of fit statistics (the means and standard deviations of the squares of residual correlations and absolute residuals).

According to parallel analysis (Humphreys & Montenelli, 1975) the eigenvalues of the given correlation matrix are compared with the eigenvalues of the random data. The random data consists of binary responses generated randomly with the same number of items and examinees as that of the given data. The largest eigenvalue from the random data is used as the cutoff point for eigenvalues from the actual data to determine the number of significant factors. That is, the number of eigenvalues of the actual data greater than the largest eigenvalue of the random data is taken as the significant number of factors underlying the given data.

The second method used to determine the number of factors is the chi–square test of goodness of fit. The third method involves comparison of means and standard deviations of squares of residuals and absolute values of residuals after fit of an m–factor model with the corresponding values from the random data. If the residuals are sufficiently small, then one can regard the fit of the model as reasonably satisfactory (McDonald, 1981; Hattie, 1985, Hambleton & Rovinelli, 1986; and Berger & Knol, 1990).

## Nonlinear Factor Analysis

McDonald (1967, 1980, 1982), McDonald and Ahlawat (1974) have demonstrated that applying linear factor analysis to unidimensional binary data yields "nonlinear factors" rather than "difficulty factors". McDonald developed the method of nonlinear factor analysis (NLFA) to account for the nonlinearity of the data as an improvement over linear factor analysis. In the context of item response theory, nonlinear factor analysis

seems appropriate because the latent variable is related to the performance in a nonlinear fashion. The variables in the model can be expressed as polynomial functions of latent traits or factors. For example, a two–factor model with linear and quadratic terms would be of the following form:

$$Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+u_i e_i,\ (i=1,2,\ldots N)$$

where $Y_i$ denotes the examinee's score on item i, $\theta$ denotes the latent trait, $b_{ijk}$ denotes the factor loadings of i–th item on j–th common factor for k–th degree element in the polynomial, and $u_i$ denotes the unique factor loading for item i. Conceptually, NLFA is very appealing and seems appropriate to assess the dimensionality of binary responses conforming to normal ogive or logistic item response models. Hambleton and Rovinelli (1986) have demonstrated the use of NLFA to assess dimensionality and found it to be a promising method. They, however, caution about the criterion for the adequacy of the fit of the model.

In the present study NLFA embodied in the computer program NOFA is used. The fit of the model is studied just as in the case of the linear factor analyses comparing the means and standard deviations of squared residuals and absolute residuals with the corresponding values of random data and linear factor analyses. The chi–square statistic values are not available and hence were not used.

## Holland and Rosenbaum's Test of Lack of Fit of a Unidimensional, Monotone, and Conditional Independent Model

Rosenbaum (1984), and Holland and Rosenbaum (1986) have proved theorems concerning conditional association that can be applied to assess dimensionality. The basic notion in Holland and Rosenbaum's (H&R) theorems is that if the items are locally

independent, unidimensional, and the ICCs are monotone, then the items are conditionally positively associated. Specifically, the conditional covariances between any pair of item response functions of a set of unidimensional dichotomous item responses given any function of the remaining item responses will be nonnegative. This can be hypothesized as

$$H_0\text{: Cov }(X_i, X_j \mid \sum_{i,j\neq k} X_k) \geq 0 \quad \text{vs.} \quad H_1\text{: Cov }(X_i, X_j \mid \sum_{i,j\neq k} X_k) < 0$$

Conditional associations for each pair of items is tested, given the number–right score on the remaining items. The Mantel–Haenszel test (M–H) (Mantel & Haenszel, 1954) is used to test this hypothesis. To perform the M–H test on a given pair of items, a 2x2 contingency table is constructed for the pair for each of the possible number–right score on the remaining items. The M–H statistic is given by:

$$Z = \frac{n_{11+} - E(n_{11+}) + 1/2}{\sqrt{V(n_{11+})}}$$

where $n_{11k}$ denotes the observed number of examinees with total score of k answering both items i and j correctly with $k = 1,2,\ldots K$. $E(n_{11+})$ and $v(n_{11+})$ are the expectation and variance of $n_{11+}$ respectively where the plus subscript denotes the summation over k. The computed Z–value is referred to the lower tail of the standard normal distribution. A statistically significant Z implies that the pair of items in question are not conditionally associated given the sum of the other items, thus inconsistent with the unidimensional model. In this manner the M–H statistic is computed for all $N(N-1)/2$ pairs of items. If a large number of pairs are shown not to be conditionally associated, then the unidimensional assumption is inappropriate.

Since H&R approach tests each item pair with significance level $\alpha$, the simultaneous

inference for all item pairs can be based on Bonferoni bounds (Holland & Rosenbaum, 1986, Junker, 1990, and Zwick, 1987). According to Bonferoni bounds one would accept $H_\circ$ if the number of rejections at level $\alpha$ is around $n\alpha$, where n is the number of tests performed; reject $H_\circ$ if at least one test is rejected at level $\alpha/n$.

Rosenbaum (1984), Zwick (1987), and Ben–Simon and Cohen (1990) have demonstrated the application of H&R approach to assess dimensionality. Ben–Simon and Cohen found the H&R approach to be conservative and erroneously misclassified nearly half of the multidimensional item pools they analyzed as unidimensional. Zwick found H&R approach to be consistent with other procedures investigated in confirming unidimensionality of NAEP reading data.

