Sep 29, 1992

92 28 077

49185

40pg

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
10 August 1992

3. REPORT TYPE AND DATES COVERED
Technical: 1990-93

4. TITLE AND SUBTITLE
Assessing Dimensionality of a Set of Items - Comparison of Different Approaches

5. FUNDING NUMBERS
N00014-90-J-1940,

6. AUTHOR(S)
Ratna Nandakumar

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Statistics
University of Illinois
725 South Wright Street
Champaign, IL 61820

8. PERFORMING ORGANIZATION REPORT NUMBER
1992 - No. 3

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Cognitive Sciences Program
Office of Naval Research
800 N. Quincy
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
4421-548

11. SUPPLEMENTARY NOTES
To be published in Journal of Educational Measurement

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
See reverse

14. SUBJECT TERMS
See reverse

15. NUMBER OF PAGES
36

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
unclassified	unclassified	unclassified	UL

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: DIMTEST, Holland and Rosenbaum's approach, linear factor analysis, and nonlinear factor analysis. Each method is examined and compared with other methods on simulated data sets and on real data sets. Seven data sets, all with 2000 examinees, were generated: three unidimensional, and four two–dimensional data sets. Two levels of correlation between abilities were considered: ρ=.3 and ρ=.7. Eight different real data sets were used: four of them were expected to be unidimensional, and the other four were expected to be two–dimensional. Findings suggest that, while the linear factor analysis often overestimated the number of underlying dimensions, the other three methods correctly confirmed unidimensionality but differed in their ability to detect lack of unidimensionality. DIMTEST 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 was low.

Subject terms: DIMTEST, unidimensionality, essential dimensionality, non-linear factor analysis, item response theory.

It is well known that most item response theory (IRT) 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 has also been long argued that responses to test items are multiply determined (Humphreys, 1981, 1985, 1986; Hambleton & Swaminathan, 1985, chap. 2; 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 responses to few items will be referred to as minor abilities. Given that item responses are multiply determined, it is intuitively clear that, in order to satisfy the assumption of unidimensionality, it is required that a given test measure a single dominant ability. A number of simulation studies have demonstrated that a dominant ability can be recovered well, using computer programs such as LOGIST, 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[1].

Stout (1987, 1990) provided a mathematically rigorous definition of dominant dimensionality referred to as **essential dimensionality** and provided a statistical test (DIMTEST) to assess whether a set of items met the requirement for essential unidimensionality. Junker (1988, 1991) further explored essential dimensionality for dichotomous and polytomous items and established consistency results for the maximum likelihood ability estimates of θ under essential unidimensionality. Essential dimensionality is the total number of abilities required to satisfy the assumption of **essential independence**.

An item pool is said to be essentially independent (*EI*) with respect to the latent variable vector $\underline{\Theta}$ if, for a given subset of items, the average absolute conditional (on $\underline{\Theta}$) covariances of responses to item pairs approaches zero as the length of the subset increases. When conditional covariances based on only one dominant ability meet the assumption of essential independence, the response data is said to be essentially unidimensional (d_E=1). In contrast, the assumption of local independence requires that the conditional covariances be zero for responses to any item pair, and the number of abilities required to those conditional covariances is the dimensionality. According to this definition of dimensionality, all major and minor abilities influencing item responses have to be considered when assessing the local independence assumption; whereas, according to the essential dimensionality, it is sufficient to consider only the influence of dominant abilities. Hence, essential independence and essential dimensionality are weaker forms of local independence and traditional dimensionality respectively.

Stout's definition of essential dimensionality is conceptually based on an infinite item pool. An infinite item pool can be conceptualized in two ways: 1. as a consequence of continuing the test construction process beyond the *N* items of the test being studied where the *N* items become a subset of the item pool; 2. as a consequence of a sequence of finite tests where each finite test is optimally constructed. For example, a 20–item test is constructed with the knowledge that the test is going to be only 20 items long and that it is not necessarily a subset of an optimal 40–item test. In this way, an item pool is a collection of optimal finite test length tests (for details see Junker, 1991; Junker & Stout, 1991).

In assessing essential unidimensionality of given item responses, DIMTEST assesses the likelihood that the given set of item responses come from an essentially unidimensional item pool. That is, DIMTEST assesses whether or not the model generating the given item responses is close to the EI, d_E= 1 model. 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 these minor abilities can be ignored when assessing essential unidimensionality.

Historically speaking, linear factor analysis has been used to assess the dimensionality of the latent space underlying the responses to 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 levels of items and guessing levels of multiple–choice items can each play a major role in affecting the factor structure of item responses (for details see Carroll, 1945; Hulin, Drasgow, & Parsons, 1983, chap. 8; 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, p. 255); residual analysis (Hambleton & Swaminathan, 1985, p. 163); Bock's full information factor analysis (Bock, Gibbons, & Muraki, 1985); Holland and Rosenbaum's test of unidimensionality, monotonicity, and conditional independence (Rosenbaum, 1984; Holland & Rosenbaum, 1986); Roznowski, Tucker, and Humphreys' procedures (1991); and Stout's unidimensionality procedure DIMTEST (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. 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: DIMTEST, Holland and Rosenbaum's procedure, nonlinear factor analysis, and linear factor analysis. Linear factor analysis was used, because of its historical importance, as a benchmark to compare other procedures. Several sets of unidimensional and multidimensional test data were simulated and used to study the performance of all four procedures for assessing dimensionality. The same procedures were 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, and 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, 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 is clear for dichotomous data that the relationship between the performance and the underlying latent variable is not linear. Hence, applying factor analysis to phi or tetrachoric correlations of binary item responses produces difficulty factors (Hulin, Drasgow, & Parsons, 1983, chap. 8). Second, in computing tetrachoric correlations, the cell entries of the fourfold table for a pair of dichotomous items sometimes equal zero, making it difficult to determine an appropriate value for the correlation. Third, determination of the number of significant factors could be problematic.

