### REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

### 1. AGENCY USE ONLY (Leave blank)

<table>
<thead>
<tr>
<th>2. REPORT DATE</th>
<th>3. REPORT TYPE AND DATES COVERED</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 August 1992</td>
<td>Technical: 1990-93</td>
</tr>
</tbody>
</table>

### 4. TITLE AND SUBTITLE

Assessing Essential Dimensionality of Real Data

### 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. 2

### 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

### 12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

### 13. ABSTRACT (Maximum 200 words)

See reverse

### 14. SUBJECT TERMS

See reverse

### 15. NUMBER OF PAGES

25

### 16. PRICE CODE

UL

### 17. SECURITY CLASSIFICATION OF REPORT

unclassified

### 18. SECURITY CLASSIFICATION OF THIS PAGE

unclassified

### 19. SECURITY CLASSIFICATION OF ABSTRACT

unclassified

### 20. LIMITATION OF ABSTRACT

unclassified

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)  
Prescribed by ANSI Std Z39-18  
298-102
Assessing Essential Dimensionality of Real Data

Abstract

The purpose of this article is to validate the capability of DIMTEST to assess essential dimensionality of the model underlying the item responses of real tests as opposed to simulated tests. A variety of real test data from different sources are used to assess essential dimensionality. Based on DIMTEST results, some test data are assessed as fitting an essential unidimensional model while others are not. Essential unidimensional test data, as assessed by DIMTEST, are then combined to form two-dimensional test data. The power of Stout's statistic T is examined for these two-dimensional data. It is shown that the results of DIMTEST on real tests replicate findings from simulated tests in that the statistic T discriminates well between essential unidimensional and multidimensional tests. It is also highly sensitive to major abilities while being insensitive to relatively minor abilities influencing item responses.

Subject terms: DIMTEST, essential independence, essential dimensionality, unidimensionality, multidimensionality, item response theory.
Most of the currently used item response theory (IRT) models require the assumption of unidimensionality. From the strict IRT perspective, unidimensionality refers to one, and only one, trait underlying test items. Yet, it is a well known fact that items are multiply determined (Humphreys, 1981, 1985, 1986; Hambleton & Swaminathan, 1985, chap. 2; Reckase, 1979, 1985; Stout, 1987; Traub, 1983). Hence from the substantive viewpoint, the assumption of unidimensionality requires that the test items measure one dominant trait. Stout (1987) coined the term essential unidimensionality to refer to a particular mathematical formulation of a test having exactly one dominant trait. Dimensionality is, however, determined by the joint influence of test items and examinees taking the test (Reckase, 1990). In addition, extraneous factors such as teaching methods, anxiety level of examinees, etc., may also influence the dimensionality of the given item response data. Thus dimensionality has to be assessed each time a test is administered to a new group of examinees.

Factor analysis has traditionally been the most popular approach to assess dimensionality (Hambleton & Traub, 1973; Lumsden 1961). Factor analysis, despite its serious limitations to analyze dichotomous data (for example, see Hulin, Drasgow, and Parsons, 1983, chap. 8), has been the popular method to study the robustness of the unidimensionality assumption (Drasgow & Parsons 1983; Harrison, 1986; Reckase, 1979). There are a number of other promising methods proposed and used in varying degrees to assess dimensionality—to name a few: full information factor analysis based on the principle of marginal maximum likelihood (Bock, Gibbons, & Muraki, 1985; TESTFACT: Wilson, Wood, & Gibbons, 1983); nonlinear factor analysis (McDonald, 1962; McDonald & Ahlawat, 1974; Jamshid & McDonald, 1983); Holland and Rosenbaum’s (1986) test of unidimensionality, monotonicity and conditional independence based on contingency tables; Tucker and Humphreys’ methods based on the principle of local independence and second factor loadings (Roznowski, Tucker, & Humphreys, 1991); and Stout’s (1987)
statistical procedure based on essential independence and essential dimensionality. Hattie (1984, 1985) has provided a comprehensive review of traditional approaches to assess dimensionality, and Zwick (1987) has applied some of the above mentioned recent procedures to assess dimensionality of National Assessment of Educational Progress data. Despite having several procedures available to assess dimensionality, there is no widespread consensus among substantive researchers for a preference for any method(s), and often there is dissatisfaction about assessing dimensionality (Berger & Knol, 1990; Hambleton & Rovinelli, 1986; Hattie, 1985).

Stout (1987) proposed a statistical test (DIMTEST) to assess essential unidimensionality of the latent space underlying a set of items. Nandakumar (1987) and Nandakumar and Stout (in press) have further modified, refined, and validated DIMTEST for assessing essential dimensionality on a variety of simulated tests. This article demonstrates the validity and usefulness of Stout's procedure on a variety of real, as opposed to simulated, tests. Test data from different sources are collected and used to assess essential unidimensionality. Essential unidimensional data are then combined to form two-dimensional data. The power of Stout's statistic T is examined for these two-dimensional data.

**DIMTEST for Assessing Essential Unidimensionality**

DIMTEST, a statistical test for assessing unidimensionality, is based on the theory of essential dimensionality and essential independence (Stout, 1987, 1990). An item pool is said to be essentially independent with respect to the latent trait vector Θ if, for a given initial segment of the item pool, the average absolute conditional (on Θ) covariances of item pairs approaches zero as the length of the segment increases. When only one dominant ability Θ meets the essential independence assumption, the item pool is said to be
essentially unidimensional. In contrast, the assumption of local independence requires the conditional covariances to be zero for all item pairs in question. The number of abilities required to satisfy the local independence assumption is the dimensionality of the test.

