(4)

A New Item Response Theory Modeling Approach with Applications to Unidimensionality Assessment and Ability Estimation

William F. Stout
Department of Statistics
University of Illinois
at Urbana-Champaign
725 S. Wright Street
Champaign, IL 61820
U.S.A.

March 1989

This research was sponsored in part by the Cognitive Science Program, Psychological Sciences Division, Office of Naval Research, under Contract N00014-87-K-0277, NR 4421-548. Approved for public release: distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States government.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION
Unclassified

1b RESTRICTIVE MARKINGS

2a SECURITY CLASSIFICATION AUTHORITY

2b DECLASSIFICATION / DOWNGRADING SCHEDULE

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release;
distribution unlimited

4 PERFORMING ORGANIZATION REPORT NUMBER(S)
ONR 89-1

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
Psychometric Group, Department
of Statistics - Univ. of Ill.

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
Cognitive Science Program

6c. ADDRESS *(City, State, and ZIP Code)*
Department of Statistics, U of Illinois
725 South Wright Street
Champaign, IL 61820

7b. ADDRESS *(City, State, and ZIP Code)*
Office of Naval Research (Code 1142PT)
Arlington, VA
22217-5000

8a. NAME OF FUNDING / SPONSORING ORGANIZATION

8b. OFFICE SYMBOL *(If applicable)*

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
N00014-87-K-0277

8c. ADDRESS *(City, State, and ZIP Code)*

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO.
61153N	RR04024	RR04204-01	4421-548

11. TITLE *(Include Security Classification)*
A New Item Response Theory Modeling Approach with Applications
to Unidimensionality Assessment and Ability Estimation

12. PERSONAL AUTHOR(S)
Stout, William

13a. TYPE OF REPORT
Technical Report

13b. TIME COVERED
FROM ________ TO ______

14. DATE OF REPORT *(Year, Month, Day)*
March 1989

15 PAGE COUNT
47

16 SUPPLEMENTARY NOTATION
To appear Psychometrika late 1989 or early 1990

17 COSATI CODES

FIELD	GROUP	SUB-GROUP
05	09	

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Local independence, essential independence, essential trait, intrinsic ability scale, marginal item response function, latent dimensionality, multidimensionality, (continued)

19 ABSTRACT *(Continue on reverse if necessary and identify by block number)*

Using an infinite item test framework, it is argued that the usual assumption of local independence be replaced by a weaker assumption, essential independence. A fortiori, the usual assumption of unidimensionality is replace by a weaker and arguably more appropriate statistically testable assumption of essential unidimensionality. Essential unidimensionality implies the existence of a "unique" unidimensional latent ability. Essential unidimensionality is equivalent to the "consistent" estimation of this latent ability in an ordinal scaling sense using any balanced linear formula scoring scheme. A variation of this estimation approach allows consistent estimation of ability on the given latent ability scale. A stochastic model for the "manufacture" of essentially unidimensional items is given.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL
Dr. Charles Davis

22b TELEPHONE *(Include Area Code)*
202-696-4046

22c OFFICE SYMBOL
ONR 1142 CS

(A)

ABSTRACT

Using an infinite item test framework, it is argued that the usual assumption of local independence be replaced by a weaker assumption, essential independence. A fortiori, the usual assumption of unidimensionality is replaced by a weaker and arguably more appropriate statistically testable assumption of essential unidimensionality. Essential unidimensionality implies the existence of a "unique" unidimensional latent ablity. Essential unidimensionality is equivalent to the "consistent" estimation of this latent ability in an ordinal scaling sense using any balanced linear formula scoring scheme. A variation of this estimation approach allows consistent estimation of ability on the given latent ability scale. (KR)

Key words: Local independence, essential independence, essential trait, intrinsic ability scale, marginal item response function, latent dimensionality, multidimensionality, essential unidimensionality, item response theory, latent trait theory, ability estimation, consistent estimation, nonparametric, balanced linear formula score, infinite item pool.

A New Item Response Theory Modeling Approach

Introduction. Until recently, most theoretical and applied item response theory (IRT) based research has uncritically assumed one of a small set of unidimensional, locally independent monotone parametric models; e.g., one-, two-, or three-parameter logistic and normal ogive models for a finite item test. (See Lord [1980] for a survey of this IRT modeling research tradition and Mislevy [1987] for a survey of current IRT modeling research.)

By contrast, this paper makes a determined case for assuming a monotone nonparametric (i.e., no specific functional form for item response functions assumed) infinite item pool IRT framework with local independence replaced by a less restrictive and, we claim, psychometrically more appropriate assumption, namely essential independence. Essential independence provides the basis for assessing the essential dimensionality of test data. Essential dimensionality, much in the spirit of counting the number of dimensions in a factor analytic model, is the number of major latent dimensions with minor dimensions ignored. Essential unidimensionality, the existence of exactly one major dimension, then provides a justification for carrying out IRT based statistical analyses that require unidimensionality. We favor the use of unidimensional IRT modeling approaches in applications when they are used subsequent to a careful statistical analysis verifying that essential unidimensionality fits the item response data sufficiently well. On the other hand, the uncritical use of the standard unidimensional three parameter logistic model in applications is the equivalent of Plato's cave dweller's attempt to interpret the outside world entirely on the basis of shadows cast on his cave wall (see Reckase, Carlson, Ackerman, and Spray [1986] and Wang [1988] for implications of the uncritical use of unidimensional models where multidimensionality holds).

Other nonparametric approaches appear in the literature. Mokken scaling, with its stress on the Loevinger homogeneity index, has received considerable attention.

See for example Mokken and Lewis (1982) and Sijtsma and Molenaar (1987). Cliff (1977, 1979) proposes a non latent trait approach stressing the degree of consistency of the order relationships between persons and between items. These approaches with their emphasis on test scalability, which is only peripherally related to essential unidimensionality, are not closely related to the nonparametric approach of this paper.

Suppose that one uses the herein proposed infinite item pool modeling approach and that essential unidimensionality is assumed — hopefully subsequent to a statistical analysis of essential dimensionality. It is established below that two major consequences follow: (i) the uniqueness of the latent ability in an ordinal scaling sense and (ii) the consistency of estimation of the unique latent ability. Thus, latent ability can be consistently estimated in the essentially independent, essentially unidimensional case, even if the usual local independence does not hold.

This paper continues the work of Stout (1987), where essential unidimensionality was first defined and a statistical test of essential unidimensionality proposed and its properties and performance investigated. Indeed, it is important to emphasize that for psychological test data the nonparametric, monotone, essentially independent, essentially unidimensional model of this paper can be tested for lack of statistical fit, as described in the 1987 paper.

Our paper is organized as follows: Section 1 reviews the traditional IRT modeling approach. Section 2 defines essential independence and essential dimensionality and presents basic properties. Section 3 considers the consistent estimation of ability, establishes the "uniqueness" of the latent trait, and introduces balanced linear empirical scoring. Section 4 considers a conceptual probabilistic framework for the generation of essentially unidimensional tests consisting of multidimensional items. Section 5 briefly discusses and summarizes the results of the paper.

1. <u>The Traditional IRT Modeling Approach</u>. According to the latent trait viewpoint, each examinee is indexed by a possibly vector valued variable $\underline{\theta}$, with many examinees permitted to be assigned to each $\underline{\theta}$. Associated with each item i is an <u>item response function</u> (IRF) $P_i(\underline{\theta})$ that denotes the probability that a randomly chosen examinee from the set of examinees with ability $\underline{\theta}$ will get the item right. (Various researchers perfer various interpretations of $P_i(\underline{\theta})$ — this one is our perference. See Hambleton and Swaminathan, 1985, pp. 26-27 for discussion.) Random sampling of examinees from a specified population induces a distribution $F(\underline{\theta})$ on the latent trait space of $\underline{\theta}$s and hence a distribution for the test response $\underline{U}_N \equiv (U_1,\cdots,U_N)$ of a randomly chosen examinee. The random test response vector $\underline{U}_N$ will often be referred to as the "test". Similarly, the random variable U_i will often be referred to as the ith "item". Observed values of $\underline{U}_N$ and U_i will be denoted by $\underline{u}_N$ and u_i respectively. $\underline{u}_N = (u_1,\cdots,u_N)$ will always be a sequence of 0s and 1s. $U_i = 1$ denotes a correct response and $U_i = 0$ denotes an incorrect response to item i for a randomly chosen examinee. The latent random vector is denoted by $\underline{\Theta}$ and particular values taken on by $\underline{\Theta}$ are denoted by $\underline{\theta}$. Note that $P_i(\theta) = P[U_i = 1|\underline{\Theta} = \underline{\theta}] = E[U_i|\underline{\Theta} = \underline{\theta}]$ for all i, $\underline{\theta}$. For notational convenience, let $P(\underline{u}_N|\underline{\theta})$ denote the conditional distribution $P[\underline{U}_N = \underline{u}_N|\underline{\Theta} = \underline{\theta}]$. It is important to stress that a "test" $\underline{U}_N$ can have many possible latent trait models. That is, there are many choices of the pair $F(\underline{\theta})$, $P(\underline{u}_N|\underline{\theta})$ such that, for all $\underline{u}_N$,

$$P[\underline{U}_N = \underline{u}_N] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} P(\underline{u}_N|\underline{\theta})\, dF(\underline{\theta}) \tag{1}$$

Note here that $(\underline{U}_N, \underline{\Theta})$ are a pair of random vectors whose joint distribution is specified by the marginal distribution $F(\underline{\theta})$ and the conditional distribution $P(\underline{u}_N|\underline{\theta})$. Latent models for $\underline{U}_N$ will be denoted by $(\underline{U}_N, \underline{\Theta}, P(\underline{u}_N|\underline{\theta}), F(\underline{\theta}))$ or for brevity by $(\underline{U}_N, \underline{\Theta})$.

Three characteristics of latent models are of considerable importance:

(i) The model $(\underline{U}_N, \underline{\Theta})$ is said to be monotone (M) if $P[U_{i_1} = 1, \cdots, U_{i_k} = 1 | \underline{\Theta} = \underline{\theta}]$ is nondecreasing in $\underline{\theta}$ for each subtest $(U_{i_1}, \cdots, U_{i_k})$ of $\underline{U}_N$ (here by definition $\underline{\theta}_1 \le \underline{\theta}_2$ if and only if $\theta_{1i} \le \theta_{2i}$ for each coordinate i; also $1 \le i_1 < \cdots < i_k \le N$). The model is said to be strictly monotone if "nondecreasing" is replaced by "strictly increasing."

(ii) The model $(\underline{U}_N, \underline{\Theta})$ is said to be d dimensional if $\underline{\Theta}$ is a d dimensional random vector. The d dimensional ability is then denoted by $(\theta_1, \cdots, \theta_d)$. The dimensionality of $\underline{\Theta}$ will be denoted by $\dim(\underline{\Theta})$ or d.

(iii) The model $(\underline{U}_N, \underline{\Theta})$ is said to be locally independent (LI) if

$$P(\underline{u}_N | \underline{\theta}) \equiv P[U_1 = u_1, \cdots, U_N = u_N | \underline{\Theta} = \underline{\theta}] = \prod_{i=1}^{N} P[U_i = u_i | \underline{\Theta} = \underline{\theta}] \qquad (2)$$

$$= \prod_{i=1}^{N} P_i(\underline{\theta})^{u_i} [1 - P_i(\underline{\theta})]^{1-u_i}$$

for all $\underline{\theta}$ and each of the 2^N choices of $(u_1, \cdots, u_N)$.

The most commonly used class of models has been the LI, M, d = 1 models. In this case M is equivalent to the IRFs all being monotone. Usually for models when M, d = 1 holds, the IRFs are typically strictly monotone. Note that in the LI, d=1 case with the latent distribution $F(\theta)$ having density $f(\theta)$ that (1) and (2) combine to produce the "usual" IRT model equation

$$P[\underline{U}_N = \underline{u}_N] = \int_{-\infty}^{\infty} \left\{ \prod_{i=1}^{N} P_i(\theta)^{u_i} [1 - P_i(\theta)]^{1-u_i} \right\} f(\theta)\, d\theta. \qquad (3)$$

2. <u>A New Conceptualization of Test Dimensionality</u>. Let us recall the <u>traditional IRT definition of test dimensionality</u> that almost always applies in IRT models:

<u>Definition 2.1</u>. The dimensionality d of a test $\underline{U}_N$ is the minimal dimensionality required for $\underline{\Theta}$ to produce a latent model $(\underline{U}_N, \underline{\Theta})$ that is both LI and M. □

Although mathematically appealing, this definition is rather impractical for mental testing because, in actual practice, individual test items clearly have multiple determinants of their respective probabilities of correct response. This position has been pursued clearly and vigorously by Humphreys (1984), who states:

> The related problems of dimensionality and bias of items are being approached in an arbitrary and oversimplified fashion. It should be obvious that unidimensionality can only be approximated. ··· The large amount of unique variance in items is not random error, although it can be called error from the point of view of the attribute that one is attempting to measure. ··· We start with the assumption that responses to items have many causes or determinants.

Humphreys (1984) points out that a dominant attribute (i.e., dominant dimension) results from an attribute overlapping many items and asserts that attributes common to relatively few items or even unique to individual items are unavoidable and indeed are not detrimental to the measurement of a dominant dimension. In his writings, Humphreys stresses that the low item intercorrelations typically observed argue strongly for viewing individual items as determined by multiply attributes. Although the existence of multiply determined items is rarely emphasized in the IRT literature, it is a theme with a long history in the factor analytic test theory literature. Classical factor analysis applied to binary test data of course implicity assumes the possibility of many determinants, allowing for many determinants specific to individual items in addition to one or more dominant dimensions. McDonald (1981) actually argues for the existence of "minor components" in factor analytic modeling of test data. That is, he argues for the existence of multiple determinants, many of which are common to relatively few items at most. Tucker, Koopman, and Linn (1969) have developed a factor analytic test simulation model that includes "minor factors" as well as dominant factors and unique factors.