## Stout's Procedure

Stout (1987) developed a statistical procedure to test the hypothesis of essential unidimensionality, the existence of one dominant dimension. The procedure has several steps. These are briefly described here (for details see Stout, 1987, Nandakumar, 1991). The hypothesis is stated as

$$H_0: d_E=1 \quad \text{vs.} \quad H_1: d_E>1$$

where $d_E$ denotes the essential dimensionality of the item pool in which the given test responses are assumed to be imbedded. The J examinees are partitioned into two groups. One group of examinees is used for exploratory factor analysis to select items for subtests, and the other group of examinees is used to compute Stout's statistic T. The N test items are split into three subsets AT1, AT2, and PT. The items of subtest AT1 are chosen such that they all measure the same dominant ability; the items of AT2 are matched in difficulty with items of AT1 to correct for difficulty and guessing factors in item responses;

and the rest of the items are used for PT. The subtest PT is used to split examinees into K subgroups based on their PT score; the subtests AT1 and AT2 are used to compute the unidimensional statistic T given by:

$$T = (T_1 - T_2)/\sqrt{2},$$

where

$$T_i = \frac{1}{K^{1/2}} \sum_{k=1}^{K} \left[ \frac{\hat{\sigma}_k^2 - \hat{\sigma}_{U,k}^2}{S_k} \right]$$

is computed using items of $AT_i$. The $\sigma_k^2$ and $\sigma_{U,k}^2$ and $S_k$ are given as follows.

The usual variance estimate for subgroup k is given by

$$\hat{\sigma}_k^{\,2} = \Sigma_{j=1}^{J_k} (Y_j^{(k)} - \overline{Y}^{(k)})^2 / J_k,$$

where

$$Y_j^{(k)} = \Sigma_{i=1}^{M} U_{ijk}/M, \text{ and } \overline{Y}^{(k)} = \Sigma_{j=1}^{J_K} Y_j^{(k)}/J_K,$$

with $U_{ijk}$ (1 or 0) denoting the response for item i by examinee j in subgroup k, and $J_k$ denoting the total number of examinees in subgroup k. The "unidimensional" variance estimate for subgroup k is given by

$$\hat{\sigma}_{U,k}^{\,2} = \Sigma_{i=1}^{M} \hat{p}_i^{(k)}(1-\hat{p}_i^{(k)})/M^2,$$

where

$$\hat{p}_i^{(k)} = \Sigma_{j=1}^{J_k} U_{ijk}/J_k.$$

And the standard error of estimate for subgroup k is given by

$$S_k = \left[(\hat{\mu}_{4,k} - \hat{\sigma}_k^4) + \hat{\delta}_{4,k}/M^4\right]^{1/2}/J_k,$$

where

$$\hat{\mu}_{4,k} = \Sigma_{j=1}^{J_k}(Y_j^{(k)} - \bar{Y}^{(k)})^4/J_k, \text{ and}$$

$$\hat{\delta}_{4,k} = \Sigma_{i=1}^{M}\, p_i^{(k)}(1-p_i^{(k)})\,(1-2p_i^{(k)})^2.$$

The computed T value is referred to the upper tail of the standard normal distribution to obtain the significance level. The p–values of unidimensional tests are expected to be large while the p–values of multidimensional tests are expected to be within the margin of the specified level of significance.

Stout's procedure, as refined by Nandakumar and Stout (1991), is used for assessing dimensionality in the present study. Stout's procedure has been found to be discriminating well between unidimensional and two–dimensional tests in a variety of simulated test data for correlation between abilities as high as .7 (Stout, 1987; Nandakumar & stout, 1991). Nandakumar (1991) has shown the usefulness of Stout's procedure to assess essential unidimensionality in the possible presence of several minor abilities. Nandakumar(1989) applied Stout's procedure on several real test data sets and found that the procedure correctly confirmed the unidimensionality of test data that were previously shown to be unidimensional by others. For two–dimensional test data, created by combining the unidimensional test data, Stout's procedure exhibited good power.

## Description of Test Data

### The Simulated Test Data

Seven data sets DATA1–DATA7 are simulated. Out of the seven, three data sets

DATA1–DATA3 are strictly unidimensional tests consisting of 25, 40, and 50 items respectively. The other four data sets DATA4–DATA7 are each two–dimensional with length N=25 and correlation between abilities $\rho$=.3, N=25 and $\rho$=.7, N=50 and $\rho$=.3, and N=50 and $\rho$=.7 respectively. All data sets have 2000 examinees. These test data are described in Table 1. The unidimensional test data are generated according to the three–parameter logistic model. The abilities are independently generated from the standard normal distribution and the item parameters $(a_i, b_i, c_i)$ of real tests as described in Nandakumar (1991) are used in generating item responses. For example, items of DATA 1 have a larger variability in discrimination power $(a_i)$ ranging from 1.22 to 2.82; items of DATA 2 have a smaller variability of $a_i$s ranging from 1.07 to 2.00. For each simulated examinee, the probability of correctly answering each item $P_i(\theta)$ was computed using the three–parameter logistic model. For each item i, a random number between 0 and 1 was generated from a pseudo–uniform distribution. If the computed probability $P_i(\theta)$ is greater than or equal to the random number generated, the examinee was said to have answered the item correctly and was given a score of 1; otherwise the examinee was given a score of 0. The two–dimensional test data were generated according to the multidimensional compensatory model (Reckase & McKinley, 1983). The abilities $\underline{\theta} = (\theta_1, \theta_2)$ were generated from a bivariate normal distribution with both means zero, and both variances one. The correlation coefficient between the abilities varied appropriately. The pseudo guessing level was taken to be .20 for all tests. The discrimination parameters $(a_{i1}, a_{i2})$ for each item were generated as follows:

$$a_{1i} \sim N\left[\frac{\mu}{2}, \frac{\sigma}{\sqrt{2}}\right], \; a_{2i} \sim N\left[\frac{\mu}{2}, \frac{\sigma}{\sqrt{2}}\right],$$

where $\mu$ and $\sigma$ are the mean and standard deviation of the distribution of discrimination parameters of the respective unidimensional tests with the same number of items. Similarly