While the traditional definition of dimensionality (Lord & Novick, 1968) counts all abilities required to respond to test items correctly to satisfy the assumption of local independence, essential dimensionality counts only dominant abilities required to satisfy the assumption of essential independence (as opposed to local independence). DIMTEST, using this definition, assesses the closeness of approximation of the model generating the given item responses to the essential unidimensional model. Nandakumar (1991) describes the theoretical differences between traditional dimensionality and essential dimensionality and establishes through Monte Carlo studies the usefulness of DIMTEST for assessing essential unidimensionality in the possible presence of several secondary dimensions.

To use DIMTEST for assessing essential unidimensionality, 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 assumed that essential independence with respect to some dominant ability $\Theta$ holds and that the item response functions are monotonic with respect to the same vector $\Theta$. The hypothesis is stated as follows:

$$H_0 : d_E = 1 \ versus \ H_1 : d_E > 1$$

where $d_E$ denotes the essential dimensionality of the latent space underlying a set of items.

In order to assess essential unidimensionality of a given test data, DIMTEST follows several steps. The steps are summarized briefly here (for details see Stout 1987; Nandakumar & Stout, in press). First, test items are split into three subtests AT1, AT2, and PT with the aid of factor analysis (FA) using part of the sample (a sample size of 500
is recommended for this purpose). Items of AT1 are selected so that they all tap the same dominant ability. Instead of using FA, it is also possible to use expert opinion (EO) to select items for AT1. If the FA method of selection is chosen, DIMTEST automatically determines the length of the subtest AT1. Once items for AT1 are chosen, items of AT2 are selected so that they have a difficulty distribution similar to those of AT1 items (for details see Stout, 1987). The remaining items form the partitioning subtest PT.

Second, examinees are assigned to K different subgroups based on their score on the partitioning subtest PT. In other words, all examinees obtaining the same PT total score are assigned to the same subgroup. When the subtest PT is "long" and the test is essentially unidimensional, within each subgroup k, examinees are assumed to be approximately of similar ability. When PT is not long, the subtest AT2 compensates for the bias in AT1 caused by PT being short. Also, AT2 compensates for the bias in AT1 caused by the presence of guessing or the difficulty factor that is often found by the factor analysis.

Third, within each subgroup k, variance estimates, $\hat{\sigma}_k^2$ and $\hat{\sigma}_{U,k}^2$, and the standard error of estimate $S_k$ are computed using item responses of AT1. These estimates are then summed across K subgroups to obtain

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

Similarly, $T_B$ is computed using items of subtest AT2. Stout's statistic $T$ is given by

$$T = (T_L - T_B)/\sqrt{2}.$$  

The decision rule is to reject $H_0$ if $T \geq Z_{\alpha}$, where $Z_{\alpha}$ is the upper 100(1-\alpha) percentile of the
When the given test data are well modeled by an essential unidimensional model, items of AT1, AT2, and PT would all be tapping the same dominant dimension. Therefore, the variance estimates \( \hat{\sigma}_k^2 \) and \( \hat{\sigma}_{U,k}^2 \) will be approximately equal resulting in a "small" T-value, suggesting the tenability of \( H_0 \). On the other hand, when the test data is not well modeled by an essential unidimensional model, the variance estimate \( \hat{\sigma}_k^2 \) will be much larger than \( \hat{\sigma}_{U,k}^2 \) resulting in a "large" T-value leading to the rejection of \( H_0 \).

Simulation studies (Stout, 1987; Nandakumar, 1987; Nandakumar & Stout in press) on a wide variety of tests have demonstrated the utility of DIMTEST in discriminating between one- and two-dimensional tests. Simulation studies by Nandakumar (1991) have particularly demonstrated the usefulness of DIMTEST in assessing essential unidimensionality with the aid of a rough index of deviation from essential unidimensionality. The tests in Nandakumar (1991) were modeled by two- and higher-dimensional IRT models as opposed to a one-dimensional model, and the test items were influenced by major and secondary abilities to varying degrees. For some tests, the secondary ability or abilities influenced a high proportion of items, and for others the secondary ability or abilities influenced only a small proportion of items. It has been shown that DIMTEST reliably accepts the hypothesis of essential unidimensionality, provided the model generating the test is close to the essential unidimensional model: established when each of the secondary abilities influences relatively few items, or if secondary abilities are influencing many items, the degree of influence on each item is small. The type-I error in these cases was within tolerance of nominal level. As the degree of influence of the secondary abilities increases, however, the approximation to an essential unidimensional model degenerates, inflating the observed type-I error of the hypothesis of essential unidimensionality. Simulation results (Stout, 1987; Nandakumar and Stout, in press) have particularly demonstrated the excellent power of the statistic T when the model generating
the item responses is two-dimensional (two major abilities) with correlation between abilities as high as .7 and items jointly influenced by both abilities.

Description of Data

The data sets used in the present study came from different sources. The U.S. history and literature data for grade 11/age 17, from the 1986 National Assessment of Educational Progress (NAEP, 1988) test data, were obtained from Educational Testing Service (ETS). The General Science data, Arithmetic Reasoning data, and Auto Shop Information data for grades 10 and 12, from the Armed Services Vocational and Aptitude Battery (ASVAB) test data, were obtained from Linn, Hastings, Hu, and Ryan (1987). The Mathematics Usage test data, the science test data, and the reading test data were obtained from American College Testing program (ACT).