Unfortunately, the traditional definition (Definition 2.1), based on local independence, makes no distinction between dominant and minor dimensions. Thus, if taken seriously, this definition compels us to take as test dimensionality the total number of all item dimensions rather than adopting the more appropriate "factor analytic viewpoint" by which only the number of dominant dimensions is counted. This is true <u>even</u> in situations with only one dominant dimension where, both from the viewpoint of psychmetric verity and of modeling parsimony, it would be desirable to ignore multiple determinants (i.e., minor and unique factors) and categorize tests as unidimensional. Thus the traditional definition requires us to assign dimensionality $d = d_0 > 1$ (d_0 possibly quite large in fact) in settings where it would be desirable to assign $d = 1$. For example, if all items of a long test depend on θ_1, but Items 1 and 2 alone also depend on θ_2, then $d = 2$.

Clearly, it is an important psychometric goal to be able to statistically assess whether or not a test $\underline{U}_N$ is driven by exactly one dominant dimension. As a necessary precursor, a mathematical conceptualization of the number of dominant dimensions is needed, namely the <u>essential dimensionality</u> of a collection of test items. In order to present a rigorous definition of essential dimensionality, it is necessary to conceptualize $\underline{U}_N$ as the initial <u>observed</u> segment of an infinite item pool $\{U_i, i \geq 1\}$. Here, it is assumed that whatever process has been used to construct the first N items of the pool making up the test $\underline{U}_N$ could be continued <u>in the same manner</u> by including further items from the pool. Thus the infinite item pool $\{U_i, i \geq 1\}$ is of the same dimensional character as $\underline{U}_N \equiv \{U_i, 1 \leq i \leq N\}$. If an actual item banking scheme with random sampling of items is being used to construct the test, then $\underline{U}_N$ and $\{U_i, i \geq 1\}$ being the same dimensional character is guaranteed. Such random sampling of items is often used for criterion referenced tests constructed from item banks; see for example Hambleton and Swaminathan (1985, Chapter 12). A latent model for $\{U_i, i \geq 1\}$ is denoted by $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$ or more completely by $\{\underline{U}_N, \underline{\Theta}, P(\underline{u}_N|\underline{\theta}), F(\underline{\theta}), N \geq 1\}$, thus emphasizing that adding successive items from the item pool generates a sequence of tests.

It will be assumed throughout the remainder of the paper that $\underline{U}_N$ consists of the first N items of an infinite item test $\{U_i. i \geq 1\}$. This will be referred to as the <u>infinite item pool formulation</u> of IRT. This proposed replacement of $\underline{U}_N$ for fixed N by $\{U_i, i \geq 1\}$ in IRT modeling is a specific instance of a standard and useful modeling device used throughout mathematical statistics. For example, in order to study the performance of estimating a population mean by the sample average, an infinite random sample $\{X_i, i \geq 1\}$ is often posited, thus enabling one to examine the asymptotic properties of $\bar{X}_N \equiv \sum_{i=1}^{N} X_i/N$ as $N \to \infty$.

Our modeling approach deliberately tolerates the test $\underline{U}_N$ (and hence the infinite item pool) containing an insignificant number of atypical items. For example every 2^k-th item of a "mathematics" item pool could be a verbal item. The allowing of an "insignificant number" of atypical items is facilitated by the introduction of the concept of a collection of nonsparse subtests: Consider the sequence of tests $\{\underline{U}_N, N \geq 1\}$ obtained from the item pool $\{U_i, i \geq 1\}$ by iterating $\underline{U}_{N+1} = \{\underline{U}_N, U_{N+1}\}$. A subtest of $\underline{U}_N$ will be denoted by $\mathcal{M}_N = \{U_{i_1}, U_{i_2}, \ldots, U_{i_{M(N)}}\}$ for each $N \geq 1$. Thus $M(N) \leq N$ denotes the length of the subtest $\mathcal{M}_N$. A particular collection of subtests $\{\mathcal{M}_N, N \geq 1\}$ is termed <u>nonsparse</u> if it is nested ($\mathcal{M}_N \subset \mathcal{M}_{N+1}$ for $N \geq 1$; that is, all items of $\mathcal{M}_{N+1}$ are also in $\mathcal{M}_N$) and there exists $\epsilon > 0$ such that

$$\frac{M(N)}{N} \geq \epsilon \tag{4}$$

for all $N \geq 1$. That is, (4) requires that the length of each subtest $\mathcal{M}_N$ of the collection must exceed a fixed possibly small proportion of the length N of the corresponding test $\underline{U}_N$. Roughly, a test is two dimensional if it has a subtest measuring a second ability different than the latent ability of interest. For our infinite item pool formulation, this translates into a sequence of subtests $\{\mathcal{M}_N, N \geq 1\}$ measuring a this second ability. The theory of essential unidimensionality then requires that these subtests be nonsparse and hence typical of the infinite item pool.

We next define a weaker type of independence than local independence called essential independence. The intuitive idea is that conditional on the latent random variable, the covariances between items are "small" on average, regardless of which collection of nonsparse subtests is being used.

Definition 2.2. The latent model $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$ is said to be essentially independent (EI) if for every collection of nonsparse subtests for each $\underline{\theta}$ in the range of $\underline{\Theta}$,

$$D_N(\underline{\theta}) \equiv \binom{M(N)}{2}^{-1} \sum_{i,j \in I_N, i<j} \operatorname{cov}(U_i, U_j | \underline{\Theta} = \underline{\theta}) \to 0 \quad \text{as } N \to \infty. \tag{5}$$

Remarks. (i) From a measure-theoretic probability perspective, we mean by $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$ that the random variables $\{U_i, i \geq 1\}$ and the random vector $\underline{\Theta}$ are defined on a common probability space. In this paper issues of measure-theoretic probability rigor, although always surmountable, are suppressed in the interest of clarity.

(ii) The content of (5) is that $\operatorname{cov}(U_i, U_j | \underline{\Theta} = \underline{\theta})$ must be small on average for a wide class of subtests, the collections of nonsparse subtests. For example taking $\mathcal{U}_N = \underline{U}_N$, an EI latent model must satisfy for each $\underline{\theta}$ in the range of $\underline{\Theta}$

$$\frac{\sum_{1 \leq i<j \leq N} \operatorname{cov}(U_i, U_j | \underline{\Theta} = \underline{\theta})}{\binom{N}{2}} \to 0 \tag{6}$$

as $N \to \infty$.

(iii) $\operatorname{Cov}(U_i, U_j | \underline{\Theta} = \underline{\theta}) \neq 0$ holds if and only if there is latent information beyond knowing $\underline{\Theta} = \underline{\theta}$ that influences examinee performance on the item pair (U_i, U_j).

(iv) Observe that a sufficient condition implying (5) for all collections of nonsparse subtests is that for each $\underline{\theta}$

$$D'_N(\underline{\theta}) \equiv \frac{\sum_{1 \leq i<j \leq N} |\operatorname{cov}(U_i, U_j) | \underline{\Theta} = \underline{\theta}|}{\binom{N}{2}} \to 0 \text{ as } N \to \infty. \tag{7}$$

It is informative to contrast the definition of essential independence given by (5) with the traditional latent trait conceptualization of local independence given in (2). LI implies independence of all pairs (U_i, U_j), given $\underline{\Theta}$, which is equivalent to $\operatorname{cov}(U_i, U_j | \underline{\Theta} = \underline{\theta}) = 0$ for all $\underline{\theta}$. By contrast, EI is a weaker assumption than LI and only requires that for each fixed $\underline{\theta}$, $\mathrm{cov}(U_i,U_j|\underline{\Theta} = \underline{\theta})$ is small in magnitude on average as the test length N grows. The psychometric interpretation of EI is that $\underline{\Theta}$ measures those individual examinee differences that are essential or dominant in influencing item pool performance. Whereas, $\underline{\Theta}$ must be augmented to $\underline{\Theta}_{LI} = (\underline{\Theta},\underline{\Theta}')$ in order that $\underline{\Theta}_{LI}$ measure all individual differences that influence any of the items of the item pool; that is LI holds for $\{\underline{U}_N,\underline{\Theta}_{LI},N \geq 1\}$. $\underline{\Theta}'$ here consists of dimensions that have an inessential or minor influence. For example a component of $\underline{\Theta}'$ might influence examinee performance on only one item.

From the mathematical viewpoint, it is not known whether it is possible to construct a latent model for which essential independence holds (i.e., (5) for nonsparse subtests) and yet (7) fails. Indeed, we conjecture but have not proved yet that under a very mild hypothesis (5) does imply (7). However, from the applied psychometric perspective, it is very plausible that (5) implies (7) in all realistic IRT models of useful tests: To see this, suppose (7) fails. For simplicity, suppose that (7) fails in the sense that there exists $\epsilon > 0$ and some value $\underline{\theta}$ of the latent variable $\underline{\Theta}$ for which $D_N'(\underline{\theta}) > \epsilon$ for <u>all</u> N, rather than the technically connect infinitely many N. Fix such a $\underline{\theta}$. Because $|\mathrm{cov}(U_i,U_j|\underline{\theta})| \leq 1$ for all (i,j) pairs and $D_N'(\underline{\theta}) > \epsilon$ for all N, it follows that for each N there must exist many item pairs (U_i,U_j), $1 \leq \ell \leq j \leq N$ for which

$$|\mathrm{cov}(U_i,U_j|\underline{\Theta} = \underline{\theta})| \geq \frac{\epsilon}{2}$$

Thus it is plausible that there exists a collection $\{\mathcal{A}_N,N \geq 1\}$ of nonsparse subtests for which the above inequality holds for all item pairs of the collection. Let, recalling the preceeding paragraph, $\underline{\Theta}_{LI} = (\underline{\Theta},\underline{\Theta}')$ be the augmentation of $\underline{\Theta}$ that produces LI for $\{\mathcal{A}_N,N \geq 1\}$. Item pairs from a set of items all with similar latent factor loadings on the components of $\underline{\Theta}_{LI}$ ("similar latent factor loadings" to be informally interpreted) tend to be non-negatively correlated conditional on a specific value of the latent vector $\underline{\Theta}$. Thus, it seems plausible that one can extract from $\{\mathcal{U}_N, N \geq 1\}$ a collection of nonsparse subtest $\{\mathcal{U}'_N, N \geq 1\}$; i.e., $\mathcal{U}'_N \subset \mathcal{U}_N$ for all N, for which $\mathrm{cov}(U_i, U_j | \underline{\Theta} = \underline{\theta}) > \frac{\epsilon}{2}$ for all items of $\{\mathcal{U}'_N, N \geq 1\}$. That is (5) fails for a collection of nonsparse subtests. Thus, it is, as claimed, very plausible that (5) imples (7).

Because it is plausible that (5) and (7) are equivalent and because (7) is so much easier to interpret, it is often useful to interpret EI using (7) instead of (5)—the reader is encouraged to do so when appropriate. Also, when it is necessary to prove mathematically that EI holds, often this is done by proving that (7) holds rather than (5).

Just as LI can be weakened to EI monotonicity can be weakened to to weak monotonicity:

Definition 2.3. $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$ is said to be weakly monotone (WM) if

$$\sum_{i=1}^{N} P_i(\underline{\theta}) \text{ is nondecreasing in } \underline{\theta} \tag{8}$$

for all N.

In most of the results of this paper, it will suffice to assume WM instead of M. Note that WM does allow for some nonmonotone IRFs in the item pool.

Now essential dimensionality can be defined.

Definition 2.4. The essential dimensionality d_E of a test $\{U_i, i \geq 1\}$ is the minimal dimensionality required for a latent trait $\underline{\Theta}$ to make the latent model $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$ an EI, WM model. When $d_E = 1$, essential unidimensionality is said to hold. If essential d_E dimensionality holds using ability $\underline{\Theta}$ then $\{U_i, i \geq 1\}$ is said to be essentially d_E dimensional with respect to ability $\underline{\Theta}$. Such a trait is called an essential trait for $\{U_i, i \geq 1\}$.

Remarks. (i) Although $d_E = 0$ is theoretically possible, it is psychometrically uninteresting and does not occur in well designed tests. Thus, to avoid irrelevant trivialities we assume $d_E \geq 1$ throughout this paper.

(ii) It is vital to note that the essential dimensionality (traditional dimensionality too) depends simultaneously on both the infinite item pool and the examinee population. The dimensionality of a test administration is determined by the interaction between items and examinees. The dimensionality of a test administration cannot be assigned without consideration of the examinee population as well as the test items. Nonetheless it is a linguistic convenience to refer to the "test dimensionality."

The theorems and examples in the remainder of Section 2 combine to produce a partial taxonomy of infinite item pools $\{U_i, i \geq 1\}$ for which essential unidimensionality holds. Many variations and combinations of these theorems and examples are easily derivable. In essence, it is shown that essential unidimensionality holds for an item pool $\{U_i, i \geq 1\}$ if

(i) only a "nondense" subsequence of the items depends on an ability (or abilities) other than the ability of interest,

(ii) each ability other than the ability of interest influences at most $K < \infty$ items and moreover these incidental abilities are "orthogonal" to each other, conditional on the ability of interest, or

(iii) the magnitude of the dependence of items on an ability (or abilities) other than the ability of interest is asymptotically negligible, even though a "dense" set of items of the test may depend on this other ability.

Note that in each case (i) - (iii) it is intuitively clear that there is one dominant dimension with possibly many minor dimensions; that is, essential unidimensionality clearly ought to hold.

Recall that we propose using $\{U_i, i \geq 1\}$ as a model for a given finite item test. From this viewpoint (i) - (iii) holding for $\{U_i, i \geq 1\}$ translates into (i′) - (iii′) respectively holding for the finite test $\underline{U}_N$:

(i′) Few items of $\underline{U}_N$ depend on an ability (or abilities) other than the ability of interest, (ii′) each ability other than the ability of interest influences at most a small number of items of $\underline{U}_N$ and moreover these incidental abilities are "orthogonal" to each other, conditional on the ability of interest,

(iii′) the magnitude of the dependence of the items of $\underline{U}_N$ on an ability (or abilities) other than the ability of interest is small, even though most of the items may depend on this other ability.