$b_{1i}$ and $b_{2i}$ were assumed to be independent of each other for each item and were generated as follows:

$$b_{1i} \sim N(\mu, \sigma), \; b_{2i} \sim N(\mu, \sigma),$$

where $\mu$ and $\sigma$ are the mean and standard deviation of the distribution of difficulty parameters of the respective unidimensional test with the same number of items. For example to generate test data DATA4 with N=25 and $\rho$=.3, the means and standard deviations of $a_i$s and $b_i$s of item parameters used for DATA1 were used. The item responses (0,1) were generated exactly as described for unidimensional case by using $P_i(\theta)$ of a two–dimensional compensatory model.

## The Real Test Data

The real test data used in this study came from two different sources. The National Assessment of Educational Progress (NAEP, 1988) data for tests US History (HIST) and Literature (LIT) for grade 11/age 17 were obtained from Educational Testing Services. The Armed Services Vocational Aptitude Battery (ASVAB) data for Arithmetic Reasoning (AR) and General Science (GS) for grade 10 were obtained from Linn, Hastings, Hu, and Ryan (1987). The details about these data sets are described in Table 1. Since all the four test data sets are considered to be unidimensional, they were combined to form pseudo two–dimensional tests (Zwick, 1987; Nandakumar, 1989). Four two–dimensional tests were formed as follows. The test data HSTLIT1 was formed by combining the data of 31 items of HIST with the data of 5 items of LIT randomly selected from 30 items. Since HIST contains more examinees than LIT, excess examinees in HIST are randomly deleted in order to make the lengths of the data sets equal. Similarly the data on 10 items from LIT is combined with the data on 31 items of HIST to form HSTLIT2; and the data on 10 items

from GS are combined with the data on 30 items of AR to form ARGS. These three are pseudo two–dimensional test data because it is not known if the same examinees took both tests. Hence the correlation between the abilities is considered to be zero. The last two–dimensional test HSTGEO consisting of 36 items differs from other two–dimensional tests in that the same examinees took both sets of items. HSTGEO contains 31 history items spanning the US history from colonization period to modern times (HIST) and in addition contains 5 map items requiring the knowledge of geographical location of different countries in the world. This is the actual history test according to NAEP. It was shown using Stout's procedure that the 5 map items formed a separate dimension (Nandakumar, 1989). Hence the data on these 5 map items were removed from the history test to form HIST with 31 items and the original history is treated as a natural (as opposed to pseudo) two–dimensional test (HSTGEO).

## Results

The results of Stout's procedure and the H&R approach will be studied together and compared because of the similarity in the underlying theory and because both of them are statistical tests. Likewise the results of linear and nonlinear factor analysis will be studied and compared together.

### The Simulated Test Data

#### Stout's and H&R Procedures

The results of Stout's procedure and the H&R approach for simulated data are presented at the top of Table 2. For all test data the p–values associated with Stout's procedure indicate that Stout's procedure is able to correctly confirm unidimensionality and detect lack of unidimensionality for both correlation (between abilities) levels $\rho=.3$ and

$\rho$=.7. For example, all three unidimensional test data DATA1–DATA3 have large p–values implying the acceptance of the null hypothesis of essential unidimensionality (here the tests are strictly unidimensional). For two–dimensional data, on the other hand, the associated p–values are very small, strongly rejecting the null hypothesis of essential unidimensionality.

The results of the H&R approach indicate that for unidimensional tests DATA1–DATA3, the number of significant negative partial associations at level $\alpha$ ($\alpha$=.05) are far below the expected number (n$\alpha$), strongly confirming the unidimensional nature of test data. Among the two–dimensional tests, DATA4 and DATA6 (for both $\rho$=.3) were correctly assessed as multidimensional. For DATA4 and DATA6 the number of significant negative partial associations at level $\alpha$ were beyond n$\alpha$ level, and the number of significant negative partial associations beyond level $\alpha$/n were 15 and 1 respectively, making them multidimensional. The test data DATA5 and DATA7 (for both $\rho$=.7), on the other hand, are assessed as unidimensional. For DATA5 and DATA7 the number of significant negative partial associations at level $\alpha$ are within n$\alpha$ level, and the number of significant negative partial associations beyond level $\alpha$/n is zero, hence making them unidimensional tests. It was disappointing to note that for many of the item pairs measuring different traits, in two–dimensional tests, the covariance did not approach significance. One reason for this could be the noise in the conditional score. More research is necessary to draw definite conclusions.

## Linear and Nonlinear Factor Analysis

The computer programs used to do the analyses, LISCOMP and NOFA are heavily computationally intensive and consume enormous CPU time. In addition, LISCOMP program can not handle more than about 40 variables. For these reasons only a selection of simulated data sets were included in the linear factor analyses but all test data were included in the nonlinear factor analyses. The results of linear and nonlinear factor analysis

are presented in Table 3.

Based on parallel analyses, one factor would be retained for DATA1, DATA2, and DATA5; two factors would be retained for DATA4. Whereas according to the p–values associated with a chi–square test of goodness of fit, in Table 3, a two–factor model fits DATA1, beyond four–factor model fits DATA2 and DATA4, and a three–factor model fits DATA5. Similar chi–square values are not available for nonlinear models and hence are not reported.