In this study the statistical package LISCOMP was used to perform exploratory linear factor analysis using tetrachoric correlations. Three different approaches were 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 & Montanelli, 1975), the eigenvalues of the given correlation matrix are compared with the eigenvalues of random data. The random data consist of binary responses generated 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 was the chi–square test of goodness of fit from LISCOMP. The third method involves comparisons 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) and McDonald and Ahlawat (1974) have demonstrated that applying linear factor analysis to unidimensional binary data yields "nonlinear factors" rather than "difficulty factors." Nonlinear factors account for nonlinear relationships among the variables by using higher order polynomials in the factor model (for example, quadratic and cubic terms). McDonald developed the method of nonlinear factor analysis (NLFA) to account for the nonlinearity of the data as an improvement over linear factor analysis. 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:

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

where Y_i denotes the examinee's score on item i, θ_1 and θ_2 denote latent traits, b_{ijk} denotes the factor loading of the i–th item on the j–th common factor for the k–th degree element in the polynomial; u_i denotes the unique factor and d_i denotes the unique factor loading for item *i*. 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, developed by Etazadi–Amoli and McDonald (1983), was used. The fit of the model is studied just as in the case of the linear factor analyses, by 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 from NOFA.

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 item characteristic curves 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. The test of this relationship can be specified as

$$\mathrm{H}_0\colon Cov\,(X_i, X_j \mid \sum_{i,j \neq k} X_k) \geq 0 \quad vs. \quad H_1\colon Cov\,(X_i, X_j \mid \sum_{i,j \neq k} X_k) < 0$$

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, 1959) 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 scores on the remaining items. The cell values of a 2×2 table for item pair i and j for examinees with total score k (k=1,2,...K) on the remaining items can be denoted as the following: the number of examinees who got both item i and item j correct (n_{11k}), the number of examinees who got both item i and item j incorrect (n_{00k}), the number of examinees who got item i correct and item j incorrect (n_{10k}), and the number of examinees who got item i incorrect and item j correct (n_{01k}). The M–H statistic is then given by

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

where $n_{11+} = \sum_{k=1}^{K} n_{11k}$, and $E(n_{11+})$ and $V(n_{11+})$ are the expectation and the variance of n_{11+} given by

$$E(n_{11+}) = \sum_{k=1}^{K} \frac{n_{1+k}\, n_{+1k}}{n_{++k}} \tag{2}$$

and

$$V(n_{11+}) = \sum \frac{n_{1+k}\, n_{0+k}\, n_{+1k}\, n_{+0k}}{n_{++k}^{2}\,(n_{++k}-1)} \tag{3}$$

The plus subscript in Equations 2 and 3 denotes the summation over that subscript. The computed Z–value is compared 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 remaining items and are thus inconsistent with the unidimensional model. In this manner, the M–H statistic is computed for all $N(N-1)/2$

pairs of items, where N is the total number of items in a test. 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 α, the simultaneous inference for all item pairs can be based on Bonferroni bounds (Holland & Rosenbaum, 1986, Junker, 1990, and Zwick, 1987). According to Bonferroni bounds, one would accept H_0 if the number of rejections at level α is around $t\alpha$, where t is the number of tests performed, which is equal to $N(N-1)/2$; one would reject H_0 if at least one test is rejected at level α/t.

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 assessing unidimensionality of NAEP reading data.

DIMTEST

Stout (1987) developed DIMTEST to test the hypothesis of essential unidimensionality: the existence of one dominant dimension. Nandakumar and Stout (in press) further modified and improved the performance of DIMTEST. The improvements have lead to the following: a robust procedure against presence of guessing in item responses; a better control of the observed level of significance, and greater power; and automation of the size of assessment subtests, as described below. The hypothesis to test unidimensionality can be stated as

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

where d_E denotes the essential dimensionality of the item pool of which the given test items are a part.

In order to apply DIMTEST, it is assumed that a group of J examinees take an N–item test. Each examinee produces a vector of responses of 1s and 0s with 1 denoting a correct response and 0 denoting an incorrect response. It is also assumed that essential independence with respect to some dominant ability Θ holds and that the item response functions are monotone with respect to the same dominant ability Θ. DIMTEST has several steps. These are briefly described here (for details see Stout, 1987; Nandakumar and Stout, in press).

Step 1: The N items of the test are split into three subtests: AT1, AT2, and PT. First, AT1 items are selected so that these items all measure the same dominant ability. This can be achieved either through factor analysis (FA) or through expert *opinion* (EO). If FA method is chosen, M items with highest loadings on the second factor (before rotation) are selected. In this case, the program automatically determines the size M of AT1 as a function of the test length and the sample size. If EO is sought, on the other hand, it is recommended that, at most, one–quarter of the total items should be selected that tap the same ability. After selecting items of AT1, items of AT2 are selected, also of the same size M, so that items of AT1 and AT2 have the same difficulty distribution (for details see Stout, 1987). The remaining items ($n=N-2M$) form the partition subtest PT. In the present study, FA is chosen to select AT1 items. For examples where EO is used to select AT1 items, see Nandakumar (in press).

When FA is used to select AT1 items, the given sample of J examinee responses are partitioned into two groups. One group of examinee responses (500 examinees recommended) is used for exploratory factor analysis to select AT1 and AT2 items, and the other group of examinee responses is used to compute the Stout's statistic T.

Step 2: The second group of examinees (if the first group of examinees is used for FA) are partitioned into K subgroups based on their PT score. That is, all examinees obtaining the same total score on PT are assigned to the same subgroup k (k=1,2,...K).

Step 3: Within each subgroup k, examinee responses to subtest items AT1 and AT2 are used to compute the unidimensional statistic T given by

$$T = (T_1 - T_2)/\sqrt{2}, \tag{4}$$

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 ATi. The σ_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

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

and

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

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

DIMTEST assesses the degree of closeness of an essentially unidimensional model to the model generating the observed data. This is done by splitting the test items into three subtests—AT1, AT2, and PT—as described above. When the model underlying the test item responses is close to essentially unidimensional, items of AT1, AT2, and PT would all be of the same dominant dimension; therefore, the value of the statistic T computed based on AT1, AT2 would be "small," leading to the tenability of H_0. When the model underlying the test responses is not essentially unidimensional, however, items of AT1 would be dimensionally different from items of AT2 and PT and the value of the statistic T will be "large" leading to the rejection of H_0.