The NAEP achievement tests are part of the so called Balanced Incomplete Block (BIB) design with spiraled administration (Rogers et al., 1988) which allows the study of interrelationships among all items within a subject area. Because the U.S. history and literature tests fall into the simplest category of BIB design, it was relatively easy to gather the response data for all examinees taking these tests. Hence, these tests were chosen for the present study. The items in each area (history and literature) were divided into four "parallel" blocks with approximately the same number of items. One block of items out of four was randomly selected in each case for the present study.

The U.S. history test data (HIST–A) with 36 items consists of items requiring knowledge from different time periods of U.S. history: Colonization to 1763; the Revolutionary War and the New Republic, 1763–1815; Civil War, 1815–1877; the rise of modern America, World War I 1877–1920; the Depression, World War II, 1920–1945; Post–World War II, 1945–to the present; and map items requiring the knowledge of
geographical location of different countries in the world. A 31-item subtest of HIST–A, named HIST was created (explained in detail in the next section) consisting of all the items of HIST–A, except the five map items. There are 2428 examinees in the HIST–A and HIST samples.

The literature test data (LIT) with 30 items consists of items requiring knowledge within four literary genres: novels, short stories, and plays; myths, epics, and Biblical characters and stories; poetry; and nonfiction. There are 2439 examinees in the LIT sample.

The ASVAB tests are used by the Department of Defense Student Testing Program in high schools and post secondary schools. The Arithmetic Reasoning test data for grades 10 and 12, with 30 items each, consists of items requiring knowledge in solving arithmetic word problems. The arithmetic reasoning test sample for grade 10 (AR10) has 1984 examinees, and for grade 12 (AR12) has 1961 examinees. The Auto and Shop Information test data for grades 10 and 12, with 25 items, each consists of items requiring knowledge of automobile, tools, and shop terminology and practices. The auto shop test sample for grade 10 (AS10) has 1981 examinees, and for grade 12 (AS12) has 1974 examinees. The General Science test data for grades 10 and 12, with 25 items each, consists of items requiring knowledge in solving high school level physical, life, and earth sciences. There are 1990 examinees in the general science test sample for grade 10 (GS10) and 1990 examinees in the general science grade 12 (GS12) sample.

The ACT mathematics usage test data (MATH) with 40 items consists of items requiring knowledge in solving different types of mathematics problems: arithmetic and algebra operations, geometry, numeration, story problems, and advanced topics. There are 2491 examinees in the MATH sample.

The ACT reading test data (READ–A) with 40 items consists of 4 passages, each followed by 10 questions. The first three passages are taken from different books all dealing with humanities, and the last passage is taken from a book about psychology. The first
passage came from *Of the Farm* by John Updike. The second passage came from *Light and Color in Nature and Art* by Samuel Williamson and Herman Cummins. The third passage came from *Theatre: the Dynamics of the Art* by Brian Hansen. And the fourth passage came from *Toward a Psychology of Being* by Abraham Maslow. A 30-item subset of READ—A named READ was created (details in the next section) consisting of the first 30 items of READ—A. There are 5000 examinees in the READ—A and READ samples.

The ACT science test data (SCI—A) with 40 items consists of 7 passages, each followed by 5 to 7 questions. The first passage dealt with the effect of the thymus gland on the development of immune system in mice. The second passage dealt with sub-surface ground water movement and its effects for waste disposal. The third passage dealt with the periods of the pendulum on the earth and the moon and its relationship to the string length and mass of the ball. The fourth passage dealt with the environmental impact of effluent. The fifth passage dealt with a bimetallic catalyst and its relationship to the speed of certain chemical reactions. The sixth passage dealt with the views of two paleontologists on the characteristics of dinosaurs. And the seventh passage dealt with the principals of osmosis and osmotic characteristics of 3 categories of organisms. A 28-item subset of SCI—A named SCI was created (explained in the next section) consisting of the first 28 items of SCI—A. There are 5000 examinees in SCI—A and SCI samples.

In addition, in order to examine the effect of sample size on DIMTEST, both SCI and READ are randomly split into four mutually exclusive data sets. The READ is split into READ1, READ2, READ3, and READ4—with 750, 1000, 1250 and 2000 examinees, respectively. Similarly SCI is split into SCI1, SCI2, SCI3, and SCI4—with 750, 1000, 1250, and 2000 examinees, respectively. In all there are 22 test data. These are listed along with the test size and sample size in the first three columns of Tables 1 and 2.
Creation of Two–Dimensional Test Data

Three different sets of two–dimensional test data from the content perspective were created by combining responses from test data that were assessed as essentially unidimensional by DIMTEST in the present study.

The two–dimensional test data, RS, was created by combining responses of 30 items of READ with the responses of 6 items of SCI forming a 36–item test with 5000 examinees. The 6 items of SCI are part of one of the passages randomly selected from its 5 passages. Just as in the unidimensional case of READ and SCI, RS is then randomly split into 4 mutually exclusive data sets RS1, RS2, RS3, and RS4— with 750, 1000, 1250 and 2000 examinees, respectively. These tests are listed along with their test sizes and sample sizes in the first four columns of Table 3.

The two–dimensional test data ARGS1, for Grade 10, was created by combining the responses of 30 items from AR10 with the responses of 5 items (randomly selected from 25 item responses) from GS10. Similarly, ARGS2 was created by combining the responses of 30 items from AR10 with the responses of 10 items from GS10. The two–dimensional test data GSAR1, for grade12, was created by combining the responses of 25 items from GS12 with the responses of 5 items from AR12; and GSAR2 was created by combining the responses of 25 items from GS12 with responses of 10 items from AR12. These test data are listed along with their test sizes and sample sizes in the first four columns of Table 4.