The goodness of fit statistics, the means and standard deviations of squared residuals and absolute residuals, are reported for all test data in Table 3. The top entry in Table 3 refers to random data (RANDOM) with 25 variables and 2000 examinees. Because of the cost of computations, only one random data is used to compare the goodness of fit statistics. Comparison of goodness of fit statistics of RANDOM with DATA1, it appears that one–factor quadratic model fits the data better than four–factor linear model. Hence nonlinear model accurately confirms the unidimensional nature of items. The one–factor cubic model is no better than the one–factor quadratic model. Similar observation can be made for DATA2. Comparison of goodness of fit statistics for linear and nonlinear factor analysis, it can be seen that for DATA4 and DATA5, two–factor quadratic model fits better than three–factor linear model, confirming two–dimensional nature of data. As expected, the means and standard deviations of squared residuals and absolute residuals is much larger for DATA4 ($\rho$=.3) than for DATA5 ($\rho$=.7), reflecting more multidimensionality. For DATA5, although two–factor quadratic model fits better than one–factor quadratic model, the difference in goodness of fit statistics is so small that one is tempted to accept one–factor quadratic model. Likewise two–factor quadratic model fits better than one–factor quadratic for DATA6 and one–factor quadratic model fits DATA7.

In summary, the linear factor analysis either underestimates or overestimates the number of factors and hence is not adequate for assessing dimensionality. The other three procedures are excellent in confirming unidimensionality. Stout's procedure has

demonstrated greater power in detecting multidimensionality for correlation between abilities as high as .7. H&R and nonlinear factor analysis methods have demonstrated good power provided the correlation between abilities is low.

## The Real Test Data

### Stout's and H&R Procedure

The results of Stout's and H&R for real test data are presented at the bottom of Table 2. For all test data the p–values associated with Stout's procedure indicate that Stout's procedure is able to correctly confirm unidimensionality and detect lack of unidimensionality in cases where a test data is contaminated by as few as 15% of the data from a second dominant dimension (for example, HSTLIT1).

For LIT data, the p–value associated with Stout's procedure is in the border line tending towards acceptance of $H_\circ$. The p–values associated with HIST, AR, and GS are large leading to acceptance of $H_\circ$. Relatively small p–values for LIT and AR suggest that there is some multidimensionality present in these test data. For all two–dimensional tests, the associated p–values are very small strongly confirming multidimensional nature of these data. This is true both for correlated abilities (HSTGEO) and for uncorrelated abilities (HSTLIT1, HSTLIT2, ARGS). The p–value for HSTLIT1 is larger than for HSTLIT2 suggesting greater degree of multidimensionality.

The results of H&R approach is consistent with Stout's procedure in assessing unidimensionality. Whereas for two–dimensional tests, the H&R approach does not seem to exhibit good power. while test data HST, and AR were clearly confirmed as unidimensional, for test data LIT the decision is not clear. Although the number of significant negative partial associations for LIT are less than the maximum allowed ($n\alpha$=22), one of the M–H tests was found to be significant beyond $\alpha/n$ level suggesting significant presence of multidimensionality in the data. For two–dimensional tests

HSTLIT1, ARGS, and HSTGEO, the number of significant negative partial associations is far below the $n\alpha$ level suggesting unidimensional nature of these data. For HSTLIT2, however, the number of significant negative partial associations is well above $n\alpha$ level suggesting presence of multidimensionality but none of the M–H tests were significant beyond level $\alpha/n$ to conform multidimensionality. Hence the decision about dimensionality is not clear although one is tempted towards multidimensionality.

On closer examination it was found that the M–H z–values for many of the item pairs where items were supposed to be measuring different traits were negative but not statistically significant. One explanation for this could be that for these item pairs the conditional score ($\Sigma X_k$), on the basis of which the examinees are classified into different groups is confounded by noise. This is especially true for HSTLIT2 and ARGS where one quarter of the test items are of second dominant dimension. Because of the noise in the conditional score distribution the covariance of item pairs measuring different abilities may not be exhibiting significant negative covariance. Proper conditional score could considerably increase the power of the H&R approach.

Linear and Nonlinear Factor Analysis

The results of linear and nonlinear factor analysis for a selection of tests are reported in Table 4. The results are consistent with the simulated test data in that for all cases nonlinear factor models fit more accurately than linear factor models. According to the chi–square test of goodness of fit, beyond four–factor model fits all test data where linear factor analysis is performed. Based on goodness of fit statistics, one factor quadratic model fits the test data LIT, AR, and HSTLIT better than three– or four–factor linear model. Also one–factor quadratic model fits as well as a two–factor quadratic model. In the interest of parsimony therefore, one–factor quadratic is the right choice. For HSTLIT2 and ARGS two–factor quadratic fits better than one–factor quadratic, and three–factor quadratic is no better than two–factor quadratic. But the distinction in the fit statistics

between one–factor quadratic and two–factor quadratic is not clear. If chi–square statistics were available along with the goodness of fit statistics, it would have aided in the interpretation.

In summary, for real test data, the results are somewhat consistent with simulated test data. Linear factor analysis over estimates the number of underlying dimensions and is not adequate for assessing the fit of the model. Whereas the other three methodologies are excellent in assessing unidimensionality but differed in assessing lack of unidimensionality. Stout procedure has demonstrated greater power than either the H&R or the nonlinear factor analysis methods. With the appropriate conditional score the power of H&R approach could be improved; and with some type of fit statistics the power of nonlinear factor analysis could be improved.