DIMTEST has been found to discriminate between unidimensional and two–dimensional tests for a variety of simulated test data when the correlation between abilities is as high as .7 (Stout, 1987; Nandakumar & Stout, in press). Nandakumar (1991) has shown the usefulness of DIMTEST to assess essential unidimensionality in the possible presence of several minor abilities. The findings indicate that essential unidimensionality is established when each of the minor abilities influence relatively few items, or, if minor abilities are influencing many items, the strength of the influence of the minor abilities is low. As the strength of the minor abilities increases, the approximation to an essentially unidimensional model degenerates, inflating the type–I error of the test of hypothesis of essential unidimensionality. Nandakumar (in press) has further replicated these findings on a wide variety of real test data. This study also demonstrates the sensitivity of DIMTEST to major and minor abilities influencing item responses.

Description of Test Data

The Simulated Test Data

Seven data sets, DATA1–DATA7, were generated. Of the seven, three data sets, DATA1–DATA3, are strictly unidimensional, consisting of 25, 40, and 50 items, respectively. The other four data sets, DATA4–DATA7, are two–dimensional with length N=25 and correlation between abilities ρ=.3, N=25 and ρ=.7, N=50 and ρ=.3, and N=50 and ρ=.7, respectively. All 7 data sets have 2000 examinees. These data set characteristics are summarized in Table 1.

Table 1 about here

The unidimensional data sets were generated using the three–parameter logistic model given by

$$P_i(\theta) = c_i + \frac{1-c_i}{1+exp\{-1.7[a_i(\theta-b_i)]\}}. \tag{5}$$

The abilities (θ) were 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) were 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_is, 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 uniform distribution. If the computed probability, $P_i(\theta)$, was 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) given by

$$P_i(\theta_1,\theta_2) = c_i + \frac{1-c_i}{1+exp\{-1.7[a_{1i}(\theta_1-b_{1i})+a_{2i}(\theta_2-b_{2i})]\}}. \tag{6}$$

The abilities $\underline{\theta} = (\theta_1,\theta_2)$ were sampled from a bivariate normal distribution with both means zero and both variances one. Two levels of correlation coefficients between the abilities were used: .3 and .7. The guessing level was taken to be .20 for all tests. The discrimination parameters (a_{i1}, a_{i2}) for each item were independently generated as follows:

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

where μ and σ 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:

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

where μ and σ 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 ρ=.3, the means and standard deviations of a_is and b_is of item parameters used for DATA1 were used. The item responses (0,1) were generated exactly as described for unidimensional case by using $P_i(\underline{\theta})$ of (6).

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 the 1986 US History (HIST) and Literature (LIT) for grade 11/age 17 were obtained from Educational Testing Service. 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). For all data sets, examinees who missed one or more items were deleted from the analyses. Test sizes and sample sizes for all real tests are given in bottom half of Table 1. Since all four test data were assessed as unidimensional by the methods employed in this article (details are provided in Results section), they were combined to form two–dimensional tests. 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. Similarly HSTLIT2 was formed by combining the responses of 31 items of HIST with the responses of 10 items of LIT, and the test data GS was formed by combining responses of 30 items of AR with the responses of 10 items of GS. The two-dimensional test HSTGEO contains 31 history items spanning US history from the 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. But it was shown using DIMTEST that the 5 map items formed a separate dimension significantly different from history items (Nandakumar, in press). Hence the data on these 5 map items were removed from the history test to form HIST with 31 items, and the original history data were treated as a natural two-dimensional test.

Results

The results of DIMTEST 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

DIMTEST and H&R Procedure

The results of DIMTEST and the H&R approach for simulated data are presented at the top of Table 2. For all data sets, the significance levels associated with DIMTEST indicate that DIMTEST is able to correctly confirm unidimensionality and detect lack of unidimensionality for both correlation (between abilities) levels ρ=.3 and ρ=.7. For example, all three unidimensional data sets, DATA1-DATA3, have small T-values and large significant values, implying the acceptance of the null hypothesis of essential unidimensionality (here the data were simulated as strictly unidimensional). Two-dimensional data, DATA4-DATA7, on the other hand, have large T-values, strongly rejecting the null hypothesis of essential unidimensionality.

Table 2 about here

The results of the H&R approach indicate that for unidimensional tests, the number of significant negative partial associations at level α (α=.05) are far below the expected number ($t\alpha$), strongly confirming the unidimensional nature of these data sets. Among the two–dimensional data sets, DATA4 and DATA6 (ρ=.3) were correctly assessed as multidimensional. For these data, the number of significant negative partial associations at level α were beyond $t\alpha$ level, and the number of significant negative partial associations beyond level α/t were 15 and 1, respectively, identifying them as multidimensional. The test data DATA5 and DATA7 (ρ=.7), on the other hand, were assessed as unidimensional. For DATA5 and DATA7, the number of significant negative partial associations at level α were within $t\alpha$ level, and the number of significant negative partial associations beyond level α/t was zero, 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 can not handle more than about 40 variables. For these reasons, not all data sets were included in the linear factor analyses, but all data sets were included in the nonlinear factor analyses. The results of linear and nonlinear factor analyses are presented in Table 3.

Table 3 about here

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 significance levels associated with a chi–square test of goodness of fit, in Table 3, a two–factor model fits DATA1, a four–factor model fits DATA2 and DATA4, and a three–factor model fits DATA5. Similar chi–square values are not available for nonlinear models.