The purpose of the statistical test proposed in Stout (1987) is to assess for a test $\underline{U}_N$ administered to a population of examinees whether essential unidimensionality provides a good data fit—for example, because (i′) (ii′) or (iii′) holds. In Section 4 a model for test item construction is proposed, making explicit how we view the dimensional character of $\underline{U}_N$ as being the same as that of $\{U_i, i \geq 1\}$. This helps demonstrate the close relationship between (i) and (i′), between (ii) and (ii′), and between (iii) and (iii′).

The basic thrust of essential unidimensionality (recall Definition 2.4) is that the <u>same unidimensional</u> latent trait Θ makes the average conditional covariances of item pairs small for <u>all</u> collections of nonsparse subtests. This will hold even if a "nondense" subsequence of items in the infinite item test depend on an ability other than that intended to be measured, as the following example illustrates.

<u>Example 2.1</u>. Assume **WM**. Suppose every 2^kth item is a pure verbal (Θ_2) item and the rest are pure mathematics (Θ_1) items, with mathematics the ability intended to be measured. Thus we suppose for all θ_1 that

$$\mathrm{cov}(U_i, U_j \mid \Theta_1 = \theta_1) = 0$$

for all pairs (U_i, U_j) of mathematics items. An easy calculation shows that (7) and hence that (5) holds; i.e., essential unidimensionality holds. □

The above example is easily abstracted into a conceptually useful theorem. First, a set of indices $\{i_1 < i_2 < \cdots\}$ is said to be <u>nondense</u> if $\frac{i_k}{k} \to \infty$ as $k \to \infty$.

Theorem 2.1. Assume WM. Let $A = \{i_1, i_2, \cdots\}$ denote a nondense set of indices. Suppose for an item pool $\{U_i,\ i \geq 1\}$ there exists a unidimensional latent ability Θ for which for all θ

$$\sup_{i \notin A, j \notin A, |i-j| \geq N} \left| \operatorname{cov}\,(U_i, U_j | \Theta = \theta) \right| \to 0$$

as $N \to \infty$. Then essential unidimensionality with respect to Θ holds.

Remark. The hypothesis of Theorem 2.1 states except for item pairs coming from a nondense sequence of items, that the magnitude of covariances, conditional on Θ asymptotically approaches 0. For example, it might be that items with indices in A depend on other dimensions besides θ.

Proof. Fix θ. Fix $1 > \epsilon > 0$. Let Num (B) denote the cardinality of any set B. Let $A_N = \{i_k \text{ in } A \text{ such that } i_k \leq N\}$. Choose N_0 such that

$$\sup_{i \notin A, j \notin A, |i-j| \geq N_0} \left| \operatorname{cov}(U_i, U_j | \Theta = \theta) \right| \leq \epsilon, \quad \frac{\operatorname{Num}(A_N)}{N} \leq \epsilon$$

for all $N \geq N_0$. Split item pairs from $\{U_i,\ 1 \leq i \leq N\}$ up into those for which at least one item is in A_N; those for which $|i-j| < N_0$ and both $i \notin A_N$, $j \notin A_N$; and those for which $|i-j| \geq N_0$ and both $i \notin A_N$, $j \notin A_N$. Then, for $N > N_0$, noting that $|\operatorname{cov}(U_i, U_j | \Theta = \theta)| \leq 1$ for all i, j, θ,

$$\begin{aligned} \sum\nolimits_{1 \leq i < j \leq N} |\operatorname{cov}(U_i, U_j | \Theta = \theta)| &\leq \epsilon N^2 + NN_0 \\ &\quad + \sum\nolimits_{1 \leq i < j \leq N, |i-j| \geq N_0, i \notin A_N, j \notin A_N} |\operatorname{cov}(U_i, U_j | \Theta = \theta)| \\ &\leq \epsilon N^2 + NN_0 + \epsilon N^2. \end{aligned}$$

Then, $D_N(\theta) \leq 3\epsilon$ for N large, thus establishing (7) and hence establishing essential unidimensionality. □

The following example suggests how essential unidimensionality can fail when two psychometric dimensions are present.

Example 2.2. Modify Example 2.1 by making every 10k-th item a verbal item. Assume there exists $\epsilon > 0$ and θ_1 such that $\operatorname{cov}(U_i,\ U_j | \Theta_1 = \theta_1) \geq \epsilon$ for all pairs of verbal items. Consider $N = 10k$; $k = 1, 2, \cdots$. Then, it is easily seen that

$D_N(\theta_1) \longrightarrow 0$ as $N \to \infty$ and hence that (5) does not hold for all collections of nonsparse verbal subtests. The point is that $\{U_{10}, U_{20}, \cdots\}$ really is measuring a different dimension (verbal) from the intended dimension to be measured (mathematics). Hence the scaling $\sum_{k=1}^{N} U_{10k}/N$ is an empirical verbal scale, $\sum_{i=1}^{N} U_{2i+1}/N$ is an empirical mathematics scale, and $\sum_{i=1}^{N} U_i/N$ is a combined mathematics and verbal scale.

□

The following commonly occuring test setting illustrates that essential dimensionality (Definition 2.4) and traditional dimensionality (Definition 2.1) can differ considerably. It also illustrates a setting where essential unidimensionality is the result of (ii).

Example 2.3. Consider the construction of a paragraph comprehension test of length $N = 5n$, where n = number of paragraphs and each paragraph is followed by five related questions. Assume total independence between questions involving different paragraphs given Θ, where for convenience we think of Θ as reading ability. Suppose VM with respect to Θ. Note, using $|\mathrm{cov}(U_i,U_j|\Theta = \theta)| \leq 1$ for all i, j, θ that

$$|D_N(\theta)| \leq \frac{2}{N(N-1)} \sum_{1\leq i<j\leq N} |\mathrm{cov}(U_i,U_j|\Theta = \theta)|$$
$$\leq \frac{2}{N(N-1)}\left[\frac{N}{5}\right] \begin{bmatrix} 5 \\ 2 \end{bmatrix} \to 0 \text{ as } N \to \infty.$$

Thus essential unidimensionality holds, whereas a traditional dimensionality of $n + 1$ seems necessary for a test of length $N = 5n$. Reading ability Θ is the essential trait for this essentially unidimensional model. □

The example, by displaying a test that clearly should be psychometrically labeled as "unidimensional", illustrates our view that minor or idiosyncratic dimensions should be ignored in assessing test dimensionality from the applications viewpoint. Our requiring EI rather than LI is the key step that makes it possible to ignore minor dimensions in assessing dimensionality.

The following theorem makes (ii) precise and hence demonstrates one way in which essential unidimensionality holds.

<u>Theorem 2.2</u>. Let $\{U_i,\ i \geq 1\}$ be given. Suppose that local independence holds with respect to $\underline{\theta} = (\theta, \theta_1, \theta_2, \cdots)$. Suppose conditional on Θ that $(\Theta_1, \Theta_2, \cdots)$ are mutually independent. Suppose that each θ_i influences at most L items and that each item depends on at most K θ_is. Suppose VM holds for $\{\underline{U}_N, \Theta, N \geq 1\}$. Then essential unidimensionality holds with respect to Θ.

<u>Proof</u>. It suffices to prove (7). Consider $\underline{\theta} = (\theta, \underline{\theta}^{(1)}, \underline{\theta}^{(2)}, \underline{\theta}^{(3)})$, a splitting of $\underline{\theta}$ up into four subsets. Consider U_i, U_j for which U_i depends on $(\theta,\ \underline{\theta}^{(1)})$ and U_j depends on $(\theta,\ \underline{\theta}^{(2)})$ with no dependence on $\underline{\theta}^{(3)}$. Then, using the standard calculus for conditional probabilities,

$$\mathrm{cov}(U_i,\ U_j|\Theta) = E[\mathrm{cov}(U_i,U_j|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)})|\Theta] + \mathrm{cov}[E(U_i|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)}),$$
$$E(U_j|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)})|\Theta] = \mathrm{cov}[E(U_i|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)}),\ E(U_j|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)})|\Theta]$$

Here we have used the fact that by local independence and the nondependence on $\underline{\theta}^{(3)}$,

$$0 = \mathrm{cov}(U_i,U_j|\underline{\Theta}) = \mathrm{cov}(U_i,U_j|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)})$$

Now,
$$\begin{aligned}
&\mathrm{cov}[E(U_i|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)}),\ E(U_j|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)}|\Theta] \\
&= E[E(U_i|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)})E(U_j|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)})|\Theta] - \\
&\qquad E[E(U_i|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)})|\Theta]E[E[(U_j|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)})|\Theta] \\
&= E[E(U_i|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)})|\Theta]E[E(U_j|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)})|\Theta] - \\
&\qquad E[E(U_i|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)})|\Theta]E[E(U_j|\Theta,\underline{\Theta}^{(1)},\underline{\Theta}^{(2)})|\Theta] \\
&= 0.
\end{aligned}$$

The above factoring was allowed because $\underline{\Theta}^{(1)}$ and $\underline{\Theta}^{(2)}$ are independent given Θ by hypothesis. Thus, we have proved that when U_i depends on $(\Theta,\ \underline{\Theta}^{(1)})$ only and U_j depends on $(\Theta,\underline{\Theta}^{(2)})$ only that

$$\mathrm{cov}(U_i,U_j|\Theta) = 0. \qquad (9)$$

Recall that

$$\left|\mathrm{cov}(U_i,U_j|\Theta)\right| \leq 1$$

for arbitrary i, j. Thus, using (9) an upper bound for $D_N'(\Theta)$ of (7) is given by

$$\frac{1}{\binom{N}{2}} \cdot C$$

where C is the number of item pairs with indices $\leq N$ for which both items depend on at least one common θ_i. Consider a fixed item with index $\leq N$. It depends on at most K θ_is. For each such θ_i there are at most L items also dependent on θ_i. Thus the total items sharing at least one common θ_i is bounded above by LK. Thus the total number of item pairs C is bounded above by LKN. I.e.,

$$\frac{1}{\binom{N}{2}} C \leq \frac{2\ LKN}{N(N-1)} \to 0$$

as $N \to \infty$, as desired. Since **VM** holds by hypothesis, the theorem is proved.

□

Remark. The mathematical assumption of LI with respect to $\underline{\theta}$ in Theorem 2.2 simply means from the psychometric viewpoint that examinee performance is completely explained by $\underline{\theta}$

Theorem 2.3 now makes explicit one way in which (iii) above implies essential unidimensionality.

Theorem 2.3. Let $\{U_i,\ i \geq 1\}$ be given. Suppose that local independence holds with respect to $\underline{\theta} \equiv (\theta,\underline{\theta}^{(1)}) \equiv (\theta,\theta_1,\theta_2,\cdots)$. Define

$$\epsilon_i(\theta) \equiv \sup_{\underline{\theta}^{(1)}} P_i(\theta,\underline{\theta}^{(1)}) - \inf_{\underline{\theta}^{(1)}} P_i(\theta,\underline{\theta}^{(1)}).$$

Suppose that the dependence of items on $\underline{\theta}^{(1)}$ is asymptotically negligible in the sense that the IRFs $P_i(\underline{\theta})$ satisfy for every θ

$$\frac{1}{N} \sum_{i=1}^{N} \epsilon_i(\theta) \to 0 \tag{10}$$

as $N \to \infty$. Suppose **VM** holds for $\{\underline{U}_N,\Theta,N \geq 1\}$. Then essential unidimensionality holds with respect to θ.

Proof. It suffices to prove (7). For an arbitrary pair of items $(U_i,\ U_j)$, using standard probability calculus, by the assumption of local independence,

$$\left|\text{cov}(U_i,U_j|\Theta)\right| = \left|E[\text{cov}(U_i,U_j)|\underline{\Theta})\Theta] + \text{cov}[E(U_i|\underline{\Theta}),\ E(U_j|\underline{\Theta})|\Theta]\right|$$
$$= |\text{cov}[E(U_i|\underline{\Theta}),\ E(U_j|\underline{\Theta})|\Theta]|.$$

Since $\text{cov}(X,Y)^2 \leq \text{Var}(X)\ \text{Var}(Y)$ holds in general,

$$\begin{aligned}&|\mathrm{cov}[E(U_i|\underline{\Theta}),\ E(U_j|\underline{\Theta})|\Theta]|\\ &\le \mathrm{Var}^{1/2}[E(U_i|\underline{\Theta})|\Theta]\ \mathrm{Var}^{1/2}[E(U_j|\underline{\Theta})|\Theta]\\ &\equiv \mathrm{Var}^{1/2}[P_i(\underline{\Theta})|\Theta]\ \mathrm{Var}^{1/2}[P_j(\underline{\Theta})|\Theta].\end{aligned}$$

Combining,

$$|\mathrm{cov}(U_i,U_j|\Theta)| \le \mathrm{Var}^{1/2}[P_i(\underline{\Theta})|\Theta]\ \mathrm{Var}^{1/2}[P_j(\underline{\Theta})|\Theta]. \tag{11}$$

Since for $a_i \ge 0$, $\sum_{i\ne j} a_i a_j \le \left[\sum a_i\right]^2$, it follows that, using (11),

$$\frac{1}{\binom{N}{2}} \sum_{1\le i<j\le N} |\mathrm{cov}(U_i,U_j|\Theta)| \le \frac{1}{\binom{N}{2}} \left[\sum_{i=1}^{N} \mathrm{Var}^{1/2}[P_i(\underline{\Theta}|\Theta)]\right]^2 .$$

Because $0 \le a \le X \le b$ implies $\mathrm{Var}\ X \le (b - a)^2$, it follows that

$$\frac{1}{\binom{N}{2}} \sum_{1\le i<j\le N} |\mathrm{cov}(U_i,U_j|\Theta)| \le \frac{1}{\binom{N}{2}} \left[\textstyle\sum_{i=1}^{N} \epsilon_i(\theta)\right]^2.$$

Thus (7) follows by (10). □

3. Application to Consistent Estimation of Ability. We turn now to the problem of estimating a particular latent ability θ of interest in the presence of other ("nuisance") abilities. The act of specifying an IRT model usually includes the choice of a specific latent ability scale $\underline{\theta}$ and hence a specific scale for the ability of interest θ. Throughout this paper "θ scale" will refer to the scale for the ability of interest that is predetermined by the specific IRT model chosen. The following dichotomy holds for ability estimation: (i) Because the θ scale has been established by its use in other test settings, because the θ scale is linked to some external criterion, or because the θ scale has an intrinsic theoretical justification, etc., one may insist on estimating the ability of interest using the θ scale; or (ii) nothing about the test setting makes the θ scale or any other scale particularly preferable to use. That is, in the case of (ii), any strictly increasing transformation $A(\theta)$ yields an equally acceptable scale for the purpose of estimating ability. In Case (ii), it is thus appropriate to choose, for statistical convenience perhaps, a particular scaling $A(\theta)$ on the real line with apparent interval scaling properties as long as inferences drawn depend only on the

ordinal nature of the real line. The point is well discussed in Section 1.6 of Lord and Novick, 1968. In summary Lord and Novick state "A major problem of mental test theory is to determine a good interval scaling to impose when the supporting theory implies only ordinal properties." For example, LOGIST and BILOG, likely the two most commonly used ability estimation programs, both create such an "interval" ability scale. Mislevy (1987) stresses that from a pure IRT model fit viewpoint (that is, without extraneous requirements such as Rasch's specific objectively), only ordinal scaling properties are defensible. We will refer in this paper to the creation of a convenient ability scale in the absence of prior scaling constraints as ordinal scaling.