The two–dimensional test data HSTLIT1 was created by combining the responses of 31 items from HIST with the responses of 5 items (randomly selected from 30 item responses) from LIT. Similarly HSTLIT2 and HSTLIT3 were created by combining the responses of 31 items from HIST with the responses of 8 and 10 items, randomly selected, from LIT respectively. These test data are listed along with their test sizes and sample sizes in the first four columns of Table 5.
ASSESSING ESSENTIAL DIMENSIONALITY–12

Results

Unidimensional Studies

All the tests in Table 1, except HIST, READ, and SCI (which are derived subtests of HIST–A, READ–A, and SCI–A, respectively as described below), were initially tested for essential unidimensionality using DIMTEST. In each case, 500 examinees were randomly selected from the given pool for the use of selecting AT1 items, using factor analysis. The rest of the items were used for computing Stout's statistic T. The size of AT1 (M) was also determined by DIMTEST. For each test, the T-value and the p-value are noted. Table 1 lists the T- and p-values for all tests in the fourth and fifth columns. The method of selection of the AT1 subtest, the value of M, and item numbers selected for AT1 are listed in the last three columns of Table 1.

Table 1 about here

It can be seen from Table 1 that the p-values associated with test data LIT, AR10, AR12, GS10, and GS12 are well above the nominal level of significance (α=.05), thereby strongly affirming essential unidimensional nature of these tests. That is, the underlying model generating the test data is judged essentially unidimensional. However, the p-values associated with HIST–A, AS10, AS12, MATH, READ–A, and SCI–A are well below the nominal level of significance of .05, thereby strongly affirming the multidimensional nature of these test data. For these tests where p-values were below the nominal level, the nature of multidimensionality was further explored.
When the test data are essentially unidimensional, items of AT1 are, by logic, of the same dominant dimension as the rest of the items; therefore, DIMTEST does not reject the null hypothesis. When the test data is not unidimensional, however, the items of AT1 are dimensionally different from the rest of the items, and DIMTEST rejects the null hypothesis of essential unidimensionality. Following this reasoning for tests where p-values were very low, the content of items of AT1 were examined. Table 1 shows that for HIST-A, items 12 through 16 and item 6 were selected for AT1. Upon studying the content of these items, it was found that items 12 through 16 were homogeneous and differed dimensionally from the rest of the items of HIST-A; these 5 items require the knowledge of location of different countries on the world map (map items), while the rest of the items deal with U.S. history. It is also possible in theory that these items were selected for AT1 due to chance alone. In order to test for this, DIMTEST was applied on the given sample of 2428 examinees 100 times repeatedly, each time randomly splitting 2428 examinees into two groups of 500 and 1928 examinees. That is, AT1 items were selected repeatedly on different random samples of 500 examinees each. The resampling results showed that items 12 through 16 were consistently selected for AT1. In addition to these items one or two more items, which varied from run to run, were selected from the rest of the items. Hence it was concluded that the map items are dimensionally different from the rest. A subset HIST was formed consisting of all items of HIST-A except for map items. It can be seen from Table 1 that the p-value associated with HIST (p=.095) shows evidence of essential unidimensionality. Furthermore, from the content perspective, items of AT1 do not form a set that is dimensionally different from the rest of the items of HIST.

A similar phenomenon was observed with test data READ-A and SCI-A. For READ-A, the last 10 items (items followed by the last passage) formed part of subtest AT1. Again these same 10 items formed part of AT1 in repeated resampling applications of DIMTEST. Upon studying the content of these items, it was found that these 10 items
tapped "psychology" content area which is different from the "literature," tapped by the first three passages. Another possibility is that, since these are the last 10 items of reading test, speededness could have caused the secondary dimension. Based on these observations, it was concluded that these items were dimensionally different from the rest, and a subset READ was formed consisting of first 30 items of READ-A. It can be seen from Table 1 that the $p$-value associated with READ ($p=.32$) shows strong evidence of an essential unidimensional model underlying the test items. In addition, items of AT1 now come from all the passages of READ.

For test data SCI-A, the 12 items following the last two passages formed part of AT1. Just as in HIST-A and READ-A, after resampling application of DIMTEST, these items were removed. The resulting subtest SCI with the first 28 items was still found to be multidimensional ($p=.002$). Thus, a unidimensional subset could not be formed. Unlike reading test items, science test items come from distinctly different content areas, with a moderate correlation among content areas, and require a higher level of abstract reasoning and analytical skills than the reading items. Thus, in addition to content areas, difficulty or speededness could have caused major secondary dimensions in this case.

For the test data MATH, AS10, and AS12, where $p$-values were low, items of AT1 did not form a subgroup tapping a secondary ability as found in HIST-A, READ-A, or SCI-A. In addition upon studying the content of the items, it was found these items tap multiple major content areas. Therefore these test data are treated as multidimensional.

Table 2 shows dimensionality results of the unidimensional READ and
multidimensional SCI test data for different sample sizes. The p-values associated with READ1 through READ4 show evidence of a high degree of essential unidimensionality underlying the test data. These results are consistent with that of READ in Table 1. The selection of items of AT1 for tests READ1 through READ4 are highly varied, and yet they consistently affirm essential unidimensionality. The results of SCI1 through SCI4 are consistent with that of SCI in Table 1 in affirming multidimensionality of the test data. Items of AT1 varied highly for all four tests and yet consistently affirmed multidimensionality, except for SCI3.