The goodness of fit statistics—the means and standard deviations of squared residuals and absolute residuals—are reported for all data sets 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 set was used to compare the goodness of fit statistics. Comparing goodness of fit statistics of RANDOM with DATA1, it appears that both one–factor quadratic and one–factor cubic models fit as well as the four–factor linear model. However, since the differences in the magnitude of residuals among models are small, one could argue that four–factor linear and one–factor quadratic or cubic models are over fit and that one should go with a more parsimonious model. Observance of the significance values of the chi–square test of goodness of fit indicates that the two–factor model fits the data. If one strictly applies the criterion of using random data residuals as a guide to determine the number of factors, however, a one–factor model with a quadratic term seems to be the right choice. Similar observations can be made for DATA2. Comparing goodness of fit statistics for linear and nonlinear factor analysis, it can be seen that for DATA4 and DATA5, the two–factor quadratic model fits better than the three–factor linear model, confirming the two–dimensional nature of data. Here again one could argue, based on the absolute residuals, that the differences in the residuals are small and that the quadratic models or three–factor and four–factor linear models are an over fit.

The significant values associated with the chi–square test indicate overestimation of factors for DATA4. As expected, the means and the standard deviations of squared residuals and absolute residuals are much larger for DATA4 (ρ=.3) than for DATA5 (ρ=.7), reflecting more deviation from unidimensionality for DATA4. For DATA5, the goodness of fit analyses support a one–factor quadratic model. Likewise the two–factor quadratic model fits DATA6, and one–factor quadratic model fits DATA7.

In summary, there are many criteria that can be used to assess dimensionality by linear factor analysis approach. The different criteria may give rise to different conclusions regarding the dimensionality of the data set in consideration. In the present study it is shown that the significant values associated with the chi–square test overestimated the number of factors in most cases. Parallel analyses correctly identified the dimensionality in some cases. Nonlinear factor analyses exhibited a better fit than the linear factor analyses. DIMTEST and H&R procedures were excellent in confirming unidimensionality. DIMTEST demonstrated greater power in detecting multidimensionality for correlations between abilities as high as .7. H&R and nonlinear factor analysis methods demonstrated good power provided the correlation between abilities was low (ρ=.3).

The Real Test Data

DIMTEST and H&R Procedure

The results of DIMTEST and H&R for real data sets are presented at the bottom of Table 2. For data sets LIT, HIST, AR, and GS, the T–values associated with DIMTEST indicate that these data can be approximated by an essentially unidimensional model. The results of H&R approach for these data are also consistent with DIMTEST results in that the number of significant negative partial associations, for each one of the tests, is less than the nominal level $t\alpha$. While both approaches strongly support that HIST, AR, and GS are essentially unidimensional, the decision is not clear for LIT because there is one negative partial association that is significant beyond level α/t, and the T–value of DIMTEST is in the border line region, indicating presence of violations to the unidimensionality hypothesis.

For two–dimensional data HSTLIT1, HSTLIT2, ARGS, and HSTGEO, the T–values associated with DIMTEST strongly indicate the multidimensional nature of these data. Relatively large T–values associated with ARGS and HSTGEO indicate that abilities within these tests are more orthogonal than abilities in HSTLIT1 and HSTLIT2. The results based on H&R approach, however, indicate that all four data sets are unidimensional. For each one of the two–dimensional data sets, the number of significant negative partial associations is well below the nominal level $t\alpha$, and none of the partial associations are significant beyond level α/t. Even with a liberal $\alpha = .10$, the number of negative partial associations did not rise above the nominal level for any of the tests. These results suggest that the H&R approach lacks power.

On further examination of H&R results, it was found that the M–H Z–values for many of the item pairs, where items were supposed to be measuring different traits, did not reach significance level. One explanation for this could be that for these item pairs, the conditional score (ΣX_k), on the basis of which the examinees are classified into different groups, may be contaminated with items tapping different abilities. This could be especially true for HSTLIT2 and ARGS where one quarter of the test items are from the 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. A proper conditional score may 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 real data sets are reported in Table 4. The results are consistent with the simulated test data in that for all cases nonlinear factor models fit better than linear factor models. According to the chi–square test of goodness of fit, the four–factor model was best fitting for all data sets where linear factor analysis was performed. Based on goodness of fit statistics, a one–factor quadratic model fits LIT, AR, and HSTLIT1 better than three– or four–factor linear models. Since a one–factor quadratic model fits as well as a two–factor quadratic model, a more parsimonious model is strongly recommended in these cases. For HSTLIT2 and ARGS, again it appears that a one–factor quadratic model is appropriate. If chi–square statistics were available along with the goodness of fit statistics for nonlinear factor analyses, it would have aided in the interpretation.

Table 4 about here

In summary, for real data sets, the results are somewhat consistent with simulated data sets. For data sets assessed as unidimensional by DIMTEST and H&R, the chi–square tests based on the linear factor analysis indicated a four–factor model for the same data. Although we do not know the true dimensionality of real data, these results suggest that linear factor analysis is overestimating the underlying dimensionality. Whereas, the other three methodologies were excellent in identifying essential unidimensionality but differed in identifying lack of unidimensionality. DIMTEST demonstrated greater power than either the H&R or the nonlinear factor analysis methods. It appears that with the appropriate conditional score the power of the H&R approach could be improved, and with some type of fit statistics and the associated significance levels, the power of nonlinear factor analysis could be improved.

Discussion

Based on this limited study, findings demonstrate that the linear factor analysis approach to assessing essential unidimensionality is not satisfactory. This finding is consistent with the previous research and theory (see for example, Hambleton & Rovinelli, 1986; Hattie, 1984). In contrast to linear factor analysis, DIMTEST, H&R, and nonlinear factor analysis were each shown to be promising methodologies to assess dimensionality.