## Discussion

Based on this limited study, findings demonstrate that the linear factor analysis approach to assess dimensionality is not adequate. This finding is consistent with the previous research (see for example, Hambleton & Rovinelli (1986); Hattie, 1984). In contrast to linear factor analysis, Stout's, H&R, and nonlinear factor analysis were each shown to be promising methodologies to assess dimensionality. The findings should be interpreted with caution, in that a limitation of this study was the feature of creation of two–dimensional real test data (except, HSTGEO). The item responses combined from two different tests were not administered to the same group of examinees. The results may have been slightly different had the same examinees taken both sets of items.

In this study all three methodologies exhibited sensitivity to discriminate between one– and two–dimensional test data. For known unidimensional test data, both simulated or real data, all three procedures were able to confirm unidimensionality. For two–dimensional tests, however, the three procedures differed in their ability to detect the

lack of unidimensionality. Stout's procedure rejected the null hypothesis of essential unidimensionality for all two–dimensional tests, both real and simulated tests. The H&R approach confirmed the lack of unidimensionality for two–dimensional simulated tests provided the correlation between abilities was low ($\rho$=.3). For simulated test data with high correlation between abilities ($\rho$=.7) the H&R approach was unable to detect multidimensionality. In addition, for all two–dimensional real test data, the H&R approach was unable to detect multidimensionality. The performance of nonlinear factor analysis methodology was similar to H&R procedure for two–dimensional tests. For simulated test data with $\rho$=.3, the two–factor model with linear and quadratic terms demonstrated adequate fit statistics (smaller means and standard deviations of squared residuals and absolute residuals). For simulated tests with $\rho$=.7, however, the distinction between fit statistics between one–factor and two–factor quadratic models was not evident. Similarly for two–dimensional real test data HSTLIT2 and ARGS, the difference in fit statistics between one–factor and two–factor models with linear and quadratic terms was not evident. The difficulty in deciding about the correct model arises because there is no concrete way of assessing what is meant by sufficiently small for goodness of fit statistics.

In this study the results associated with the H&R approach were consistent with the findings of the Ben–Simon and Cohen's (1990) and Zwick's (1987) studies. The number of significant negative partial associations for unidimensional tests were far below the expected five percent level, making it a very conservative test. Consequently it did not exhibit high power. According to the theorems proved by Holland and Rosenbaum (1986), the conditional score used to compute the covariances can be any function of the latent trait. An appropriate choice of conditional score therefore could maximize the power of H&R approach.

The results of nonlinear factor analyses were consistent with the findings of Hambleton and Rovinelli (1986). Factor models with linear and quadratic terms were able to fit the data better than models with just linear terms. The problem with nonlinear

factor analysis is the appropriate number of polynomial terms to retain in the model. This suggests that some type of adequacy of fit statistics with associated p–values would be necessary to aid in assessing the fit of nonlinear models.

In terms of assessing the degree of multidimensionality, both Stout's and nonlinear factor analysis approaches can be useful. The p–values associated with Stout's procedure and the fit statistics associated with nonlinear factor analysis can be helpful in assessing the degree of multidimensionality. For example, both HIST and AR are unidimensional tests but the associated p–values are .937 and .118 respectively. By contrast for a two–dimensional test HSTLIT2, p=.000. The difference in the p–values mirror the degree of multidimensionality present in the data. Similarly, the difference in fit statistics between one–factor and two–factor quadratic models for DATA1 and DATA4 reflect the degree of multidimensionality.

Just as linear and nonlinear methodologies share the same philosophical theory, Stout's and H&R approaches share the same theoretical framework. The basic rationale for the H&R approach is to reject the locally independent, monotone, unidimensional model if the conditional covariances are significantly negative. By contrast, Stout's procedure rejects the essentially independent, monotone, essentially unidimensional model if the conditional covariances are significantly positive (it can be shown that the expected value of the numerator of Stout's statistic T is mathematically equivalent to average conditional covariances among AT1 items, Stout (1987)). This apparent contradiction in the criterion for assessing unidimensionality may be resolved by noting the subtle difference in item pair covariances under consideration. In the H&R approach one expects the conditional covariance between items measuring different traits to be negative; whereas in Stout's approach one expects the asymptotic conditional covariance between items measuring the same trait to approach zero. Stout's procedure is specifically designed to assess unidimensionality and hence looks for the existence of at least two dominant dimensions. By contrast, the H&R approach looks at all item pairs and detects items that are not

measuring the same trait as other items of the test.

As for the computational time involved, Stout's procedure is most efficient. The computational time involved for other procedures is significantly more. For example, for a 25 item test with 2000 examinees, Stout's procedure uses 4 seconds of CPU time, H&R approach uses 24 seconds, and nonlinear factor analysis uses 42 seconds; for a 50 items test with 2000 examinees, Stout's procedure uses 8 seconds, H&R approach uses 106 seconds, and nonlinear factor analysis uses 191 seconds. As the test length increases, H&R approach requires disproportionately more time, and the same is true for nonlinear factor analysis as test length increases and/or the model gets more complex.

## References


Bejar, I. I. (1980). A procedure for investigating the unidimensionality of achievement tests based on item parameter estimates. Journal of Educational Measurement, 17, 283–296.

Ben–Simon, A., & Cohen, Y. (1990). Rosenbaum's test of unidimensionality: Sensitivity analysis. Paper presented at the annual AERA meeting, Boston.

Berger, M. P., & Knol, D. L. (1990). On the assessment of dimensionality in multidimensional item response theory models. Paper presented at the annual AERA meeting, Boston.