Two-dimensional Studies

Results of two-dimensional reading and science test data are reported in Table 3. Since items that tap a distinct second dimension, from the content perspective, are clearly known (in this case, 6 SCI items), the science items were forced to be selected for AT1. This is an example where expert opinion is used to select AT1 items. The T- and p-values for RS1, RS2, RS3, RS4, and RS strongly confirm the two-dimensional nature of these test data. As expected, as the sample size increases, the power also increases.

The results of the two-dimensional test data of ARGS and GSAR are reported in Table 4. Also in this case, since items that are used to create these two-dimensional data are known (GS items for ARGS and AR items for GSAR), these items were forced to be selected for AT1. The T- and p-values associated with all the four tests strongly confirm
the multidimensionality of these test data. For ARG1 and ARG2, there is a sharp increase in $T$- and $p$-values as the degree of contamination, as measured by the number of item responses contaminated, increases from 5 to 10.

The results of the two-dimensional history and literature test data are reported in Table 5. As with other two-dimensional tests, LIT items were forced to be selected for AT1. Also in this case, the $T$- and $p$-values confirm the multidimensional nature of these data.

DIMTEST was again applied to a sample of test data selected from two-dimensional tests. This time FA was used as the method of selection for AT1 items. The purpose of this analysis was to check if the FA method of selection of AT1 items would lead to the similar $p$-values as with EO. The findings revealed that for these tests FA could not always ferret out purely unidimensional items from content perspective. The subtest AT1 had a mixture of items tapping both dimensions, and DIMTEST was then able to correctly assess dimensionality only when there were 1000 or more examinees for computing the statistic.

Discussion and Conclusions

None of the tests examined in the present study are strictly unidimensional in the sense of measuring only one ability. Items, in every test, are influenced by several secondary abilities in addition to the major ability intended to be measured. Based on DIMTEST analysis, some test data were assessed as fitting an essential unidimensional model while others were not. This depends upon whether the secondary abilities were major or minor.

The unidimensionality analysis of HIST–A, READ–A, and SCI–A present interesting findings. For HIST–A, the map items had high second factor loadings and thus were selected for AT1. Consequently, the computed $T$-statistic was large, leading to the
rejection of \( H_0 \) and implying that AT1 items are dimensionally different from the rest of the test. Content analysis of HIST–A reveals that HIST–A consists of items of United States history for different time periods spanning from 1763 to present time. These items cover such a large span of time that the test is surely slightly multidimensional for this reason alone. In addition, the test contains map items. The map items, however, were isolated and statistically confirmed as not measuring the same trait as the rest of the test. This shows that the statistic \( T \) is highly sensitive to distinct major dimensions (in this case, map items). The analysis of HIST, with map items removed, reveals that it is essentially unidimensional. Thus the statistic \( T \) seems to be robust against relatively minor correlated abilities influencing test items while being sensitive to major abilities. Likewise, for the test data READ–A, multidimensionality was caused by items tapping psychology topic (scientific) versus literature topics (humanities). Once the psychology item responses were removed, the remaining item responses could be well modeled by an essential unidimensional model. In contrast, the multidimensionality in SCI–A was due to not only distinct major abilities but also likely due to speededness of the test, which in itself is a major determinant. Moreover, an essential unidimensional subtest could not be formed for SCI–A.

Another interesting feature of these analyses is that although both READ and SCI are paragraph comprehension type test data, they differ widely in the degree of their approximation to essential dimensionality. The READ test data has 3 passages each followed by 10 items, all dealing with humanities. Although these passages come from different sources, the model underlying the item responses approximates an essential unidimensional model. This is an example where a few secondary abilities (possibly highly correlated) each influence a large group of items. In contrast, the SCI test data has 5 passages each followed by 5 or 6 items. These passages, although they deal with science in general, come from widely different and conceptually difficult topics, and the model
underlying the item responses does not approximate an essential unidimensional model. This is an example where many secondary abilities each influence a small groups of items, but the strength of the influence of these secondary abilities is such that item responses can not be well modeled by an essential unidimensional model. These results are consistent with simulation results of Nandakumar (1991) in that the number of items influenced by secondary abilities and the strength of the secondary abilities present determine the degree to which the assumption of essential unidimensionality is violated.

The results obtained in this study are similar to the results obtained by other researchers who have analyzed some of these data using different statistical methodologies. Zwick (1987) performed dimensionality analyses of HIST—A and LIT by various techniques to assess dimensionality and concluded that these are unidimensional. Regarding the ACT data, it is believed that MATH and SCI are multidimensional. Bock, Gibbons, and Muraki (1985) have analyzed ASVAB test data for a different sample and found a significant second factor for arithmetic reasoning, general science, and auto shop information. Since the sample used here is not the same it is hard to develop a meaningful comparison.

The results of two-dimensional tests demonstrate a very good power of the statistic T. The statistic T has the capability to ignore minor secondary traits, which should be largely discounted, from the major dominant traits. This is evidenced in several cases. The test data HIST illustrates this. There is inherent multidimensionality in HIST as it covers a range of time periods in history. However, the p-value is above the nominal level of significance, suggesting acceptance of unidimensionality. By contrast, with the additional contamination of only 5 LIT items or 5 map items, the T-value shoots up, indicating essential multidimensionality of the data. This remarkable sensitivity of the statistic T to major dimensions illustrates its power.

These results, for the first time, have illustrated both the factor analysis approach and the expert opinion approach to select items for the subtest AT1. Tables 1 and 2 use FA
to select AT1 items, and Tables 3, 4, and 5 use EO. It is evident that FA serves as an exploratory tool and EO serves as a confirmatory tool in selecting items for AT1 to assess essential dimensionality.