In this study, all three methodologies exhibited sensitivity to discriminate between one– and two–dimensional test data. For simulated unidimensional test data, all three procedures were able to confirm unidimensionality. For the real data, all three procedures were consistent in identifying unidimensionality of HIST, AR, and GS. For two–dimensional test data, however, the three procedures differed in their ability to detect the lack of unidimensionality. DIMTEST rejected the null hypothesis of essential unidimensionality for all two–dimensional tests: both real and simulated. The H&R approach confirmed the lack of unidimensionality for two–dimensional simulated tests, provided the correlation between abilities was low (ρ=.3). For simulated test data with high correlation between abilities (ρ=.7), the H&R approach was unable to detect multidimensionality. Also, for all two–dimensional real test data, the H&R approach was unable to detect multidimensionality.

The performance of the nonlinear factor analysis methodology was similar to the H&R procedure for two–dimensional data sets. For simulated test data with ρ=.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 ρ=.7, however, the difference in 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 was far below the expected five percent level, making it a very conservative test. Consequently, it did not exhibit high power. The reason one observes fewer than the nominal level of negative partial associations is that the conditional score used in computing the covariances is not perfectly correlated with the latent variable (Zwick, 1987). According to the theorems proved by Holland and Rosenbaum (1986), the conditional score used to compute the covariances can be <u>any</u> 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 determining the appropriate number of polynomial terms to retain in the model. This problem suggests that some type of adequacy of fit statistics with associated sampling distribution would be necessary to aid in assessing the fit of nonlinear models.

In terms of assessing the degree of multidimensionality, both the DIMTEST and nonlinear factor analysis approaches can be useful. The T–values associated with DIMTEST 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 considered as essentially unidimensional data sets, but the associated T–values are –1.53 and 1.18 respectively. By contrast, for a two–dimensional data set HSTLIT2, T=2.03. The difference in the T–values mirrors 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 reflects the degree of multidimensionality.

In the present study, the test length is more than 25 items, and the sample sizes are around 2000 examinees. It is not known if the results would hold up for small test lengths and sample sizes. De Champlain and Gessaroli (1991) have shown that DIMTEST loses power when both the test length and the sample size are small (for example, N=25 and J=500). Their results show support for the use of incremental fit index (IFI) using the nonlinear factor analysis program, NOHARM II, to assess dimensionality in cases of smaller test lengths and sample sizes. Ben–Simon and Cohen (1990) have found that the test length and the sample size had a marked effect on the M–H Z–statistic in the detection of multidimensionality. In their study they tried test lengths of 20, 30, 40, and 50 and sample sizes of 1000, 2000, 3000, and 4000. They found that larger samples and larger tests facilitated the detection of multidimensionality. They urge a cautious interpretation of M–H test results in light of test lengths and sample sizes.

Just as linear and nonlinear methodologies share the same philosophical theory, DIMTEST 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, DIMTEST 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. DIMTEST is specifically designed to assess unidimensionality and thus 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, DIMTEST is most efficient. The computational time involved for other procedures is significantly more. For example, for a 25 item test with 2000 examinees, DIMTEST 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, DIMTEST uses 8 seconds, H&R approach uses 106 seconds, and nonlinear factor analysis uses 191 seconds. As the test length increases, the H&R approach requires disproportionately more time, and the same is true for the nonlinear factor analysis as test length increases and/or the model gets more complex.

Notes

[1]The reader is reminded that testing for unidimensionality is not synonymous to testing for model–data fit. If a unidimensional model is to be applied to the data, testing for unidimensionality is the first step. If item responses are essentially unidimensional, then as a second step, one can test for model–data fit, such as, one–parameter logistic, two–parameter logistic, etc.

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., & Muraki, 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.

De Champlain, A., & Gessaroli, M. E. (1991). Assessing test dimensionality using an index based on nonlinear factor analysis. Paper presented at the annual AERA meeting, Chicago.

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

Etazadi–Amoli, J., & McDonald, R. P. (1983). A second generation nonlinear factor analysis. Psychometrika, 48, 315–342.

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. (1988). Statistical aspects of a new latent trait theory, Unpublished doctoral dissertation, University of Illinois at Urbana–Champaign.

Junker, B. (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.

Junker, B. (1991). Essential independence and likelihood–based ability estimation for polytomous items. Psychometrika, 56, 255–278.

Junker, B., & Stout, W.F. (1991). Robustness of ability estimation when multiple traits are present with one trait dominant. Paper presented at the International Symposium on Modern Theories in Measurement: Problems and Issues. Montebello, Quebec.

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.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the retrospective study of disease. Journal of the National Cancer Institute, 22, 719–748.

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

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.

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

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

Nandakumar, R. (in press). Assessing essential dimensionality of real data. Applied Psychological Measurement.

Nandakumar, R., & Stout, W. F. (in press). Refinement of Stout's procedure for assessing latent trait dimensionality. Journal of Educational Statistics.

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

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.

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

Roznowski, M. A., Tucker, L. R., & Humphreys, L. G. (1991). Three approaches to determining the dimensionality of binary data. Applied Psychological Measurement, 15, 109–128.

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.

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[1]	Traits	ρ[2]	N[3]	Number of items of each trait: Trait1	Trait2	Mixed[4]
Simulated data sets							
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 data sets							
LIT	2439	1		30	30	0	0
HIST	2428	1		31	31	0	0
AR	1984	1		30	30	0	0
GS	1990	1		25	25	0	0
HSTLIT1	2428	2	–	36	31	5	0
HSTLIT2	2428	2	–	41	31	10	0
ARGS	1853	2	–	40	30	10	0
HSTGEO	2440	2	–	36	31	5	0

[1] J denotes the number of examinees

[2] ρ denotes the correlation between traits

[3] N denotes the test length

[4] mixed items are a combination of both traits 1 and 2

Table 2
Results of DIMTEST and H&R Analyses

	DIMTEST			H&R Test			
	H.: $d_E=1$			H.: $cov(X_i,X_j \mid \sum_{k\neq i,j} X_k)\geq 0$			
Name	T	$p<$	Decision based on DIMTEST	No.of item pairs t	No. of* pairs significant at level α	No.of pairs significant at level α/t	Decision based on Bonferoni bounds
Simulated test data							
DATA1	–1.05	.85	accept H.	300	1	0	accept H.
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	3.01	.036	reject	630	17	0	accept
HSTLIT2	2.03	.021	reject	820	18	0	accept
ARGS	6.15	.000	reject	780	4	0	accept
HSTGEO	6.19	.000	reject	630	17	0	accept