Bock, R. D., Gibbons, R., & Murake, E. (1985) Full–information item factor analysis (MRC Report No. 85–1). Chicago: National Opinion Research Center

Carroll, J. B. (1945) The effect of difficulty and chance success on correlation between items and between tests. Psychometrika, 26, 347–372.

Drasgow, F. & parsons, C. (1983). Applications of unidimensional item response theory models to multidimensional data. Applied Psychological Measurement, 7, 189–199.

Hambleton, R. K. & Swaminathan, H. (1985) Item Response Theory: Principles and applications, Kluwer–nyjhoff Publishers, Boston.

Hambleton, R. K., & Rovinelli, R. J. (1986). Assessing the dimensionality of a set of test items. Applied Psychological Measurement, 10, 287–302.

Harrison, D. (1986). Robustness of IRT parameter estimation to violations of the unidimensionality assumption, Journal of Educational Statistics, 11, 91–115.

Hattie, J. (1984). An empirical study of various indices for determining unidimensionality. Multivariate Behavioral Research, 19, 49–78.

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

Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. Annals of Statistics, 14, 1523–1543.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). Item response theory: Application to psychological measurement. Homewood, Illinois: Irwin.

Humphreys, L. G. (1981). The primary mental ability. In M. P. Friedman, J. P. Das, & N. O'Connor (Eds). Intelligence and learning (pp. 87–102). New York: Plenum Press.

Humphreys, L. G. (1985). General intelligence: An integration of factor, test, and simplex theory. In B. B. Wolman (Ed.), Handbook of intelligence. John Wiley, New York.

Humphreys, L. G. (1986). An analysis and evaluation of test and item bias in the

prediction context. Journal of Applied Psychology, 71, 327–333.

Humphreys, L. G., & Montanelli, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. Multivariate Behavioral Research, 10, 193–205.

Junker, B. W. (1990). Progress in characterizing the monotone unidimensional IRT representation. Paper presented at the annual Office of Naval Research contractor's meeting on model–based psychological measurement, Portland, Oregon.

Linn, R. L., Hastings, N. C., Hu, G., & Ryan, K. E. (1987). Armed Services Vocational Aptitude Battery: Differential item functioning on the high school form, Dayton, OH: USAF Human Resources Laboratory.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores (pp. 359–382). Reading Mass: Addison–Wesley.

McDonald, R. P. (1967). Non–linear factor analysis. Psychometrika Monograph (No. 15)

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. British Journal of Mathematical and Statistical Psychology, 27, 82–89

McDonald, R. P. (1980). The dimensionality of tests and items. British Journal of Mathematical and Statistical Psychology, 34, 100–117.

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, 6, 379–396.

Nandakumar, R. (1989). Validation of Stout's procedure for assessing latent trait unidimensionality with real tests. Paper presented at the 1989 European meeting of the Psychometric Society, Leuven, Belgium.

Nandakumar, R. & Stout, W. F. (1991). Refinement of Stout's procedure for assessing latent trait dimensionality. (under review)

Nandakumar, R. (1991). Traditional dimensionality vs. essential dimensionality. Journal of Educational Measurement, 28, 1–19.

NAEP (1988). National Assessment of Educational Progress 1985–86 public–use data tapes. Version 2.0. Users Guide. Educational Testing Service.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. Journal of Educational Statistics, 4, 207–230.

Reckase, M. D., & McKinley, R. L. (1983). The definition of difficulty and discrimination for multidimensional item response theory models. Paper presented at the meeting of the American Educational Research Association, Montreal.

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

Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. Psychometrika, 49, 425–435.

Stout, W. F. (1987). A nonparametric approach for assessing latent trait unidimensionality. Psychometrika, 52, 589–617.

Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensional assessment and ability estimation. Psychometrika, 55, 293–326.

Traub, R. E. (1983). A priori considerations in choosing an item response model. In R. K. Hambleton (Ed.), Applications of item response theory. British Columbia: Educational Research Institute of British Columbia.

Tucker, L. R., Humphreys, L. G., & Roznowski, M. A. (1986) Comparative accuracy of five indices of dimensionality of binary types. Office of Naval Research Technical Report (N00014–84–K–0186).

Wise, S. L. (1981) A modified order–analysis for determining unidimensional item sets. Doctoral dissertation, University of Illinois, Urbana–Champaign.

Yen, W. M. (1985). Increasing item complexity: A possible cause of scale shrinkage for unidimensional item response theory. Psychometrika, 50, 399–410.

Zwick, R. (1987). Assessing the dimensionality of NAEP reading data. Journal of Educational Measurement, 24, 293–308.

Table 1
Description of Data Sets

| Name | J* | Traits | $\rho$*** | N | Number of items of each trait: Trait1 | Trait2 | Mixed**** |
|---|---|---|---|---|---|---|---|
| **Simulated test data** | | | | | | | |
| DATA1 | 2000 | 1 | | 25 | 25 | 0 | 0 |
| DATA2 | 2000 | 1 | | 40 | 40 | 0 | 0 |
| DATA3 | 2000 | 1 | | 50 | 50 | 0 | 0 |
| DATA4 | 2000 | 2 | .3 | 25 | 8 | 8 | 9 |
| DATA5 | 2000 | 2 | .7 | 25 | 8 | 8 | 9 |
| DATA6 | 2000 | 2 | .3 | 50 | 16 | 16 | 17 |
| DATA7 | 2000 | 2 | .7 | 50 | 16 | 16 | 17 |
| **Real test data** | | | | | | | |
| LIT | 2380 | 1 | | 30 | 30 | 0 | 0 |
| HIST | 2425 | 1 | | 31 | 31 | 0 | 0 |
| AR | 1984 | 1 | | 30 | 30 | 0 | 0 |
| GS | 1990 | 1 | | 25 | 25 | 0 | 0 |
| HSTLIT1 | 2380 | 2 | 0 | 36 | 31 | 5 | 0 |
| HSTLIT2 | 2380 | 2 | 0 | 41 | 31 | 10 | 0 |
| ARGS | 1984 | 2 | 0 | 40 | 30 | 10 | 0 |
| HSTGEO | 2425 | 2 | ? | 36 | 31 | 5 | 0 |