The dimensionality of a given set of item responses in certain sense is a continuum—one cannot determine whether a given data of responses generated by a set of items to an examinee sample is truly essentially unidimensional or truly multidimensional; one can only approximate. Although the exact number of dimensions in an IRT model is rigorously defined for a finite length test, the number of dominant dimensions—whether determined by Stout’s essential dimensionality conceptualization or by some other conceptualization—is only rigorously definable for an infinitely long test. In other words, for a finite test (that is, for any real test data) it is a judgment call whether a particular IRT model is seen as having one, or more than one, dominant dimension, based upon where on the continuum the amount of multidimensionality falls. One consequence of this is that the performance of ability estimation procedures such as LOGIST or BILOG needs to be addressed in the context of the assessment of the amount of lack of unidimensionality. In this regard, indices of lack of essential unidimensionality developed by Junker and Stout (1991) will be extremely useful. These indices can be used to decide when it is safe to use unidimensional estimation procedures such as LOGIST and BILOG to arrive at accurate estimates of ability.

In cases where approximation of essential unidimensional model to the data is in question, there are various alternatives. The test items can be split into essential unidimensional subtests (for example, HIST—A and READ—A). Another possible approach is to investigate the applicability of the concept of "testlet" to the data (Rosenbaum, 1988; Thissen, Steinberg, and Mooney, 1989). If the assumption of local independence is violated within the passages but maintained among the passages, the theory of testlets promises unidimensional scoring for such tests. The test data SCI—A and SCI could fall into this
category. Multidimensional modeling can be applied if either of the above procedures cannot be applied (Reckase, 1989).
References


ASSESSING ESSENTIAL DIMENSIONALITY

theory. Paper presented at the 7th European Meeting of the Psychometric Society, Trier, Germany.


Acknowledgements

The author thanks Bill Stout and Brian Junker for their helpful comments and many suggestions on this research, and Mark Reckase and Tim Miller for providing ACT data.
Table 1
Results of $H_0: d_E = 1, \alpha = .05$

<table>
<thead>
<tr>
<th>Test</th>
<th>No. of items</th>
<th>No. of examinees</th>
<th>$T$</th>
<th>$p$</th>
<th>Selection of $M^{**}$</th>
<th>Items of $AT1$ items</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIST-A</td>
<td>36</td>
<td>2428</td>
<td>6.19</td>
<td>.00001</td>
<td>FA</td>
<td>6,12,13,14,15,16</td>
</tr>
<tr>
<td>HIST</td>
<td>31</td>
<td>2428</td>
<td>1.31</td>
<td>.095</td>
<td>FA</td>
<td>5,7,23,24,26,30</td>
</tr>
<tr>
<td>LIT</td>
<td>30</td>
<td>2439</td>
<td>.71</td>
<td>.234</td>
<td>FA</td>
<td>5,9,18,20,22,26</td>
</tr>
<tr>
<td>AR10</td>
<td>30</td>
<td>1984</td>
<td>-.75</td>
<td>.727</td>
<td>FA</td>
<td>6,1,3,4,5,6,8</td>
</tr>
<tr>
<td>AR12</td>
<td>30</td>
<td>1961</td>
<td>.64</td>
<td>.260</td>
<td>FA</td>
<td>4,1,4,6,14</td>
</tr>
<tr>
<td>GS10</td>
<td>25</td>
<td>1990</td>
<td>.96</td>
<td>.168</td>
<td>FA</td>
<td>5,4,16,19,23,25</td>
</tr>
<tr>
<td>GS12</td>
<td>25</td>
<td>1988</td>
<td>-.26</td>
<td>.601</td>
<td>FA</td>
<td>6,14,15,19,23,24,25</td>
</tr>
<tr>
<td>AS10</td>
<td>25</td>
<td>1981</td>
<td>2.27</td>
<td>.012</td>
<td>FA</td>
<td>5,4,16,19,23,25</td>
</tr>
<tr>
<td>AS12</td>
<td>25</td>
<td>1974</td>
<td>3.64</td>
<td>.000</td>
<td>FA</td>
<td>5,3,4,8,14,22</td>
</tr>
<tr>
<td>MATH</td>
<td>40</td>
<td>2491</td>
<td>2.79</td>
<td>.003</td>
<td>FA</td>
<td>10,1,5,25,27,29,30,32,34,35,39</td>
</tr>
<tr>
<td>READ-A</td>
<td>40</td>
<td>5000</td>
<td>8.67</td>
<td>.00001</td>
<td>FA</td>
<td>10,31,32,33,34,35,36,37,38,39,40</td>
</tr>
<tr>
<td>READ</td>
<td>30</td>
<td>750</td>
<td>.05</td>
<td>.480</td>
<td>FA</td>
<td>7,1,2,6,11,12,13,21</td>
</tr>
<tr>
<td>SCI-A</td>
<td>40</td>
<td>5000</td>
<td>.48</td>
<td>.32</td>
<td>FA</td>
<td>12,29,30,31,32,33,34,35,36,37,38,39,40</td>
</tr>
<tr>
<td>SCI</td>
<td>28</td>
<td>5000</td>
<td>3.19</td>
<td>.0007</td>
<td>FA</td>
<td>5,2,3,5,8,12</td>
</tr>
</tbody>
</table>

*AT1 items can be selected by using factor analysis (FA) or by expert opinion (EO).