*significant at .05 level

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

	$\overline{r_{ij}^2}$*	SD(r_{ij})	$\overline{\lvert r_{ij}\rvert}$	SD($\lvert r_{ij}\rvert$)	$p<$**
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	
($Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+d_iu_i$)					
1 Factor Cubic	.0003	.0185	.0147	.0113	
($Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}\theta^3+d_iu_i$)					
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	.0115	
($Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+d_iu_i$)					
1 Factor Cubic	.0003	.0185	.0148	.0113	
($Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}\theta^3+d_iu_i$)					
DATA3					
Nonlinear Factor Analysis					
1 Factor Quadratic	.0003	.0186	.0147	.0115	
($Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+d_iu_i$)					
1 Factor Cubic	.0003	.0175	.0138	.0108	
($Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}\theta^3+d_iu_i$)					

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	.0021	.0465	.0523	.0379	
$(Y_i = b_{i0} + b_{i1}\theta + b_{i2}\theta^2 + d_i u_i)$					
2 Factor Quadratic	.0003	.0171	.0131	.0109	
$(Y_i = b_{i0} + b_{i11}\theta_1 + b_{i12}\theta_1^2 + b_{i21}\theta_2 + b_{i22}\theta_2^2 + d_i u_i)$					
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	.0009	.0307	.0246	.0186	
$(Y_i = b_{io} + b_{i1}\theta + b_{i2}\theta^2 + d_i u_i)$					
2 Factor Quadratic	.0003	.0174	.0138	.0107	
$(Y_i = b_{i0} + b_{i11}\theta_1 + b_{i12}\theta_1^2 + b_{i21}\theta_2 + b_{i22}\theta_2^2 + d_i u_i)$					
DATA6					
Nonlinear Factor Analysis					
1 Factor Quadratic	.0005	.0242	.0204	.0172	
$(Y_i = b_{i0} + b_{i1}\theta + b_{i2}\theta^2 + d_i u_i)$					
2 Factor Quadratic	.0003	.0182	.0145	.0111	
$(Y_i = b_{i0} + b_{i11}\theta_1 + b_{i12}\theta_1^2 + b_{i21}\theta_2 + b_{i22}\theta_2^2 + d_i u_i)$					
DATA7					
Nonlinear Factor Analysis					
1 Factor Quadratic	.0005	.0223	.0176	.0137	
$(Y_i = b_{i0} + b_{i1}\theta + b_{i2}\theta^2 + d_i u_i)$					
2 Factor Quadratic	.0003	.0175	.0140	.0105	
$(Y_i = b_{i0} + b_{i11}\theta_1 + b_{i12}\theta_1^2 + b_{i21}\theta_2 + b_{i22}\theta_2^2 + d_i u_i)$					

* 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

	$\overline{r_{ij}^2}$ *	SD(r_{ij})	$\overline{\lvert r_{ij}\rvert}$	SD($\lvert r_{ij}\rvert$)	$p<$ **
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	
$(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+d_iu_i)$					
2 Factor Quadratic	.0004	.0207	.0162	.0130	
$(Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+d_iu_i)$					
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	
$(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+d_iu_i)$					
2 Factor Quadratic	.0004	.0190	.0146	.0122	
$(Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+d_iu_i)$					
HSTLIT1					
Nonlinear Factor Analysis					
1 Factor Quadratic	.0008	.0275	.0213	.0175	
$(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+d_iu_i)$					
2 Factor Quadratic	.0003	.0185	.0143	.0118	
$(Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i23}\theta_1\theta_2+d_iu_i)$					

Table 4 continued...

HSTLIT2				
Nonlinear Factor Analysis				
1 Factor Quadratic	.0006	.0236	.0181	.0152
$(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+d_iu_i)$				
2 Factor Quadratic	.0004	.0191	.0150	.0119
$(Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i23}\theta_1\theta_2+d_iu_i)$				
ARGS				
Nonlinear Factor Analysis				
1 Factor Quadratic	.0021	.0462	.0268	.0376
$(Y_i = b_{i0}+b_{i1}\theta+b_{i2}\theta^2+b_{i3}e_i)$				
2 Factor Quadratic	.0004	.0192	.0003	.0123
$(Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i23}\theta_1\theta_2+d_iu_i)$				
3 Factor Quadratic	.0004	.0175	.0003	.0111
$(Y_i = b_{i0}+b_{i11}\theta_1+b_{i12}\theta_1^2+b_{i21}\theta_2+b_{i22}\theta_2^2+b_{i31}\theta_3+$ $b_{i32}\theta_3^2+b_{i33}\theta_1\theta_2+b_{i34}\theta_1\theta_3+b_{i35}\theta_2\theta_3+d_iu_i)$				

* r_{ij} are residual correlations

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

Dr. Terry Ackerman
Educational Psychology
260C Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Terry Allard
Code 1142CS
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217-5000

Dr. Nancy Allen
Educational Testing Service
Princeton, NJ 08541

Dr. Gregory Anrig
Educational Testing Service
Princeton, NJ 08541

Dr. Phipps Arabie
Graduate School of Management
Rutgers University
92 New Street
Newark, NJ 07102-1895

Dr. Isaac I. Bejar
Law School Admissions
Services
Box 40
Newtown, PA 18940-0040

Dr. William O. Berry
Director of Life and
Environmental Sciences
AFOSR/NL, N1, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Thomas G. Bever
Department of Psychology
University of Rochester
River Station
Rochester, NY 14627

Dr. Menucha Birenbaum
Educational Testing
Service
Princeton, NJ 08541

Dr. Bruce Bloxom
Defense Manpower Data Center
99 Pacific St.
Suite 155A
Monterey, CA 93943-3231