*J denotes the number of examinees

**$\rho$ denotes the correlation between traits

***N denotes the test length

****mixed items are a combination of both traits 1 and 2

Table 2
Results of Stout and H&R Analyses

| | Stout's Test $H_0: d_E=1$ | | | H&R Test $H_0: Cov(X_i, X_j \mid \sum_{k \neq i,j} X_k) \geq 0$ | | | |
|---|---|---|---|---|---|---|---|
| Name | T | p< | Decision based on Stout's procedure | No.of item pairs $\eta$ | No. of* pairs significant at level $\alpha$ | No.of pairs significant at level $\alpha/n$ | Decision based on Bonferoni bounds |
| **Simulated test data** | | | | | | | |
| DATA1 | -1.05 | .85 | accept $H_0$ | 300 | 1 | 0 | accept $H_0$ |
| DATA2 | -0.75 | .77 | accept | 780 | 3 | 0 | accept |
| DATA3 | -0.94 | .83 | accept | 1225 | 10 | 0 | accept |
| DATA4 | 7.19 | .000 | reject | 300 | 71 | 15 | reject |
| DATA5 | 3.62 | .000 | reject | 300 | 10 | 0 | accept |
| DATA6 | 10.13 | .000 | reject | 1225 | 206 | 1 | reject |
| DATA7 | 2.41 | .008 | reject | 1225 | 56 | 0 | accept |
| **Real test data** | | | | | | | |
| LIT | 1.70 | .045 | accept | 435 | 16 | 1 | undecided |
| HIST | -1.53 | .937 | accept | 465 | 6 | 0 | accept |
| AR | 1.18 | .118 | accept | 435 | 3 | 0 | accept |
| GS | -0.14 | .555 | accept | 300 | 6 | 0 | accept |
| HSTLIT1 | 2.75 | .003 | reject | 630 | 18 | 0 | accept |
| HSTLIT2 | 8.9 | .000 | reject | 820 | 83 | 0 | undecided |
| ARGS | 8.34 | .000 | reject | 780 | 37 | 0 | accept |
| HSTGEO | 6.83 | .000 | reject | 630 | 16 | 0 | accept |

*significant at .05 level

Table 3
Results of Linear and Nonlinear Factor Analysis
For Simulated Test data: Goodness of Fit Statistics

| | $\overline{rij^2}$* | $SD(rij^2)$ | $\overline{\lvert rij \rvert}$ | $SD(\lvert rij \rvert)$ | p<** |
|---|---|---|---|---|---|
| **RANDOM** | | | | | |
| Linear Factor Analysis | | | | | |
| 1 Factor | .0009 | .0308 | .0250 | .0182 | |
| 2 Factor | .0008 | .0283 | .0225 | .0169 | |
| 3 Factor | .0007 | .0246 | .0207 | .0160 | |
| 4 Factor | .0006 | .0245 | .0196 | .0147 | |
| **DATA1** | | | | | |
| Linear Factor Analysis | | | | | |
| 1 Factor | .0017 | .0412 | .0333 | .0242 | .006 |
| 2 Factor | .0013 | .0359 | .0286 | .0218 | .350 |
| 3 Factor | .0011 | .0332 | .0262 | .0204 | .610 |
| 4 Factor | .0009 | .0303 | .0236 | .0191 | .860 |
| Nonlinear Factor Analysis | | | | | |
| 1 Factor Quadratic | .0003 | .0185 | .0147 | .0113 | |
| $(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i)$ | | | | | |
| 1 Factor Cubic | .0003 | .0185 | .0147 | .0113 | |
| $(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}\theta^3+b_{i4}e_i)$ | | | | | |
| **DATA2** | | | | | |
| Linear Factor Analysis | | | | | |
| 1 Factor | .0110 | .1049 | .0982 | .0369 | .000 |
| 2 Factor | .0091 | .0954 | .0896 | .0327 | .000 |
| 3 Factor | .0070 | .0834 | .0774 | .0310 | .000 |
| 4 Factor | .0061 | .0779 | .0720 | .0278 | .000 |
| Nonlinear Factor Analysis | | | | | |
| 1 Factor Quadratic | .0003 | .0186 | .0148 | .0113 | |
| $(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i)$ | | | | | |
| 1 Factor Cubic | .0003 | .0185 | .0148 | .0113 | |
| $(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}\theta^3+b_{i4}e_i)$ | | | | | |
| **DATA3** | | | | | |
| Nonlinear Factor Analysis | | | | | |
| 1 Factor Quadratic | .0003 | .0186 | .0147 | .0115 | |
| $(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i)$ | | | | | |
| 1 Factor Cubic | .0003 | .0175 | .0138 | .0108 | |
| $(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}\theta^3+b_{i4}e_i)$ | | | | | |

Table 3 continued...