The modeling framework we adopt to investigate ability estimation is the WM, EI infinite item pool framework $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$ introduced in Section 2. It is assumed that the ability of interest Θ is determined by $\underline{\Theta}$; that is, that Θ is a function of $\underline{\Theta}$. Mathematically speaking, there may well exist item pools $\{U_i, i \geq 1\}$ for which $d_E = \infty$. But, psychometrically this is unrealistic: Recall that it is assumed that the $\{U_i, i > N\}$ is constructed by continuing the same process that produced $\underline{U}_N$ and thus that $\{U_i, i \geq 1\}$ will be of the same dimensional character as $\underline{U}_N$. Taking this into consideration, from the psychometric viewpoint it is clear for virtually all tests to be modeled that it is realistic to assume at most a finite number of dominant latent dimensions. Thus we assume for our infinite item pool framework that $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$ has $d_E < \infty$ with the assurance that this assumption is not psychometrically restrictive.

In order to illuminate certain theoretical issues most clearly, in Section 3.1 ability estimation will be considered first in its simplest setting, namely where only ordinal scaling (recall (ii) above) is required. Then, in Section 3.2 we consider ability estimation for the perhaps more useful case where use of the θ scale is desirable or required.

3.1 Estimation of ability in the ordinal scaling case. We first select a natural scale for the ability θ of interest. Let

$$P_i(\theta) = E[P_i(\underline{\Theta})|\Theta = \theta] \qquad (12)$$

where $P_i(\underline{\theta})$ is the ith item response function of the latent model $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$, Θ is the random latent variable of interest, and the conditional expectation in (12) is with respect to the conditional distribution of $\underline{\Theta}$, given $\Theta = \theta$.

The following probabilistic calculation justifies calling $P_i(\theta)$ in (12) the ith marginal item response function with respect to ability θ: By definition

$$P_i(\underline{\theta}) = P[U_i = 1|\underline{\Theta} = \underline{\theta}].$$

Thus

$$\begin{aligned} E\{P_i(\underline{\Theta})|\Theta = \theta\} &= E\{P[U_i = 1|\underline{\Theta}]|\Theta = \theta\} \\ &= P[U_i = 1|\Theta = \theta], \end{aligned}$$

this last equality holding since expectation is a "projection operator." We have thus proved that for all θ,

$$P_i(\theta) = P[U_i = 1|\Theta = \theta]. \qquad (13)$$

That is, $P_i(\theta)$, while defined as the probabilistic average of getting Item i correct with averaging over all $\underline{\theta}$ for which $\Theta = \theta$, is also an item response function in the ordinary sense of being the probability of getting Item i correct for a randomly sample examinee of ability $\Theta = \theta$.

Let

$$A_N(\theta) = \textstyle\sum_{i=1}^N P_i(\theta)/N \qquad (14)$$

$A_N(\theta)$ is called the intrinsic ability scale for θ relative to the test $\underline{U}_N$ and the examinee population $\underline{\Theta}$. $A_N(\theta)$ has an interpretation bridging classical test theory and IRT: $A_N(\theta)$ is the expected test score, that is, true score, among all examinees with latent ability θ. Under the assumption of strict monotonicity for $\sum_{i=1}^N P_i(\underline{\theta})/N$ of the latent model, a slight modification of Theorem 3.1 below implies that $A_N(\theta)$ is strictly increasing in θ and hence is an acceptable scale for estimating the ability of interest θ in the ordinal scaling case. As such, $A_N(\theta)$

for fixed N is our recommended choice of scale in the ordinal scaling case. It helps to recall that the assumption of VM is that $\sum_{i=1}^{N} P_i(\underline{\theta})$ is monotone.

Considerable recent attention has been focused on nonmonotone unidimensional item response functions. It has been shown that "attractive distractors" are a source of nonmonotonicity in multiple choice items. It has been suggested that the existence of attractive distractors may be explainable by multidimensionality of the ability space. In this regard, it is interesting to note that a multidimensional item response function $P(\underline{\theta})$ can be monotone and yet the corresponding marginal item response function $P(\theta)$ be nonmonotone:

Example 3.1. Let $P(\theta_1, \theta_2) = (\theta_1 + \theta_2)/17$, $1/4 \le \theta_1 \le 1$, $0 \le \theta_2 \le 16$. Define the conditional distribution of Θ_2 given $\Theta_1 = \theta_1$ by $f(\theta_2|\theta_1) = \theta_1/4$ if $0 \le \theta_2 \le 4/\theta_1$; $= 0$ otherwise.

Then the marginal item response function with respect to θ_1 is given by

$$P(\theta_1) = \int_0^{4/\theta_1} \left[\frac{\theta_1 + \theta_2}{68}\right] \theta_1 \, d\theta_2$$

$$= \left[\frac{\theta_1}{2} + \frac{1}{\theta_1}\right] \frac{2}{17} \qquad 1/4 < \theta_1 < 1.$$

But $P(\theta_1)$ is decreasing in θ_1 for all θ_1. □

Of course, as is intuitively clear, mild and natural regularity conditions preclude this nonmonotone behavior. Indeed, the nonmonotonicity of marginal item response function can occur only when the multidimensional ability $\underline{\theta}$ has some sort of negative association among its components, as Theorem 3.1 below makes clear by setting $N = 1$.

Definition 3.1. A random vector $\underline{Y}$ is said to be stochastically larger than a random vector $\underline{X}$ if, for all $\underline{t}$,

$$P[\underline{X} \ge \underline{t}] \le P[\underline{Y} \ge \underline{t}] \tag{15}$$

with strict inequality for at least one $\underline{t}$.

The following fact is well known:

Lemma 3.1. Let $\underline{Y}$ be stochastically larger than $\underline{X}$, and let f be a nonnegative real valued function that is nondecreasing in each of it arguments. Then

$$Ef(\underline{X}) \leq Ef(\underline{Y})$$

Theorem 3.1. Let $\sum_{i=1}^{N} P_i(\underline{\theta})$ for the latent model $(\underline{U}_N, \underline{\Theta})$ be monotone. Let $\underline{\Theta} \equiv (\Theta, \underline{\Theta}^{(2)})$. Suppose that, for every $\theta' < \theta''$ pair of real numbers that the distribution of $\underline{\Theta}^{(2)}$ given $\Theta = \theta''$ is stochastically larger than the distribution of $\underline{\Theta}^{(2)}$ given $\Theta = \theta'$. Then, $A_N(\theta)$ is monotone.

Proof. It must be shown for each i that

$$\int \sum_{i=1}^{N} P_i(\theta, \underline{\theta}^{(2)}) dP(\underline{\theta}^{(2)} | \theta)$$

is nondecreasing in θ where $P(\underline{\theta}^{(2)} | \theta)$ denotes the distribution of $\underline{\Theta}^{(2)}$, given $\Theta = \theta$. Fix real numbers $\theta' < \theta''$. By Lemma 3.1, noting that $\sum_{i=1}^{N} P_i(\theta, \underline{\theta}^{(2)})$ is for fixed θ a nondecreasing function of $\underline{\theta}^{(2)}$ by the assumption of monotonicity,

$$\int \sum_{i=1}^{N} P_i(\theta', \underline{\theta}^{(2)}) dP(\underline{\theta}^{(2)} | \theta') \leq \int \sum_{i=1}^{N} P_i(\theta', \underline{\theta}^{(2)}) dP(\underline{\theta}^{(2)} | \theta''). \qquad (16)$$

But, because $\sum_{i=1}^{N} \equiv P_i(\theta, \underline{\theta}^{(2)})$ is nondecreasing in θ for each fixed $\underline{\theta}_2$,

$$\int \sum_{i=1}^{N} P_i(\theta', \underline{\theta}^{(2)}) dP(\underline{\theta}^{(2)} | \theta'') \leq \int \sum_{i=1}^{N} P_i(\theta'', \underline{\theta}^{(2)}) dP(\underline{\theta}^{(2)} | \theta''). \qquad (17)$$

The combination of (16) and (17) yields the desired result. □

Remark. If a LI model $(\underline{U}_N, \underline{\Theta})$ has $\sum_{i=1}^{N} P_i(\underline{\theta})$ monotone then Theorem 3.1 makes clear that assuming $A_N(\theta)$ monotone is a mild assumption.

We now turn directly to the ability estimation problem in the ordinal scaling case. As we shall see, essential unidimensionality characterizes the consistent estimation of the unidimensional latent ability on some scale; moreover, it implies, from the ordinal scaling viewpoint, that the latent ability is unique. It is in this spirit that an essential trait (recall Definition 2.4) can be referred to as the essential trait with respect to which the items are essentially unidimensional.

Theorem 3.2 below asserts that essential unidimensionality is precisely the condition needed for consistent estimation of ability. Before stating Theorem 3.2, we must carefully state what it means to consistently estimate ability using an infinite item pool formulation. Recall our ordinal scaling viewpoint that any strictly monotone transformation of θ — for example $A_N(\theta)$, which is strictly monotone when the marginal IRFs are — is an acceptable scale on which to estimate θ.

Definition 3.2. (i) It is said that Θ may be consistently estimated (in probability) if for every collection of nonsparse subtests, for each θ, given $\Theta = \theta$

$$\left[\frac{1}{M(N)} \sum_{j=1}^{M(N)} U_{i_j}\right] - \left[\frac{1}{M(N)} \sum_{j=1}^{M(N)} P_{i_j}(\theta)\right] \to 0 \qquad (18)$$

in probability as $N \to \infty$.

Remarks. Specialized to the present setting, the traditional statistical definition of consistency applies separately to each estimator $\delta_N(\underline{U}_N)$ of θ as follows: $\delta_N(\underline{U}_N)$ is a consistent estimator of θ if for all θ, given $\Theta = \theta$,

$$\delta_N(\underline{U}_N) \to \theta$$

in probability as $N \to \infty$. By contrast, our psychometric notion of item pool consistency proposed by *Definition* 3.2 *is:* (i) stronger than the traditional definition in that convergence is required simultaneously for all collections of nonsparse subtests; (ii) weaker in the sense that the required convergence for each scaling $\{\delta_N(\underline{U}_N) \equiv \sum_{j=1}^{M(N)} U_{i_j}/M(N)\}$ is not to θ but rather to a convenient rescaling of θ that varies with $\{\delta_N(\underline{U}_N)\}$, and hence the psychometric notion of consistency is an ordinal scaling concept; and (iii) different in the sense that statistical consistency is a property that each estimator $\{\delta_N(\underline{U}_N),\ N \ge 1\}$ either has or doesn't have while psychometric consistency is a property that the infinite item pool (that is the sequence $\{U_i,\ i \ge 1\}$) either has or doesn't have.

The intuitive idea of Definition 3.2 is that any reasonable collection of subtests should be able to estimate θ in our ordinal scaling sense. Not all collections of subtests need necessarily be usable to estimate θ in the ordinal scaling sense. For example if every 2^kth item in a "mathematics" test were a "verbal" item, then $\delta_{2^M} \equiv \sum_{m=1}^{M} U_{2^m}/M$ would estimate verbal ability. But this is developed from a nondense collection of subtests and hence is not required to asymptotically estimate θ for consistency to hold. However, if every 10th item were a verbal item, then Definition 3.2 implies that consistency does not hold because obviously the corresponding subtests are nonsparse.

A reasonable question to ask is whether in formulating our definition of test consistency, it would suffice to merely require for each θ that, given $\Theta = \theta$,

$$\bar{U}_N - \sum_{i=1}^{N} P_i(\theta)/N \to 0 \tag{19}$$

in probability as $N \to \infty$ instead of requiring (18). However merely requiring (19) is vacuous: Any test $\underline{U}_N$ of fixed length N can be appropriately embedded in an essentially unidimensional sequence of tests $\{\underline{U}_N, \Theta, N \geq 1\}$ for which (19) holds for a judicious choice of θ. The following example (suggested by remarks of D. R. Divgi) illustrates this embedding for a test where 50% of the items measure one trait and 50% measure a second trait. It presents an essentially two dimensional infinite item pool where (19) is satisfied for a mathematically judicious choice of θ, θ being some function of the dimensions (θ_1, θ_2). This is a situation where most psychometricians would prefer to split the test up into two unidimensional tests and only then address the issue of consistency. Requiring (18) instead of (19) as a definition of consistency reflects this consideration.