Dr. Gwyneth Boodoo
Educational Testing Service
Princeton, NJ 08541

Dr. Richard L. Branch
HQ, USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago, IL 60064

Dr. Robert Brennan
American College Testing
Programs
P. O. Box 168
Iowa City, IA 52243

Dr. David V. Budescu
Department of Psychology
University of Haifa
Mount Carmel, Haifa 31999
ISRAEL

Dr. Gregory Candell
CTB/MacMillan/McGraw-Hill
2500 Garden Road
Monterey, CA 93940

Dr. Paul R. Chatelier
Perceptronics
1911 North Ft. Myer Dr.
Suite 800
Arlington, VA 22209

Dr. Susan Chipman
Cognitive Science Program
Office of Naval Research
800 North Quincy St.
Arlington, VA 22217-5000

Dr. Raymond E. Christal
UES LAMP Science Advisor
AL/HRMIL
Brooks AFB, TX 78235

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
Los Angeles, CA 90089-1061

Director
Life Sciences, Code 1142
Office of Naval Research
Arlington, VA 22217-5000

Commanding Officer
Naval Research Laboratory
Code 4827
Washington, DC 20375-5000

Dr. John M. Cornwell
Department of Psychology
I/O Psychology Program
Tulane University
New Orleans, LA 70118

Dr. William Crano
Department of Psychology
Texas A&M University
College Station, TX 77843

Dr. Linda Curran
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn, VA 22209

Dr. Timothy Davey
American College Testing Program
P.O. Box 168
Iowa City, IA 52243

Dr. Charles E. Davis
Educational Testing Service
Mail Stop 22-T
Princeton, NJ 08541

Dr. Ralph J. DeAyala
Measurement, Statistics,
and Evaluation
Benjamin Bldg., Rm. 1230F
University of Maryland
College Park, MD 20742

Dr. Sharon Derry
Florida State University
Department of Psychology
Tallahassee, FL 32306

Hei-Ki Dong
Bellcore
6 Corporate Pl.
RM: PYA-1K207
P.O. Box 1320
Piscataway, NJ 08855-1320

Dr. Neil Dorans
Educational Testing Service
Princeton, NJ 08541

Dr. Fritz Drasgow
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
(2 Copies)

Dr. Richard Duran
Graduate School of Education
University of California
Santa Barbara, CA 93106

Dr. Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence, KS 66045

Dr. George Engelhard, Jr.
Division of Educational Studies
Emory University
210 Fishburne Bldg.
Atlanta, GA 30322

ERIC Facility-Acquisitions
2440 Research Blvd., Suite 550
Rockville, MD 20850-3238

Dr. Marshall J. Farr
Farr-Sight Co.
2520 North Vernon Street
Arlington, VA 22207

Dr. Leonard Feldt
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Richard L. Ferguson
American College Testing
P.O. Box 168
Iowa City, IA 52243

Dr. Gerhard Fischer
Liebiggasse 5
A 1010 Vienna
AUSTRIA

Dr. Myron Fischl
U.S. Army Headquarters
DAPE-HR
The Pentagon
Washington, DC 20310-0300

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152-6800

Chair, Department of
Computer Science
George Mason University
Fairfax, VA 22030

Dr. Robert D. Gibbons
University of Illinois at Chicago
NPI 909A, M/C 913
912 South Wood Street
Chicago, IL 60612

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01003

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Susan R. Goldman
Peabody College, Box 45
Vanderbilt University
Nashville, TN 37203

Dr. Timothy Goldsmith
Department of Psychology
University of New Mexico
Albuquerque, NM 87131

Dr. Sherrie Gott
AFHRL/MOMJ
Brooks AFB, TX 78235-5601

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Prof. Edward Haertel
School of Education
Stanford University
Stanford, CA 94305-3096

Dr. Ronald K. Hambleton
University of Massachusetts
Laboratory of Psychometric
and Evaluative Research
Hills South, Room 152
Amherst, MA 01003

Dr. Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign, IL 61820

Dr. Patrick R. Harrison
Computer Science Department
U.S. Naval Academy
Annapolis, MD 21402-5002

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 13
San Diego, CA 92152-6800

Dr. Thomas M. Hirsch
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Paul W. Holland
Educational Testing Service, 21-T
Rosedale Road
Princeton, NJ 08541

Prof. Lutz F. Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Ms. Julia S. Hough
Cambridge University Press
40 West 20th Street
New York, NY 10011

Dr. William Howell
Chief Scientist
AFHRL/CA
Brooks AFB, TX 78235-5601

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Martin J. Ippel
Center for the Study of
Education and Instruction
Leiden University
P. O. Box 9555
2300 RB Leiden
THE NETHERLANDS

Dr. Robert Jannarone
Elec. and Computer Eng. Dept.
University of South Carolina
Columbia, SC 29208

Dr. Kumar Joag-dev
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign, IL 61820

Professor Douglas H. Jones
Graduate School of Management
Rutgers, The State University
of New Jersey
Newark, NJ 07102

Dr. Brian Junker
Carnegie-Mellon University
Department of Statistics
Pittsburgh, PA 15213

Dr. Marcel Just
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213

Dr. J. L. Kaiwi
Code 442/JK
Naval Ocean Systems Center
San Diego, CA 92152-5000

Dr. Michael Kaplan
Office of Basic Research
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Jeremy Kilpatrick
Department of
Mathematics Education
105 Aderhold Hall
University of Georgia
Athens, GA 30602

Ms. Hae-Rim Kim
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Jwa-keun Kim
Department of Psychology
Middle Tennessee State
University
Murfreesboro, TN 37132

Dr. Sung-Hoon Kim
KEDI
92-6 Umyeon-Dong
Seocho-Gu
Seoul
SOUTH KOREA

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Department
501 North Dixon Street
P. O. Box 3107
Portland, OR 97209-3107