**$M$ is the size of AT1

Table 2
Results of $H_0: d_E = 1, \alpha = .05$

<table>
<thead>
<tr>
<th>Test</th>
<th>No. of items</th>
<th>No. of examinees</th>
<th>$T$</th>
<th>$p$</th>
<th>Selection of $M$</th>
<th>Items of $AT1$ items</th>
</tr>
</thead>
<tbody>
<tr>
<td>READ1</td>
<td>30</td>
<td>750</td>
<td>.05</td>
<td>.480</td>
<td>FA</td>
<td>5,11,12,13,15,17</td>
</tr>
<tr>
<td>READ2</td>
<td>30</td>
<td>1000</td>
<td>.48</td>
<td>.317</td>
<td>FA</td>
<td>7,1,2,6,11,12,13,21</td>
</tr>
<tr>
<td>READ3</td>
<td>30</td>
<td>1250</td>
<td>-.06</td>
<td>.524</td>
<td>FA</td>
<td>7,2,4,6,9,11,12,13</td>
</tr>
<tr>
<td>READ4</td>
<td>30</td>
<td>2000</td>
<td>1.01</td>
<td>.155</td>
<td>FA</td>
<td>5,1,11,12,13,16</td>
</tr>
<tr>
<td>SCI1</td>
<td>28</td>
<td>750</td>
<td>1.99</td>
<td>.029</td>
<td>FA</td>
<td>7,1,3,4,5,17,20,21</td>
</tr>
<tr>
<td>SCI2</td>
<td>28</td>
<td>1000</td>
<td>3.19</td>
<td>.007</td>
<td>FA</td>
<td>6,8,12,14,18,20,24</td>
</tr>
<tr>
<td>SCI3</td>
<td>28</td>
<td>1250</td>
<td>1.38</td>
<td>.080</td>
<td>FA</td>
<td>7,6,9,10,11,19,25,28</td>
</tr>
<tr>
<td>SCI4</td>
<td>28</td>
<td>2000</td>
<td>2.91</td>
<td>.001</td>
<td>FA</td>
<td>7,8,9,10,11,12,19,22</td>
</tr>
</tbody>
</table>
### Table 3
Results of $H_0: \delta = 1$ for two-dimensional tests:
READ & SCI; $\alpha = .05$

<table>
<thead>
<tr>
<th>Test</th>
<th>No. of Items</th>
<th>No. of Examinees</th>
<th>$T$</th>
<th>$p$</th>
<th>Selection of AT1 items</th>
<th>$M$</th>
<th>Items of AT1</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS1</td>
<td>30</td>
<td>6</td>
<td>1.92</td>
<td>.020</td>
<td>EO</td>
<td>6</td>
<td>31,32,33,34,35,36</td>
</tr>
<tr>
<td>RS2</td>
<td>30</td>
<td>6</td>
<td>2.72</td>
<td>.003</td>
<td>EO</td>
<td>6</td>
<td>31,32,33,34,35,36</td>
</tr>
<tr>
<td>RS3</td>
<td>30</td>
<td>6</td>
<td>3.71</td>
<td>.0001</td>
<td>EO</td>
<td>6</td>
<td>31,32,33,34,35,36</td>
</tr>
<tr>
<td>RS4</td>
<td>30</td>
<td>6</td>
<td>3.32</td>
<td>.0005</td>
<td>EO</td>
<td>6</td>
<td>31,32,33,34,35,36</td>
</tr>
<tr>
<td>RS</td>
<td>30</td>
<td>6</td>
<td>6.83</td>
<td>.000</td>
<td>EO</td>
<td>6</td>
<td>31,32,33,34,35,36</td>
</tr>
</tbody>
</table>

### Table 4
Results of $H_0: \delta = 1$ for two-dimensional tests:
AR & GS; $\alpha = .05$

<table>
<thead>
<tr>
<th>Test</th>
<th>No. of Items</th>
<th>No. of Examinees</th>
<th>$T$</th>
<th>$p$</th>
<th>Selection of AT1 items</th>
<th>$M$</th>
<th>Items of AT1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARGSl</td>
<td>30</td>
<td>5</td>
<td>2.85</td>
<td>.002</td>
<td>EO</td>
<td>5</td>
<td>31,32,33,34,35</td>
</tr>
<tr>
<td>ARGs2</td>
<td>30</td>
<td>10</td>
<td>6.15</td>
<td>.000</td>
<td>EO</td>
<td>10</td>
<td>31,32,33,34,35,36,37,38,39,40</td>
</tr>
<tr>
<td>GSAR1</td>
<td>25</td>
<td>5</td>
<td>4.29</td>
<td>.000</td>
<td>EO</td>
<td>5</td>
<td>26,27,28,29,30</td>
</tr>
<tr>
<td>GSAR2</td>
<td>25</td>
<td>10</td>
<td>4.06</td>
<td>.000</td>
<td>EO</td>
<td>10</td>
<td>26,27,28,29,30,31,32,33,34,35</td>
</tr>
</tbody>
</table>

### Table 5
Results of $H_0: \delta = 1$ for two-dimensional tests:
HIST & LIT; $\alpha = .05$

<table>
<thead>
<tr>
<th>Test</th>
<th>No. of Items</th>
<th>No. of Examinees</th>
<th>$T$</th>
<th>$p$</th>
<th>Selection of AT1 items</th>
<th>$M$</th>
<th>Items of AT1</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSTLIT1</td>
<td>31</td>
<td>5</td>
<td>3.01</td>
<td>.036</td>
<td>EO</td>
<td>5</td>
<td>32,33,34,35,36</td>
</tr>
<tr>
<td>HSTLIT2</td>
<td>31</td>
<td>8</td>
<td>3.38</td>
<td>.000</td>
<td>EO</td>
<td>8</td>
<td>32,33,34,35,36,37,38,39</td>
</tr>
<tr>
<td>HSTLIT3</td>
<td>31</td>
<td>10</td>
<td>2.03</td>
<td>.021</td>
<td>EO</td>
<td>10</td>
<td>32,33,34,35,36,37,38,39,40,41</td>
</tr>
</tbody>
</table>
Distribution List