Example 3.2. Let $\{\underline{U}_K, \underline{\Theta}, K \geq 1\}$ be a LI, M latent model with $\underline{\theta} = (\theta_1, \theta_2)$ and for $i \geq 1$,

$$P_{2i}(\underline{\theta}) = \theta_1,\ P_{2i-1}(\underline{\theta}) = \theta_2$$

where the distribution of $\underline{\Theta}$ is given by Θ_1, Θ_2 independent identically distributed with Θ_1 uniformly distributed on $[0,1]$. Let $\Theta = \Theta_1 + \Theta_2$. Fix $\theta < 1$. Then, standard multivariable calculus yields for $1 \leq i, j$; $1 \leq k$

$$E(U_k | \Theta = \theta) = \frac{\theta}{2},\ \mathrm{cov}(U_{2i-1}, U_{2j} | \Theta = \theta) = -\frac{\theta^2}{12} \tag{20}$$

and, if $i \neq j$,

$$\mathrm{cov}(U_{2i}, U_{2j} | \Theta = \theta) = \frac{\theta^2}{12},\ \mathrm{cov}(U_{2i-1}, U_{2j-1} | \Theta = \theta) = \frac{\theta^2}{12}.$$

Thus, using (20),

$$\mathrm{Var}(\bar{U}_{2K}|\Theta = \theta) = \frac{1}{(2K)^2}\left[\sum_{i=1}^{2K} \mathrm{Var}(U_i|\Theta = \theta) + \sum_{1\le i\ne j\le 2K} \mathrm{cov}(U_i,U_j|\Theta = \theta)\right]$$
$$= \frac{1}{(2K)^2}\left[(2k)\ \frac{\theta}{2}\left[1 - \frac{\theta}{2}\right] + \frac{2\theta^2}{12}\left[\binom{K}{2} + \binom{K}{2} - K^2\right]\right]$$
$$= \frac{1}{2K}\ \frac{\theta}{2}\left[1 - \frac{\theta}{2}\right] - \frac{\theta^2}{24K} \to 0 \quad \text{as} \quad K \to \infty.$$

Further, $E[\bar{U}_{2K}|\Theta = \theta] = \theta/2$. Thus $P[|\bar{U}_{2K} - \frac{\theta}{2}| > \epsilon\ |\Theta = \theta] \le \mathrm{Var}(\bar{U}_{2K})/\epsilon^2 \to 0$ as $K \to \infty$. Hence, for each θ, given $\Theta = \theta$,

$$\bar{U}_{2K} - \frac{\theta}{2} \to 0$$

in probability as $K \to \infty$. A similar analysis holds for $\bar{U}_{2K-1}$ and also for $1 \le \theta \le 2$. Thus (19) does hold in what is clearly an essentially two dimensional sequence of tests. However, it is intuitively clear that (18) fails since the even items can be averaged to consistently estimate θ_1 and odd items to consistently estimate θ_2. □

Remark. Example 3.2 illustrates that $\bar{U}_N$ will estimate a mixture of the essential dimensions when they exist in asymptotically fixed proportions as N increases.

Now Theorem 3.2 can be stated and proved.

Theorem 3.2. Let $\{U_i, i \ge 1\}$ be essentially unidimensional $(d_E = 1)$ with respect to ability Θ. Then, θ may be consistently estimated. In particular, for each given $\Theta = \theta$, (19) holds.

Conversely, if for some VM latent model $\{\underline{U}_N, \Theta, N \ge 1\}$ the unidimensional θ may be consistently estimated then $\{\underline{U}_N, \Theta, N \ge 1\}$ is an EI, VM representation and hence $d_E = 1$ holds.

Proof. Assume $d_E = 1$ with repect to θ. Let $\{\mathcal{M}_N\} = \{(U_{i_1}, \ldots, U_{i_{M(N)}})\}$ be a collection of nonsparse subtests. Fix $\epsilon > 0$ and θ. Then

$$P\left[\left|\frac{1}{M(N)}\sum_{j=1}^{M(N)} U_{i_j} - \frac{1}{M(N)}\sum_{j=1}^{M(N)} P_{i_j}(\theta)\right| > \epsilon|\Theta = \theta\right] \le \mathrm{Var}\left[\frac{1}{M(N)}\sum_{j=1}^{N} U_{i_j}|\Theta = \theta\right]/\epsilon^2 \quad (21)$$

because $E[\frac{1}{M(N)}\sum_{j=1}^{M(N)} U_{i_j}|\Theta = \theta] = \frac{1}{M(N)}\sum_{j=1}^{M(N)} P_{i_j}(\theta)$. But, noting by nonsparseness that $N/M(N) \le C < \infty$ for all N and that

$$\frac{1}{M(N)}\sum_{i=1}^{M(N)} \mathrm{Var}(U_i|\Theta = \theta) \le 1 \tag{22}$$

it follows that

$$0 \le \mathrm{Var}\left[\frac{1}{M(N)}\sum_{j=1}^{M(N)} U_{i_j}|\Theta = \theta\right] = \frac{1}{M(N)^2}\sum_{i=1}^{M(N)} \mathrm{Var}(U_i|\Theta = \theta) + \frac{M(N)-1}{M(N)} D_N(\theta) \le \frac{C}{N} + D_N(\theta) \to 0 \tag{23}$$

by the assumption of essential unidimensionality. Combining with (21) then establishes test consistency, as desired.

Conversely suppose θ may be consistently estimated. Fix θ and a collection of nonsparse subtests $\{\mathcal{M}_N\}$. Then, by (23)

$$D_N(\theta) \ge -\frac{C}{N}\,. \tag{24}$$

Thus, $D_N(\theta)$ cannot have any negative limit points.

For any bounded random variable X, denoting the bound by a (i.e., $|X| \le a$) the well-known converse to Chebychev's inequality states that

$$P[|X| \ge \epsilon] \ge \frac{E(X^2) - \epsilon^2}{a^2}\,. \tag{25}$$

Let

$$X_N = \frac{1}{M(N)}\left[\sum_{j=1}^{M(N)} U_{i_j} - \sum_{j=1}^{M(N)} P_{i_j}(\theta)\right]$$

Then noting that $|X_N| \le 1$,

$$P[|X_N| > \epsilon|\Theta = \theta] \ge E|X_N|^2|\Theta = \theta) - \epsilon^2 \ge \mathrm{Var}(X_N|\Theta = \theta) - \epsilon^2. \tag{26}$$

By consistency, $X_N \to 0$ in probability as $N \to \infty$. Thus,

$$P[|X_N| > \epsilon|\Theta = \theta] \to 0 \quad \text{as} \quad N \to \infty.$$

Thus, using (23) and (26), for each $\epsilon > 0$,

$$\frac{1}{M(N)^2}\sum_{i=1}^{M(N)} \mathrm{Var}(U_i|\Theta = \theta) + \frac{M(N)-1}{M(N)} D_N(\theta) - \epsilon^2$$

has no positive limit points. But, by (22) and the fact that $M(N) \to \infty$ it then follows that $D_N(\theta)$ has no positive limit points. Since it was also concluded that $D_N(\theta)$ has no negative limit points, $D_N(\theta) \to 0$ as $N \to \infty$. Thus $d_E = 1$ has been established, as desired. □

<u>Remark</u>. It is interesting to note that Theorem 3.2 guarantees the consistent estimation of ability even if the IRFs are unknown to the practitioner. That is, use of $\bar{U}_N$ to estimate $\sum_1^N P_i(\theta)/N$ <u>does not</u> require knowledge of $\sum_1^N P_i(\theta)/N$. As long as no attempt is being made to establish a standardized ability scale across tests (e.g., as a precurser to equating tests) knowledge of the IRFs is not required.

It is a foundationally relevant fact that essential unidimensionality implies under a mild and natural regularity condition for $\{U_i, i \geq 1\}$ that the latent ability is, in the ordinal sense, unique, as Theorem 3.3 below asserts.

<u>Definition 3.3</u>. Let $\{U_i, i \geq 1\}$ be essentially unidimensional with respect to ability Θ and let R be the range of Θ (i.e., $P[\Theta \subset R] = 1$ with R "minimal"). Suppose for every θ_1 in R that there exists $\epsilon_{\theta_1} > 0$ and an open neighborhood H_{θ_1} of θ_1 such that for all $\theta_2 \in H_{\theta_1}$ in the range R of Θ that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{P_i(\theta_2) - P_i(\theta_1)}{\theta_2 - \theta_1} \geq \epsilon_{\theta_1} > 0 \text{ for all } N. \tag{27}$$

Then $\{\underline{U}_N, \Theta, N \geq 1\}$ is said to be <u>locally asymptotically discriminating</u> (LAD) with respect to Θ.

<u>Remarks</u>. What LAD really supposes is that $\sum_{i=1}^{N} P_i(\theta)/N$ is increasing faster than some positive-slope linear function in some neighborhood of θ for every θ, independent of N. LAD guarantees that <u>on average</u> the items of the test $\{U_i, i \geq 1\}$ are sufficiently discriminating locally with respect to θ. Note that LAD is a strengthening of WM.

<u>Theorem 3.3</u>. Suppose $\{U_i, i \geq 1\}$ is essentially unidimensional with respect to both Θ and Θ'. Let the corresponding marginal item response functions be denoted by

$$P_i(\theta) = E[U_i | \Theta = \theta], \quad P_i'(\theta) = E[U_i | \Theta' = \theta]$$

for all θ. Suppose $\{U_i, i \geq 1\}$ is LAD with respect to Θ. There then exists a function g defined on the range R' of Θ' such that

$$\theta = g(\theta'), \ g \ \text{nondecreasing}$$

and the range of g is $\mathbb{R}$.

Remarks. (i) Theorem 3.3 states that in the sense of ordinal scaling, that all scales with LAD holding are the same. In this precise sense the latent variable is unique.

(ii) Since a $d = 1$, VM, LI model is also an EI model, note that Theorem 3.3 holds for $d = 1$, LAD, LI models as well. Thus Theorem 3.3 may be of interest even if one does not wish to use EI in IRT modeling.

Proof of Theorem 3.3. By Theorem 3.2, for each θ and θ'

$$\bar{U}_N - A_N(\theta) \to 0 \tag{28}$$

in probability given $\Theta = \theta$ (and hence on any subset of $\Theta = \theta$) and

$$\bar{U}_N - A'_N(\theta') \to 0 \tag{29}$$

in probability given $\Theta' = \theta'$ (and hence on any subset of $\Theta' = \theta'$) where

$$A_N(\theta) = E[\bar{U}_N | \Theta = \theta] \quad \text{and} \quad A'_N(\theta') = E[\bar{U}_N | \Theta' = \theta'].$$

Let

$$G_{\theta,\theta'} = [\Theta = \theta] \cap [\Theta' = \theta']$$

for all θ, θ'. Then, for each θ, θ' such that $G_{\theta,\theta'} \neq \phi$, (28) and (29) imply on $G_{\theta,\theta'}$ that

$$A_N(\theta) - A'_N(\theta') \to 0. \tag{30}$$

Fix $\theta' \in \mathbb{R}'$ and let, denoting the empty set by ϕ,

$$B_{\theta'} = \{\theta \in \mathbb{R} | G_{\theta,\theta'} \neq \phi\}.$$

Note that $B_{\theta'} \neq \phi$ for each $\theta' \in \mathbb{R}$ because each examinee has an ability value for both Θ and Θ'. Suppose $\theta_1 \neq \theta_2$ with $\theta_1 \in B_{\theta'}$, $\theta_2 \in B_{\theta'}$ and $\theta_2 > \theta_1$ without loss of generality. Then (30) implies that

$$A_N(\theta_2) - A_N(\theta_1) \to 0 \ \text{as} \ N \to \infty.$$

That is,

$$\sum_{i=1}^{N} \frac{P_i(\theta_2) - P_i(\theta_1)}{N} \to 0,$$

contradicting (27). Thus $B_{\theta'}$ consists of a unique $\theta \in \mathbb{R}$ for each θ': i.e., a function g is defined:

$$\theta = g(\theta') \quad \text{for all} \quad \theta' \in \mathbb{R}'.$$

Choose $\theta_2' > \theta_1'$ with $\theta_1' \in \mathbb{R}'$, $\theta_2' \in \mathbb{R}'$. Then define

$$\theta_2 = g(\theta_2'),\ \theta_1 = g(\theta_1').$$

Now,

$$A_N'(\theta_2') - A_N'(\theta_1') \geq 0$$

because the essentially unidimensional model $\{\underline{U}_N, \Theta', N \geq 1\}$ is WM. By the definition of g, recalling (30), it follows that $A_N(\theta_2) - A_N(\theta_1)$ has no negative limit points. Thus $\theta_2 \geq \theta_1$ by monotonicity of $A_N(\theta)$. That is, g is monotone nondecreasing and well defined for all $\theta' \in \mathbb{R}'$.

Because $[\Theta' = \theta'] \subset [\Theta = g(\theta')]$ the probability space, say Ω, satisfies

$$\Omega = \bigcup_{\theta' \in \mathbb{R}'} (\Theta' = \theta') \subset \bigcup_{\theta' \in \mathbb{R}'} (\Theta = g(\theta'))$$

and Ω can be partitioned

$$\Omega = \bigcup_{\theta \in \mathbb{R}} (\Theta = \theta),$$

it follows that the range of g is $\mathbb{R}$. □

<u>Remarks</u>. (i) Note that Theorem 3.3 does not claim that g is strictly increasing. That is, the rescaling given by g could assign many θ' to the same θ. Because no assumption analogous to (27) was made for Θ', this is of course expected. For, the Θ' scale could produce a finer partition of the latent ability space than needed to achieve essential unidimensionality. Thus the collapsing of distinct θ' into a single θ by $g(\theta')$ cannot be ruled out. For example, if for the Θ' scale there exists an interval [a,b] such that,

$$\frac{d}{d\,\theta'} P_i'(\theta') = 0 \quad \text{for all} \quad i, \quad \theta' \in [a,b]$$

then the Θ' scale should be rescaled so that all $\theta' \in [a,b]$ should be collapsed to a single point, say θ_a'. However, assuming (27) for θ' as well does imply a strictly increasing g.