Dr. William Koch
Box 7246, Meas. and Eval. Ctr.
University of Texas-Austin
Austin, TX 78703

Dr. James Kraatz
Computer-based Education
Research Laboratory
University of Illinois
Urbana, IL 61801

Dr. Patrick Kyllonen
AFHRL/MOEL
Brooks AFB, TX 78235

Ms. Carolyn Laney
1515 Spencerville Rod
Spencerville, MD 20868

Richard Lanterman
Commandant (G-PWP)
US Coast Guard
2100 Second St., SW
Washington, DC 20593-0001

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
1310 South Sixth Street
University of IL at
Urbana-Champaign
Champaign, IL 61820-6990

Dr. Charles Lewis
Educational Testing Service
Princeton, NJ 08541-0001

Mr. Hsin-hung Li
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Library
Naval Training Systems Center
12350 Research Parkway
Orlando, FL 32826-3224

Dr. Marcia C. Linn
Graduate School
of Education, EMST
Tolman Hall
University of California
Berkeley, CA 94720

Dr. Robert L. Linn
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

Logicon Inc. (Attn: Library)
Tactical and Training Systems
Division
P.O. Box 85158
San Diego, CA 92138-5158

Dr. Richard Luecht
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. George B. Macready
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Evans Mandes
George Mason University
4400 University Drive
Fairfax, VA 22030

Dr. Paul Mayberry
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. James R. McBride
HumRRO
6430 Elmhurst Drive
San Diego, CA 92120

Mr. Christopher McCusker
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Dr. Robert McKinley
Educational Testing Service
Princeton, NJ 08541

Dr. Joseph McLachlan
Navy Personnel Research
and Development Center
Code 14
San Diego, CA 92152-6800

Alan Mead
c/o Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Timothy Miller
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Dr. Ivo Molenaar
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Grote Kruisstraat 2/1
9712 TS Groningen
The NETHERLANDS

Dr. E. Muraki
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Ratna Nandakumar
Educational Studies
Willard Hall, Room 213E
University of Delaware
Newark, DE 19716

Academic Progs. & Research Branch
Naval Technical Training Command
Code N-62
NAS Memphis (75)
Millington, TN 30854

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Norman, OK 73071

Head, Personnel Systems Department
NPRDC (Code 12)
San Diego, CA 92152-6800

Director
Training Systems Department
NPRDC (Code 14)
San Diego, CA 92152-6800

Library, NPRDC
Code 041
San Diego, CA 92152-6800

Librarian
Naval Center for Applied Research
in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington, DC 20375-5000

Office of Naval Research,
Code 1142CS
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Special Assistant for Research
Management
Chief of Naval Personnel (PERS-01JT)
Department of the Navy
Washington, DC 20350-2000

Dr. Judith Orasanu
Mail Stop 239-1
NASA Ames Research Center
Moffett Field, CA 94035

Dr. Peter J. Pashley
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

Dept. of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey, CA 93943-5026

Dr. Peter Pirolli
School of Education
University of California
Berkeley, CA 94720

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Mr. Steve Reise
Department of Psychology
University of California
Riverside, CA 92521

Mr. Louis Roussos
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Donald Rubin
Statistics Department
Science Center, Room 608
1 Oxford Street
Harvard University
Cambridge, MA 02138

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg.
Knoxville, TN 37966-0900

Dr. Mary Schratz
4100 Parkside
Carlsbad, CA 92008

Mr. Robert Semmes
N218 Elliott Hall
Department of Psychology
University of Minnesota
Minneapolis, MN 55455-0344

Dr. Valerie L. Shalin
Department of Industrial
Engineering
State University of New York
342 Lawrence D. Bell Hall
Buffalo, NY 14260

Mr. Richard J. Shavelson
Graduate School of Education
University of California
Santa Barbara, CA 93106

Ms. Kathleen Sheehan
Educational Testing Service
Princeton, NJ 08541

Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr. Randall Shumaker
Naval Research Laboratory
Code 5500
4555 Overlook Avenue, S.W.
Washington, DC 20375-5000

Dr. Judy Spray
ACT
P.O. Box 168
Iowa City, IA 52243

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Kikumi Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

Dr. David Thissen
Psychometric Laboratory
CB# 3270, Davie Hall
University of North Carolina
Chapel Hill, NC 27599-3270

Mr. Thomas J. Thomas
Federal Express Corporation
Human Resource Development
3035 Director Row, Suite 501
Memphis, TN 38131

Mr. Gary Thomasson
University of Illinois
Educational Psychology
Champaign, IL 61820

Dr. Howard Wainer
Educational Testing Service
Princeton, NJ 08541

Elizabeth Wald
Office of Naval Technology
Code 227
800 North Quincy Street
Arlington, VA 22217-5000

Dr. Michael T. Waller
University of
Wisconsin-Milwaukee
Educational Psychology Dept.
Box 413
Milwaukee, WI 53201

Dr. Ming-Mei Wang
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

Dr. Thomas A. Warm
FAA Academy
P.O. Box 25082
Oklahoma City, OK 73125

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

Dr. Douglas Wetzel
Code 15
Navy Personnel R&D Center
San Diego, CA 92152-6800

German Military
Representative
Personalstammamt
Koelner Str. 262
D-5000 Koeln 90
WEST GERMANY

Dr. David Wiley
School of Education
and Social Policy
Northwestern University
Evanston, IL 60208

Dr. Bruce Williams
Department of Educational
Psychology
University of Illinois
Urbana, IL 61801

Dr. Mark Wilson
School of Education
University of California
Berkeley, CA 94720

Dr. Eugene Winograd
Department of Psychology
Emory University
Atlanta, GA 30322

Dr. Martin F. Wiskoff
PERSEREC
99 Pacific St., Suite 4556
Monterey, CA 93940

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
03-0T
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Ms. Duanli Yan
Educational Testing Service
Princeton, NJ 08541

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

Dr. Joseph L. Young
National Science Foundation
Room 320
1800 G Street, N.W.
Washington, DC 20550