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

Dr. Susan Eastwood
University of Kansas
Psychology Department
430 Fraser
Lawrence, KS 66045

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

ERIC Facility-Acquisitions
2440 Research Blvd, Suite 550
Rockville, MD 20850-3338

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

Dr. Leonard Feith
Lincoln Center for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Richard L. Ferguson
American College Testing
P.O. Box 158
Iowa City, IA 52243

Dr. Gerhard Fischer
LIEbiggasse 5
A-1010 Vienna
AUSTRIA

Dr. Myron Fishel
U.S. Army Headquarters
DAFE-HR
The Pentagon
Washington, DC 20318-0300

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152-6000

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

Dr. Robert Gibbons
University of Illinois at Chicago
NPF 999A, M/C 913
912 South Wood Street
Chicago, IL 60612

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

Dr. Robert Glaser
Learning Research & Development Center
University of Pittsburgh
3599 O'Hara Street
Pittsburgh, PA 15260

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

Dr. Timothy Goldsmith
Department of Psychology
University of New Mexico
Albuquerque, NM 87131
Dr. Joseph McLauchlan
Navy Personnel Research
and Development Center
Code 14
San Diego, CA 92152-6000

Alan Mulvihill
Director
Dr. Michael Levens
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Timothy Miller
ACT
P. O. Box 164
Iowa City, IA 52243

Dr. Robert Maloney
Educational Testing Service
Princeton, NJ 08541

Dr. Ivo Molenaar
Faculteit Sociale Werkenhousing
Rutgers University
Groningen
Groote Kruislaan 271
NL-9712 TS Groningen
The NETHERLANDS

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

Dr. Rama Nandakumar
Educational Studies
University of Delaware
Newark, DE 19716

Academic Programs & Research Branch
Naval Technical Training Command
Code N-45
NAS Memphis (73)
Memphis, TN 38054

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

Head, Personnel Systems Department
NPRDC (Code 121)
San Diego, CA 92152-6000

Director
Training Systems Department
NPRDC (Code 14)
San Diego, CA 92152-6000

Library, NPRDC
Code 041
San Diego, CA 92152-6000

Library
Naval Center for Applied Research
in Artificial Intelligence
Naval Research Laboratory
Code 5310
Washington, DC 20375-5000

Office of Naval Research
Code 14232S
114 N. Quincy Street
Arlington, VA 22217-5000
(4 Copies)

Special Assistant for Research
Management
Chief of Naval Personnel (PERS-OUT)
Department of the Navy
Washington, DC 20356-2000

Dr. Judith Oramo
Mail Stop 236-N
NASA Ames Research Center
Moffett Field, CA 94035

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

Wayne M. Paliare
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 Pecoll
School of Education
University of California
Berkeley, CA 94720

Dr. Mark D. Reckase
ACT
P. O. Box 164
Iowa City, IA 52243

Mr. Steve Reise
Department of Psychology
University of California
Riverside, CA 92511

Mr. Louis Ronca
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61801

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

Dr. Tetsuji Sato
Department of Psychology
University of Tennessee
3108 Austin Peay Blvd.
Knoxville, TN 37996-0000

Dr. Mary Schubert
E120 Parkside
Carbondale, IL 62901

Dr. Steven Semmes
N121 Elliott Hall
Department of Psychology
University of Minnesota
Minneapolis, MN 55455-0344

Dr. Valerie L. St. Atian
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

Mr. Keith Sheehan
Educational Testing Service
Princeton, NJ 08541

Dr. Kazuo Shiga
Tokyo, JAPAN

Dr. Randall Shumaker
Naval Research Laboratory
Code 5300
4551 Overlook Avenue, S.W.
Washington, DC 20375-3000

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

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. William Swing
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61801

N. Takasuka
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

Dr. David Thissen
Psychometric Laboratory
CIB 3720, Dave Hall
University of North Carolina
Chapel Hill, NC 27599-3370

Mr. Garnett Thomas
Federal Express Corporation
Human Resource Development
3035 Director Row, Suite 501
Memphis, TN 38133

Mr. Gary Throthman
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

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

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

Dr. Thomas A. Wirtsch
FAC Academy
P.O. Box 25982
Oklahoma City, OK 73125

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0341

Dr. Douglas Wesel
Code 15
Naval Personnel R&D Center
San Diego, CA 92152-6000

German Military
Representative
Personalmakleramt
Koetner 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  
Hollywood, CA 90020

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

Dr. Mano F. Witeoff  
PbSf SEC  
A Pine St., Suite 4556  
Monterey, CA 93940

Mr. John H. Wolfe  
Nav Personnel R&D Center  
San Diego, CA 42121-5600

Dr. Kicacho Yamamoto  
WAT  
Educational Testing Service  
Rosedale Road  
Princeton, NJ 08541

Mr. Duan Yen  
Educational Testing Service  
Princeton, NJ 08541

Dr. Wendy Yen  
CTB McGraw Hill  
The Minute Research Park  
Monterey, CA 93940

Dr. Joseph L. Young  
National Science Foundation  
Washington, D.C. 20550