(ii) In a private communication, Brian Junker has pointed out that an alternate proof of Theorem 3.3 can be given that produces g explicitly. See Junker (1988) for details.

(iii) Note that the infinite item pool formulation was essential for establishing the uniqueness of ability scale. It is the author's position that an infinite item pool formulation greatly aids the study of many foundational IRT issues. Indeed, that is a major point of this research.

(iv) For unidimensional EI and hence *a fortiori* for unidimensional LI models Theorem 3.3 formalizes the well known notion that the Θ scale is not completely pinned down. Indeed unless there are solid psychometric grounds for preferring one interval scale over the rest, the situation is one of a unique ordinal scale with the choice of a convenient interval scale left up to the practitioner to be decided on pragmatic grounds.

It is an often quoted axiom of psychometrics that a "test" should be unidimensional. If not, it should be broken up into a battery of unidimensional subtests, each to be analyzed separately. Thus, in the context of this paper, the axiom becomes that a test should be essentially unidimensional.

Theorem 3.2 has an interesting multidimensional analogue. Denote the multidimensional IRFs by $\{P_i(\underline{\theta}), 1 \geq 1\}$.

Theorem 3.4. Suppose essential d_E dimensionality with respect to ability $\underline{\Theta}$ for $\{U_i, i \geq 1\}$. Then, $\underline{\theta}$ is able to be consistently estimated in probability in the sense of (18) with θ replaced by $\underline{\theta}$ in (18).

Suppose that the essential dimensionality exceeds d for $\{U_i, i \geq 1\}$. Then there does not exist a d dimensional $\underline{\Theta}$ such that $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$ is WM and for each collection of nonsparse subtests for all $\underline{\theta}$, given $\underline{\Theta} = \underline{\theta}$

$$\left(\frac{1}{M(N)} \sum_{j=1}^{M(N)} U_{i_j}\right) - \left(\frac{1}{M(N)} \sum_{j=1}^{M(N)} P_{i_j}(\underline{\theta})\right) \to 0 \qquad (31)$$

in probability as $N \to \infty$.

Proof. Analogous to that of Theorem 3.2 and omitted. □

__Remark__. Assume essential d_E dimensionality with respect to $\underline{\theta}$ for some $d_E > 1$. Then according to Theorem 3.4 every collection of nonsparse subtests estimates a __unidimensional__ latent scale in the sense that for each $\underline{\theta}$ given $\underline{\Theta} = \underline{\theta}$, (31) holds. Let $g_{A,N}(\underline{\theta})$ denote the latent scale $\sum P_{i_j}(\underline{\theta})/M(N)$ of (31). Consider two different collections of nonsparse subtests resulting in two different choices of $g_{A,N}(\underline{\theta})$. Then the two different gs may be ranking examinees on the basis of two (or more) totally different abilities, which is unacceptible from the viewpoint of requiring test consistency. For, consistency requires that all of the gs indeed rank examinees on the basis of the __same__ unidimensional latent ability. Thus the unidimensionality of each $\{g_{A,N}(\underline{\theta}), N \geq 1\}$ in (31) is useless because it masks an inherent and unacceptable (from the practical viewpoint) multidimensionality.

Essential unidimensionality for the essential trait θ clearly guarantees the consistent estimation of the unique latent trait θ by the results of Section 3.1. The practitioner who wishes to estimate θ then needs to assess for a particular test $\underline{U}_N$ administered to a particular population whether it is reasonable to adopt an essentially unidimensional model. The author (Stout, 1987) has developed a statistical procedure specifically designed to assess whether essential unidimensionality holds or not. This procedure is based on a test statistic that basically is large or small according as $D_N(\cdot)$ of (7) is large or small. Thus the issue of essential unidimensionality is an __empirically verifiable__ one. It is the position of this paper to recommend that the usually untested assumption of a LI, d = 1, M IRT model be replaced by testing whether the more realistic $d_E = 1$ IRT modeling approach fits the data well or not.

3.2. __Balanced linear empirical scaling__. Sections 2 and 3.1 combine to produce a theory that applies to scoring examinees using proportion correct over all collections of nonsparse subtests. It seems useful to generalize this to a theory that applies to __all__ reasonable linear formula scoring schemes instead of just proportion correct. We can do so by a minor modification of the concepts of essential independence, essential unidimensionality, and consistency. In each case the term "strong" will be used to distinguish these concepts as defined here in Section 3.2 from Sections 2 and 3.1.

Definition 3.4. (i) A triangular array of item coefficients $\{a_{Ni}\} \equiv \{a_{Ni}, 1 \leq i \leq N, N \geq 1\}$ is said to be balanced if there exists $C > 0$ such that

$$0 \leq a_{Ni} \leq \frac{C}{N} \tag{32}$$

for all i and N.

(ii) For a given infinite item pool $\{U_i, i \geq 1\}$, the linear formula scoring sequence $\left\{\sum_{i=1}^{N} a_{Ni}U_i, N \geq 1\right\}$ is called a balanced empirical ability scaling provided $\{a_{Ni}\}$ is balanced.

Remarks. (i) Definition 3.4 needs interpretation. First, given any triangular coefficient array $\{a_{Ni}\}$, $\sum_{i=1}^{N} a_{Ni}U_i$ specifies an ability scale for the test $\underline{U}_N$ in the sense that it ranks examinees. This ability scale is not a latent ability scale: it scales examinees entirely on the basis of their item response patterns and hence is an empirical (manifest) ability scale. Assumption (32) guarantees that the scaling is reasonable in that (a) each correct answer can only increase examinee rank and (b) no single item is allowed to dominate the scaling. Intuitively, as made precise below, $\sum_{i=1}^{N} a_{Ni}U_i$ for large N can be thought of as an empirical scale that approximates some unidimensional latent scale.

(ii) Several special cases of linear formula scores are balanced. First $a_{i,N} = 1/N$ for $i \leq N$, yielding $\{\bar{U}_N, N \geq 1\}$, is clearly balanced. That is, proportion correct is balanced. Second it is easy to modify the proportion correct empirical scaling by using only items from nonsparse subtests to form the proportion correct scaling. To motivate the third scoring scheme, suppose a two parameter logistic model for $\{U_i, i \geq 1\}$ with discrimination parameters a_i satisfying

$$0 < \epsilon \leq a_i \leq K \text{ for all } i.$$

Then, the normalized sufficient statistic

$$\frac{\sum_{i=1}^{N} a_i U_i}{\sum_{i=1}^{N} a_i}$$

is clearly balanced with $a_{i,N} = a_i / \sum_{i=1}^{N} a_i$.

(iii) In addition to the special cases discussed in (ii), Junker (1988) has shown that the theory of balanced linear scores plays a central role in his establishing the robustness result that the traditional maximum likelihood estimator $\hat{\Theta}$ of θ is consistent (in the statistical sense) for θ even when only EI is known to hold rather than LI.

Definition 3.5. Strong essential independence holds if for every balanced $\{a_{Ni}\}$, for each $\underline{\theta}$, given $\underline{\Theta} = \underline{\theta}$,

$$\sum_{1 \le i \ne j \le N} a_{Ni} a_{Nj} \operatorname{cov}(U_i, U_j \mid \underline{\Theta} = \underline{\theta}) \to 0. \tag{33}$$

Remark. (33) is easily seen to be equivalent to

$$\operatorname{Var}\left[\sum_{i=1}^{N} a_{Ni} U_i \mid \underline{\Theta} = \underline{\theta}\right] \to 0 \quad \text{as} \quad N \to \infty \tag{34}$$

for each θ. Note that (7) implies strong essential independence, which in turn implies essential independence as defined by (5). Recall that (7) and (5) can be considered the same for practical purposes. Hence (7), (5), and strong essential independence can be thought of as the same for practical purposes. It follows that (33) and hence (34) can be thought of as characterizing essential independence with respect to $\underline{\Theta}$.

When one views essential independence as (33) and hence as (34) holding, then for each balanced $\{a_{Ni}\}$, by Chebychev's well-known probability inequality,

$$\sum_{i=1}^{N} a_{Ni} U_i - g_N(\underline{\theta}) \to 0$$

in probability as $N \to \infty$ for some latent scaling $g_N(\underline{\theta})$. Thus balanced empirical scalings do, as suggested above, approximate some latent scale. Thus (33) emphasizes the ability of any admissible empirical scaling to "recover" a latent scale. This will be expanded upon in Section 3.3.

Let $A \equiv \{a_{Ni}\}$ denote an arbitrary balanced sequence. Define the notation $g_{A,N}(\theta)$ for a unidimensional θ by

$$g_{A,N}(\theta) = \sum_{i=1}^{N} a_{Ni} P_i(\theta) \tag{35}$$

Definition 3.6. It is said that Θ may be strongly consistently estimated if for every balanced empirical scaling $\{\sum_{i=1}^{N} a_{Ni} U_i, N \geq 1\}$, for each θ, given $\Theta = \theta$,

$$\sum_{i=1}^{N} a_{Ni} U_i - g_{A,N}(\theta) \to 0 \tag{36}$$

in probability as $N \to \infty$.

Remark. Clearly strong consistency implies consistency as defined in Section 3.1. It is not known whether consistency can hold and strong consistency not hold.

Replacing collections of nonsparse subtests by balanced linear formula scores yields a slightly modified version of Theorem 3.2, which is stated below. First we define strong essential dimensionality.

Definition 3.7. (i) An item pool is said to be weakly monotone for all balanced sequences (WMB) with respect to $\underline{\theta}$ if

$$\sum_{i=1}^{N} a_{Ni} P_i(\underline{\theta})$$

is monotone for all N and for every balanced $\{a_{Ni}\}$.

(ii) The strong essential dimensionality d_E of a an item pool $\{U_i \; i \geq 1\}$ is the minimal dimensionality required for a latent trait $\underline{\Theta}$ to make the latent model $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$ a strongly essentially independent WMB model. When $d_E = 1$ in the strong sense, strong essential unidimensionality is said to hold.

Theorem 3.5. Let $\{U_i, i \geq 1\}$ be strongly essentially unidimensional with respect to ability Θ. Then θ may be strongly consistently estimated. In particular, for each given $\Theta = \theta$, (18) and (19) hold.

Conversely, if for some monotone latent model $\{\underline{U}_N, \Theta, N \geq 1\}$ the unideminsional θ may be strongly consistently estimated, then $\{\underline{U}_N, \Theta, N \geq 1\}$ is a strongly essentially independent, WMB model and hence strong essential unidimensionality holds.

Proof. Analogous to that of Theorem 3.2. Omitted. □

3.3. Estimation of ability when the ability scale is specified. The use of $\{\bar{U}_N\}$ or more generally of a balanced empirical scale $\{\sum_{i=1}^{N} a_{i,N} U_i\}$ as a sequence of estimators of θ in the ordinal scaling case was developed in Sections 3.1 and 3.2. Estimation in the ordinal scaling sense is inappropriate when it is either desirable or required that the θ scale be used for the ability of interest. In many such applications, the items (or at least a common core of them) have been calibrated relative to the constructed standardized ability scale θ. Estimation of θ with known IRFs has been widely treated in the literature (see for example Hambleton and Swaminathan, 1985, Section 5.3). Maximum likelihood estimation (MLE) is one method of choice in this setting. The MLE $\hat{\Theta}$ congerges in probability to θ under suitable regularity conditions in the sense that, given $\Theta = \theta$, $\hat{\Theta} \to \theta$ in probability as the number of items $N \to \infty$. Only rarely however, is it possible to provide a simple formula for the MLE as a function of $\underline{U}_N$. The MLE is usually a highly non-linear function of $\underline{U}_N$. Thus in the case of known IRFs it seems desirable to seek alternatives to MLE that are based on linear formula scoring and for which simple formulae are available. We now propose a family of such estimators, using the results of Sections 3.1.

Let

$$A_N(\theta) = \sum_{i=1}^{N} P_i(\theta)/N, \quad \tilde{A}_N(\theta) = \sum_{i=1}^{N} a_{Ni} P_i(\theta)$$

and

$$\bar{U}_N = \sum_{i=1}^{N} U_i/N, \; \tilde{U}_N = \sum_{i=1}^{N} a_{Ni}\, U_i.$$

Recall from Theorem 3.2 that when $\{U_i, i \geq 1\}$ is essentially unidimensional with respect to Θ then for each given $\Theta = \theta$,

$$\bar{U}_N - A_N(\theta) \to 0$$

in probability and $N \to \infty$. This suggests estimating θ by $\{A_N^{-1}(\bar{U}_N)\}$ and also suggests for each given $\Theta = \theta$ that

$$A_N^{-1}(\bar{U}_N) \to \theta$$

in probability and $N \to \infty$ should hold. Moreover, recalling Theorem 3.5 this result should generalize to balanced scoring with, for each $\Theta = \theta$,

$$\tilde{A}_N^{-1}(\tilde{U}_N) \to \theta$$

in probability and $N \to \infty$. Theorem 3.6 below states that this is true provided a slightly modified local asymptotic discrimination holds. Definition 3.8 is the appropriate analogue of Definition 3.3.