DATA4

| | | | | | |
|---|---|---|---|---|---|
| Linear Factor Analysis | | | | | |
| 1 Factor | .0203 | .1425 | .1108 | .0900 | .000 |
| 2 Factor | .0017 | .0412 | .0334 | .0240 | .000 |
| 3 Factor | .0012 | .0346 | .0276 | .0212 | .008 |
| Nonlinear Factor Analysis | | | | | |
| 1 Factor Quadratic ($Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i$) | .0021 | .0465 | .0523 | .0379 | |
| 2 Factor Quadratic ($Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i3}e_i$) | .0003 | .0171 | .0131 | .0109 | |

DATA5

| | | | | | |
|---|---|---|---|---|---|
| Linear Factor Analysis | | | | | |
| 1 Factor | .0047 | .0686 | .0556 | .0409 | .000 |
| 2 Factor | .0014 | .0374 | .0313 | .0218 | .011 |
| 3 Factor | .0012 | .0346 | .0289 | .0199 | .245 |
| 4 Factor | .0010 | .0316 | .0254 | .0181 | .600 |
| Nonlinear Factor Analysis | | | | | |
| 1 Factor Quadratic ($Y_i = b_{io}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i$) | .0009 | .0307 | .0246 | .0186 | |
| 2 Factor Quadratic ($Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i3}e_i$) | .0003 | .0174 | .0138 | .0107 | |

DATA6

| | | | | |
|---|---|---|---|---|
| Nonlinear Factor Analysis | | | | |
| 1 Factor Quadratic ($Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i$) | .0005 | .0242 | .0204 | .0172 |
| 2 Factor Quadratic ($Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i3}e_i$) | .0003 | .0182 | .0145 | .0111 |

DATA7

| | | | | |
|---|---|---|---|---|
| Nonlinear Factor Analysis | | | | |
| 1 Factor Quadratic ($Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i$) | .0005 | .0223 | .0176 | .0137 |
| 2 Factor Quadratic ($Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i3}e_i$) | .0003 | .0175 | .0140 | .0105 |

---

* $r_{ij}$ are the residual correlations

** p-value associated with the chi-square test of goodness of fit.

Table 4
Results of Linear and Nonlinear Factor Analysis
For Real Test data: Goodness of Fit Statistics

| | $\overline{rij^2}$* | SD($rij^2$) | $\overline{\lvert rij\rvert}$ | SD($\lvert rij\rvert$) | p<** |
|---|---|---|---|---|---|
| **LIT** | | | | | |
| Linear Factor Analysis | | | | | |
| 1 Factor | .0034 | .0584 | .0465 | .0354 | .000 |
| 2 Factor | .0028 | .0526 | .0428 | .0307 | .000 |
| 3 Factor | .0019 | .0439 | .0349 | .0267 | .000 |
| 4 Factor | .0015 | .0391 | .0310 | .0240 | .000 |
| Nonlinear Factor Analysis | | | | | |
| 1 Factor Quadratic | .0008 | .0278 | .0216 | .0176 | |
| ($Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i$) | | | | | |
| 2 Factor Quadratic | .0004 | .0207 | .0162 | .0130 | |
| ($Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i3}e_i$) | | | | | |
| **AR** | | | | | |
| Linear Factor Analysis | | | | | |
| 1 Factor | .0047 | .0683 | .0569 | .0378 | .000 |
| 2 Factor | .0032 | .0561 | .0468 | .0310 | .000 |
| 3 Factor | .0024 | .0489 | .0400 | .0281 | .000 |
| 4 Factor | .0020 | .0447 | .0362 | .0262 | .000 |
| Nonlinear Factor Analysis | | | | | |
| 1 Factor Quadratic | .0007 | .0265 | .0200 | .0174 | |
| ($Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i$) | | | | | |
| 2 Factor Quadratic | .0004 | .0190 | .0146 | .0122 | |
| ($Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i3}e_i$) | | | | | |
| **HSTLIT1** | | | | | |
| Linear Factor Analysis | | | | | |
| 1 Factor | .0053 | .0729 | .0574 | .0450 | .000 |
| 2 Factor | .0043 | .0657 | .0545 | .0368 | .000 |
| 3 Factor | .0033 | .0578 | .0457 | .0354 | .000 |
| 4 Factor | .0022 | .0469 | .0380 | .0279 | .000 |
| Nonlinear Factor Analysis | | | | | |
| 1 Factor Quadratic | .0009 | .0298 | .0213 | .0209 | |
| ($Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i$) | | | | | |
| 2 Factor Quadratic | .0004 | .0204 | .0157 | .0129 | |
| ($Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i23}\theta_1\theta_2+b_{i3}e_i$) | | | | | |

Table 4 continued...

| | | | | |
|---|---|---|---|---|
| **HSTLIT2** | | | | |
| Nonlinear Factor Analysis | | | | |
| 1 Factor Quadratic | .0013 | .0358 | .0228 | .0276 |
| $(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i)$ | | | | |
| 2 Factor Quadratic | .0003 | .0182 | .0140 | .0117 |
| $(Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i23}\theta_1\theta_2+b_{i3}e_i)$ | | | | |
| **ARGS** | | | | |
| Nonlinear Factor Analysis | | | | |
| 1 Factor Quadratic | .0011 | .0335 | .0239 | .0235 |
| $(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i)$ | | | | |
| 2 Factor Quadratic | .0003 | .0184 | .0143 | .0117 |
| $(Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i23}\theta_1\theta_2+b_{i3}e_i)$ | | | | |
| 3 Factor Quadratic | .0003 | .0175 | .0136 | .0111 |
| $(Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i31}\theta_3+$ $b_{i32}\theta_3^2+b_{i33}\theta_1\theta_2+b_{i34}\theta_1\theta_3+b_{i35}\theta_2\theta_3+b_{i4}e_i)$ | | | | |

* $r_{ij}$ are residual correlations

** p-value associated with the chi-square test of goodness of fit.