Definition 3.8. Let an infinite item pool be essentially unidimensional with respect to ability Θ. Let $\tilde{A}_N(\theta) = \sum_{i=1}^{N} a_{ni} P_i(\theta)$ be formed from a balanced sequence $\{a_{Ni}\}$. Suppose for every fixed θ_1 such that θ_1 is in the range $\mathbb{R}$ of Θ that there exists $\epsilon_{\theta_1} > 0$ and an open neighborhood H_{θ_1} of θ_1 such that for all $\theta_2 \in H_{\theta_1}$ and in the range $\mathbb{R}$ of Θ that

$$\frac{\sum_{i=1}^{N} a_{iN} P_i(\theta_2) - \sum_{i=1}^{N} a_{iN} P_i(\theta_1)}{\theta_2 - \theta_1} \geq \epsilon_{\theta_1} \quad \text{for all } N. \tag{37}$$

Then $\{\underline{U}_N, \Theta, \tilde{A}_N(\theta), N \geq 1\}$ is said to be locally asymptotically discriminating (LAD) with respect to Θ and $\{a_{Ni}\}$.

Usually $\tilde{A}_N(\theta)$ is continuous in applications, thus making its inverse well defined over its range. However, in order to have a theory that allows for discontinuities, the following definition of $\tilde{A}_N^{-1}(u)$ will be used

$$\tilde{A}_N^{-1}(u) \equiv \inf_{\theta \in \mathbb{R}} \{\theta : \tilde{A}_N(\theta) \geq u\}.$$

Here $\mathbb{R}$ denotes the range of Θ. Note that $\tilde{A}_N(\theta) = -\infty$ or ∞ is possible; e.g., if $u = 1/5$ and $A_N(\theta) \geq 1/4$ for all θ.

Theorem 3.6. Let $\{\underline{U}_N, \Theta, N \geq 1\}$ be strongly essentially unidimensional with respect to Θ. Suppose $\tilde{A}_N(\theta) = \sum_{i=1}^{N} a_{ni} P_i(\theta)$ is formed from a balanced ability scaling $\tilde{U}_N = \sum_{i=1}^{N} a_{ni} U_i$. Suppose $\{\underline{U}_N, \Theta, \tilde{A}_N(\theta), N \geq 1\}$ is LAD with respect to Θ and $\{a_{Ni}\}$. Then, for each given $\Theta = \theta$,

$$\tilde{A}_N^{-1}(\tilde{U}_N) \to \theta \tag{38}$$

in probability as $N \to \infty$.

Proof. Fix θ. By Theorem 3.5, given $\Theta = \theta$,

$$\tilde{U}_N - \tilde{A}_N(\theta) \to 0 \tag{39}$$

in probability as $N \to \infty$. It is an elementary lemma of probability theory that $X_N \to X$ in probability as $N \to \infty$ if and only if each subsequence $X_{N(j)}$ contains a further subsequence $X_{N(j(k))} \to X$ with probability one as $k \to \infty$. Thus to prove the theorem, it suffices to select an arbitrary subsequence $\{N(j)\}$ and then prove there exists a further subsequence $N(j(k))$ such that

$$\tilde{A}_{N(j(k))}^{-1}(\tilde{U}_{N(j(k))}) \to \theta$$

with probability one as $k \to \infty$. Choose $\{N(j)\}$. Then (39) implies that

$$\tilde{U}_{N(j)} - \tilde{A}_{N(j)}(\theta) \to 0$$

in probability as $j \to \infty$. Then, using the above mentioned lemma, there exists a further subsequence $N(j(k))$ for which

$$\tilde{U}_{N(j(k))} - \tilde{A}_{N(j(k))}(\theta) \to 0 \tag{40}$$

with probability one. By (37) and the definition of the inverse, for all $\mu_2 - \tilde{A}_N(\theta)$ sufficiently small in magnitude and satisfying $\mu_2 > \inf_\theta \tilde{A}_N(\theta)$, there exists $K_\theta < \infty$ such that

$$|\tilde{A}_N^{-1}(\mu_2) - \theta| \le K_\theta|\mu_2 - \tilde{A}_N(\theta)| \quad \text{for all} \quad \theta. \tag{41}$$

Fix a typical point in the probability space. Now, it may be that for some arbitrarily large k

$$\tilde{U}_{N(j(k))} \le \inf_\theta \tilde{A}_{N(j(k))}(\theta). \tag{42}$$

By (40) and LAD, there exists $\epsilon_k \to 0$ such that for all large k

$$\tilde{U}_{N(j(k))} > \tilde{A}_{N(j(k))}(\theta - \epsilon_k).$$

Thus $\tilde{U}_{N(j(k))} > \inf_\theta \tilde{A}_{N(j(k))}(\theta)$ for all sufficiently large k using LAD. Hence (42) cannot hold for arbitrarily large k and thus (41) can be applied with $\mu_2 = \tilde{U}_{N(j(k))}$.

Thus, combining (40) with (41),

$$|\tilde{A}^{-1}_{N(j(k))}(\bar{U}_{N(j(k))}) - \theta| \le K_\theta|\bar{U}_{N(j(k))} - \tilde{A}_{N(j(k))}(\theta)| \to 0$$

with probability one, as required. □

<u>Remarks</u>. (i) Theorem 3.6 provides a large class of sequences of estimators of θ, including $\{A_N^{-1}(\bar{U}_N)\}$, based on linear formula scoring. In practice, one needs to compute $\tilde{A}_N(\theta)$ and its inverse $\tilde{A}_N^{-1}(\theta)$ to make use of one of these estimators.

(ii) It is to be noted that Holland, Junker, and Thayer (1987) have proposed using $\{A_N^{-1}(\bar{U}_N)\}$ to estimate the <u>distribution</u> of Θ and have proved a convergence in distribution result to justify this. Their motivation for suggesting $\{A_N^{-1}(\bar{U}_N)\}$ is different from ours.

(iii) It is elementary to show that (38) holding for all θ implies

$$\tilde{A}_N^{-1}(\bar{U}_N) \to \Theta \tag{43}$$

in probability as $N \to \infty$. Given the IRT context, (38) is perhaps a more interesting formulation than (43). It does of course follow from (43) that $\tilde{A}_N^{-1}(\bar{U}_N)$ can be used as a method of estimating the distribution of Θ.

(iv) Note that (38) states that convergence in probability to individual ability holds <u>regardless</u> of which of a large class of estimators is used. That is, convergence in probability to individual ability holds for every balanced scaling. Theorem 3.5 and 3.6 show how close the traditional statistical notion of consistency and our psychometric notion of consistency really are. By Theorem 3.5, strong consistency is equivalent to strong essential unidimensionality, which by Theorem 3.6 implies that a wide class of natural estimates is consistent in the ordinary statistical sense.

(v) A version of Theorem 3.6 is possible that only deals with $\bar{U}_N$:

<u>Theorem 3.6′</u>. Let $\{\bar{U}_N,\theta,N \ge 1\}$ be EI and LAD with respect to θ. Then, for each gevin $\theta = 0$,

$$A_N^{-1}(\bar{U}_N) \to \theta$$

in probability as $N \to \infty$.

4. <u>A Stochastic Model for the Construction of Essentially Unidimensional Tests</u>.

In Section 2 and 3, the case has been made for using essentially unidimensional IRT models both in applications to real test data and in the investigation of theoretical issues. This approach leads to the uniqueness of the unidimensional latent ability scale and the consistent estimation of latent ability both in the ordinal and the fixed scaling case. This new modeling approach requires the replacement of modeling a fixed finite length test $\underline{U}_N$ by modeling an infinite item pool $\{U_i, i \geq 1\}$. $\{U_i, i \geq 1\}$ is the test that would result were one to continue constructing items $\{U_i, i > N\}$, "in the same manner" as $\underline{U}_N \equiv \{U_i, 1 \leq i \leq N\}$ was constructed. It has been stressed throughout that essential unidimensionality is an empirically testable property, using Stout's (1987) statistical test of unidimensionality.

Now it seems appropriate to present and study a model showing a plausible way in which essentially unidimensional infinite item pools $\{U_i, i \geq 1\}$ can be constructed. The actual observed test $\underline{U}_N$ is then obtained by terminating the process of constructing items after N items have been obtained.

We assume Θ is the ability to be measured and that the items also depend on finitely or infinitely many other abilities $(\Theta_1, \Theta_2, \ldots)$ and that the resulting ability space $\underline{\Theta} = (\Theta, \Theta_1, \Theta_2, \ldots)$ is "complete" in the sense that $\underline{\Theta}$ explains the variation between individuals in item/test performance. That is, we assume that $\{U_i, i \geq 1\}$ has a WM, LI IRT model $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$ where $\underline{\Theta} = (\Theta, \Theta_1, \ldots)$. This assumption is neither mathematically nor psychometrically restrictive.

Assume thoughout Section 4, consistent with Theorem 3.1, that $(\underline{U}_N, \Theta, N \geq 1)$ is an essentially unidimensional WM model. It is assumed that the representation for $\underline{\Theta}$ is orthogonal in the sense that all Θ_j, $\Theta_{j'}$ pairs are independent given Θ. This assumption basically amounts to choosing an orthogonal coordinate system for the latent ability space and hence is not unduly restrictive.

It is assumed that all items are inherently multiply determined, as motivated by discussion in Section 2. That is, each item can depend on one or more of the other abilities Θ_i besides Θ. Consider the construction of the ith item. Let

$$p_{ij} = P[\text{Item } i \text{ depends on } \Theta_j].$$

The assumption of multiply determined items then translates into the assumption that for each i

$$p_{ij} > 0 \text{ for some (possibly many) } j \text{ can hold.}$$

Implicit in the introduction of the p_{ij}s is the assumption that the determination of which abilities in addition to Θ influence each item can be viewed as a random process. (This is not the same as saying it is inherently a random process like radioactive decay — for example the digits of π can from the statistical perspective be well-modeled by a random process). That is, the "context" of each item can be viewed as randomly determined.

There are <u>deliberately</u> no model *assumptions* made here about other characteristics of the items such as discrimination, difficulty, guessing, etc. Also, no assumptions are made about the amount of item dependencies on various dimensions Θ_i, although such a refinement is possible and could be helpful. Because in a typical aptitude or achievement test, different items are often written by different individuals and item selection is controlled by factors such as discrimination, difficulty, congruence with intended content domain Θ, etc., the assumption of random item context together with no assumptions about item characteristics seems appropriate, even if there is no explicit random mechanism for choosing item context.

Weak and natural restrictions placed on the magnitudes of the $\{p_{ij}\}$ suffice to guarantee that essential unidimensionality holds; as is now established. Let

$$N_i = \text{Number of abilities besides } \theta \text{ influencing item } i$$

and

$$N_{\theta_j} = \text{Number of the first } N \text{ items dependent on } \theta_j.$$

Theorem 4.1. Suppose for $\{U_i, i \geq 1\}$ that

$$E(N_i) \leq K < \infty \text{ for all } i \tag{44}$$

and that for all j,

$$\frac{E(N_{\theta_j})}{N} \leq \epsilon_N \to 0 \text{ as } N \to \infty \tag{45}$$

Suppose independence of the "assignment" of abilities to item pairs in the sense that for all j

$$P[\text{Ability } \Theta_j \text{ assigned to item } i \text{ and to item } i'] = p_{ij}p_{i'j} \tag{46}$$

Then, essential unidimensionality with respect to Θ holds in the precise sense that for each θ, $D_N(\theta)$ of (7) satisfies $D_N(\theta) \to 0$ in probability as $N \to \infty$.

Remark. Here, "in probability" refers to the random process partially specified by the p_{ij}s that determines which θ_js influence which items. Thus according to Theorem 4.1, for fixed large N, "most" of the infinite item pools constructed will produce a "small" $D_N(\theta)$.

Proof. According to the proof of Theorem 2.2, $\text{cov}(U_i, U_{i'}|\Theta = \theta) = 0$ for all θ unless U_i, $U_{i'}$ depend on at least one common Θ_j. Thus, using (46)

$$E[|\text{cov}(U_i, U_{i'}|\Theta = \theta)|] \leq P[U_i, U_{i'} \text{ depend on some } \Theta_j] \leq \sum_{j=1}^{\infty} p_{ij}p_{i'j}.$$

Note that

$$E(N_i) = \sum_{j=1}^{\infty} p_{ij}, \; E(N_{\theta_j}) = \sum_{i=1}^{N} p_{ij}.$$

Thus,

$$E\left[\frac{1}{\binom{N}{2}} \sum_{1\leq i<i'\leq N} |\text{cov}(U_i, U_{i'}|\Theta = \theta)|\right]$$

$$\leq \frac{2}{N(N-1)} \sum_{1\leq i<i'\leq N} \sum_{j=1}^{\infty} p_{ij}p_{i'j}$$

$$\leq \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{\infty} p_{ij} \sum_{i'=1}^{N} p_{i'j}$$

$$\leq \frac{2}{N(N-1)} NKN\epsilon_N \to 0$$

as $N \to \infty$. But $X_N \geq 0$, $EX_N \to 0$ implies $X_N \to 0$ in probability holds for arbitrary random variables $\{X_N\}$. □

Suppose the designers of the test are deliberately being careful not to let too many pairs of items depend on any one θ_j, in an effort to create contextual balance. Then clearly assumption (46) is inappropriate and should be replaced. An appropriate assumption is for all i, i', j that

$$P[\text{Ability } \Theta_j \text{ assigned to Item } i \text{ and Item } i'] \le p_{ij}p_{i'j}. \tag{47}$$

Clearly the above proof is valid if (47) replaces (46). This yields:

<u>Corollary 4.1</u>. Suppose for $\{U_i, i \ge 1\}$ that (44), (45), and (47) hold. Then essential unidimensionality with respect to Θ holds in the sense stated in Theorem 4.1 □

There is a deterministic version of Corollary 4.1 that makes no assumptions about randomness and generalizes Theorem 2.2.

<u>Corollary 4.2</u>. Suppose for $\{U_i, i \ge 1\}$ that the dependence of abilities on items is such that

$$N_i \le K < \infty \quad \text{for all} \quad i \tag{48}$$

and that for all j

$$\frac{N_{\theta_j}}{N} \le \epsilon_N \to 0 \quad \text{as} \quad N \to \infty. \tag{49}$$

Then essential unidimensionality with respect to θ holds.

<u>Proof</u>. $\text{cov}(U_i, U_{i'} | \Theta = \theta) = 0$ unless U_i, $U_{i'}$ depend on at least one common Θ_j. Thus a simple counting argument yields

$$\frac{1}{\binom{N}{2}} \sum_{1 \le i < i' \le N} |\text{cov}(U_i, U_{i'} | \Theta = \theta)|$$

$$\le \frac{2}{N(N-1)} NKN\epsilon_N \to 0 \quad \text{as} \quad N \to \infty.$$ □

<u>Remark</u>. The mild hypotheses of Corollaries 4.1 and 4.2 suggest one strategy for essentially unidimensional test construction in the face of multiply determined items: Keep the number of abilities per item as low as possible (see (44) or (48)); keep the number of items influenced by any one ability other than θ as low as possible (see (45) or (49); and, subject to these constraints, keep the number of item pairs assigned to each ability other than θ as low as possible (see (46) or

(47)). This last constraint can either be accomplished by a random or pseudorandom assignment of abilities to items so that (46) tends to hold, or a deliberate effort can be made to balance, in an experimental design sense, the assignment of minor abilities to various items in the sense that (47) tends to hold.

There is an unavoidable paradoxical aspect to the model of this section and indeed to the infinite item pool model of this paper. In order to arrive at a <u>rigorous and useful</u> conception of essential unidimensionality it was necessary to replace modeling of the observable finite length test $\underline{U}_N$ by the unobservable (except for its initial segment $\underline{U}_N$) infinite length pool $\{U_i, i \geq 1\}$. But in order to apply this new modeling framework, one must assess the degree to which essential unidimensionality well-models the observable test $\underline{U}_N$ rather than the unobservable item pool $\{U_i, i \geq 1\}$. To address this paradox, recall that essential unidimensionality holds provided for all θ

$$D_N(\theta) \to 0 \quad \text{as} \quad N \to \infty.$$

Thus, the practitioner needs to assess, based on actual test data for $\underline{U}_N$ whether, for the actual test length N, for all θ

$$D_N(\theta) \approx 0 \tag{50}$$

A close examination of the statistical test for essential unidimensionality in Stout (1987) — see Section 5 and Formula (17) of that paper in particular — shows that the test is designed precisely to assess the degree to which (50) holds for nonsparse subtests. The Monte Carlo simulations presented in that paper justify using the statistical test of essential unidimensionality for ability tests as short as 25 items with as few as 750 examinees.

The model of this section with its large number of parameters $\{p_{ij}\}$ is intended for conceptual purposes and is not intended to facilitate analyses of real test data. However consequences of the model such as Theorem 4.1, Corollary 4.1, and Corollary 4.2 may be useful, as remarked, as guides to good unidimensional test construction.

5. Discussion and Summary of Results.

The purpose of the paper is to present a new IRT modeling approach based on the embedding of the test $\underline{U}_N$ into an infinite item pool $\{U_i, i \geq 1\}$ and then to show the usefulness of this approach to certain fundamental test measurement topics such as dimensionality and ability estimation. The paper provides a new conceptualization of latent dimensionality, essential dimensionality. This conceptualization depends on the replacement of local independence by the weaker and, in our opinion, psychometrically more appropriate notion of essential independence. Essential dimensionality, designed to dovetail with the empirical reality of multiply determined items, attempts to count only the dominant dimensions. Theorems 2.1–2.3 present conditions that guarantee that essential unidimensionality holds. In particular, dimensions distributed nondensely over items or dimensions having a minor influence on possibly many items do not negate essential unidimensionality.

In Section 3.1, essential unidimensionality is shown in Theorem 3.2 to characterize the consistent estimation of a unidimensional ability in the ordinal scaling case: The ordinal scaling case holds when any monotone transformation of the given ability scale is an acceptable choice for the ability scale to be used. The consistent estimation of ability is precisely defined in Definition 3.2 and a slight variant in Definition 3.6. Roughly, a test $\{U_i, i \geq 1\}$ consistently estimates ability if all reasonable-to-use linear formula scores asymptotically estimate different monotone transformations of the same unidimensional latent ability. "Reasonable-to-use" is formalized by examining collections of nonsparse subtests and balanced linear empirical scalings. In order to facilitate this development, the concepts of marginal item response function and intrinsic ability scale are presented. The estimation of ability in the ordinal scaling sense does not require the IRFs to be known (i.e., calibrated).

Theorem 3.3 shows that essential unidimensionality guarantees, under the mild regularity condition of local asymptotic discrimination of $\{U_i, i \geq 1\}$, that the latent ability is unique up to a monotone transformation, that is, in the ordinal scaling sense.

Section 3.2 extends the theory from collections of nonsparse subtests to balanced linear empirical scalings, thus yielding results for a wide and natural class of empirical scalings.

Section 3.3 addresses the estimation of ability θ <u>on the specified</u> θ <u>scale</u> when IRFs are assumed known. Theorem 3.6 presents a large class of estimators of θ that consistently estimate θ on the θ scale in the sense that each such estimator $\tilde{A}_N^{-1}(\tilde{U}_N)$ satisfies for each θ, given $\Theta = \theta$

$$\tilde{A}_N^{-1}(\tilde{U}_N) \quad \to \theta$$

in probability as $N \to \infty$. THis includes in particular the estimator $A_N^{-1}(\bar{U}_N)$. Each such estimator is computable, has a simple formula, and is based on an admissible linear formula scoring scheme in an intuitively natural way.

Section 4 presents a *conceptual* model *for* the construction of essentially unidimensional tests in the presence of the unavoidable empirical reality of multidimensional items. A test developers' prescription for essentially unidimensional test construction emerges: keep the number of abilities per item small; keep the number of items dependent on the same ability (other than the to-be-measured θ) small; and keep the number of item pairs assigned to the same ability other than θ small. It is stressed in Section 4 and throughout the paper that essential unidimensionality, while defined for the unobservable $\{U_i, i \geq 1\}$, is statistically testable based on data from $\underline{U}_N$ and that the statistical test given in Stout (1987) is precisely designed to assess the degree to which $\underline{U}_N$ is well modeled by the assumption of essential unidimensionality.

<u>Acknowledgements</u>. I wish to thank Brian Junker for useful comments and discussions resulting from his careful and thoughtful reading of the manuscript. I wish to thank Charles Davis for useful comments that lead to several improvements, in particular in

Section 3 and 4. I wish to thank D. R. Divgi for an insight that lead to a useful reformulation of essential independence. Lloyd Humphreys' steadfast insistence that tests items are inherently "multiply determined" provided the empirical grounding for essential dimensionality. Remarks by Mark Reckase and Ming-mei Wang were also helpful.

References

Cliff, N. (1977). A theory of consistency of ordering generalizable to tailored testing. Psychometrika, 42, 375-399.

Cliff, N. (1979). Test theory without true scores? Psychometrika, 44, 373-393.

Hambleton, R.K. and Swaminathan, H. (1985). Item Response Theory: Principles and Applications, Kluwer Nijhoff, Boston.

Holland, P., Junker, B., and Thayer, D. (1987). Recovering the ability distribution from test scores. ETS preprint.

Humphreys, L. (1984). A theoretical and empirical study of the psychometric assessment of psychological test dimensionality and bias, ONR Research Proposal.

Junker, B. (1988). Statistical Aspects of a New Latent Trait Model, Ph.D. Thesis, Department of Statistics, University of Illinois at Urbana-Champaign.

Lord, F.M. (1980). Applications of Item Response Theory to Practical Testing Problems, Lawrence Erlbaum, Hillsdale, New Jersey.

McDonald, R.P. (1981). The dimensionality of tests and items. British Journal of Mathematical and Statistical Psychology, 34, 100-117.

Mislevy, R.J. (1987). Recent developments in IRT. In E. Z. Rothkoph, Ed., Preprint. Educational Testing Service. Review of Research in Education (Vol. 14, 239-275). Washington D.C.: American Educational Research Association.

Mokken, R. and Lewis, C. (1982). A nonparametric approach to dichotomous item responses. Applied Psychological Measurement, 6, 417-430.

Reckase, M. D., Carlson, J. E., Ackerman, T. A. and Spray, J. A. (1986). The interpretation of unidimensional IRT parameters when estimated from multidimensional data. Paper presented at the Psychometric Society Annual Meeting, Tronoto.

Sijtsman, K. and Molenaar, I. (1987). Reliability of test scores in nonparametric item response theory. Psychometrika, 52, 79-98.

Stout, William (1987). A nonparametric approach for assessing latent trait dimensionality. Psychometrika, 52, 589-618.

Tucker, L.R., Koopman, R.F., and Linn, R.L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. Psychometrika, 34, 421-459.

Wang, M. (1988). Measurement bias in the application of a unidimensional model to multidimensional item-response data. Manuscript, University of Iowa.

Dr. Terry Ackerman
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

Dr. Robert Ahlers
Code N711
Human Factors Laboratory
Naval Training Systems Center
Orlando, FL 32813

Dr. James Algina
1403 Norman Hall
University of Florida
Gainesville, FL 32605

Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Isaac Bejar
Mail Stop: 10-R
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Ramat Aviv 69978
ISRAEL

Dr. Arthur S. Blaiwes
Code N712
Naval Training Systems Center
Orlando, FL 32813-7100

Dr. Bruce Bloxom
Defense Manpower Data Center
550 Camino El Estero,
Suite 200
Monterey, CA 93943-3231

Dr. R. Darrell Bock
University of Chicago
NORC
6030 South Ellis
Chicago, IL 60637

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid
Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code 7B
Naval Training Systems Center
Orlando, FL 32813-7100

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

Dr. James Carlson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52243

Dr. John B. Carroll
409 Elliott Rd., North
Chapel Hill, NC 27514

Dr. Robert M. Carroll
Chief of Naval Operations
OP-01B2
Washington, DC 20350

Dr. Raymond E. Christal
UES LAMP Science Advisor
AFHRL/MOEL
Brooks AFB, TX 78235

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

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Hans F. Crombag
Faculty of Law
University of Limburg
P.O. Box 616
Maastricht
The NETHERLANDS 6200 MD

Dr. Timothy Davey
Educational Testing Service
Princeton, NJ 08541

Dr. C. M. Dayton
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Ralph J. DeAyala
Measurement, Statistics,
and Evaluation
Benjamin Bldg., Rm. 4112
University of Maryland
College Park, MD 20742

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

Dr. Hei-Ki Dong
Bell Communications Research
6 Corporate Place
PYA-1K226
Piscataway, NJ 08854

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
Attn: TC
(12 Copies)

Dr. Stephen Dunbar
224B Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. James A. Earles
Air Force Human Resources Lab
Brooks AFB, TX 78235

Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. John M. Eddins
University of Illinois
252 Engineering Research
Laboratory
103 South Mathews Street
Urbana, IL 61801

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

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

Dr. Benjamin A. Fairbank
Performance Metrics, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

Dr. P-A. Federico
Code 51
NPRDC
San Diego, CA 92152-6800

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 5/3
A 1010 Vienna
AUSTRIA

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

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

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

Dr. Alfred R. Fregly
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Robert D. Gibbons
Illinois State Psychiatric Inst.
Rm 529W
1601 W. Taylor 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. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

DORNIER GMBH
P.O. Box 1420
D-7990 Friedrichshafen 1
WEST GERMANY

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. Grant Henning
Senior Research Scientist
Division of Measurement
Research and Services
Educational Testing Service
Princeton, NJ 08541

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 63
San Diego, CA 92152-6800

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

Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 92010

Mr. Dick Hoshaw
OP-135
Arlington Annex
Room 2834
Washington, DC 20350

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
3-104 Educ. N.
University of Alberta
Edmonton, Alberta
CANADA T6G 2G5

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

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

Dr. Douglas H. Jones
Thatcher Jones Associates
P.O. Box 6640
10 Trafalgar Court
Lawrenceville, NJ 08648

Dr. Milton S. Katz
European Science Coordination
Office
U.S. Army Research Institute
Box 65
FPO New York 09510-1500

Prof. John A. Keats
Department of Psychology
University of Newcastle
N.S.W. 2308
AUSTRALIA

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. Leonard Kroeker
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Dr. Jerry Lehnus
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn, VA 22209

Dr. Thomas Leonard
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, WI 53705

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

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

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

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

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

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

Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

Dr. James R. McBride
The Psychological Corporation
1250 Sixth Avenue
San Diego, CA 92101

Dr. Clarence C. McCormick
HQ, USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago, IL 60064

Dr. Robert McKinley
Educational Testing Service
16-T
Princeton, NJ 08541

Dr. James McMichael
Technical Director
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Barbara Means
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Dr. William Montague
NPRDC Code 13
San Diego, CA 92152-6800

Ms. Kathleen Moreno
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Headquarters Marine Corps
Code MPI-20
Washington, DC 20380

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

Deputy Technical Director
NPRDC Code 01A
San Diego, CA 92152-6800

Director, Training Laboratory,
NPRDC (Code 05)
San Diego, CA 92152-6800

Director, Manpower and Personnel
Laboratory,
NPRDC (Code 06)
San Diego, CA 92152-6800

Director, Human Factors
& Organizational Systems Lab,
NPRDC (Code 07)
San Diego, CA 92152-6800

Library, NPRDC
Code P201L
San Diego, CA 92152-6800

Commanding Officer,
Naval Research Laboratory
Code 2627
Washington, DC 20390

Dr. Harold F. O'Neil, Jr.
School of Education - WPH 801
Department of Educational
Psychology & Technology
University of Southern California
Los Angeles, CA 90089-0031

Dr. James B. Olsen
WICAT Systems
1875 South State Street
Orem, UT 84058

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

Office of Naval Research,
Code 125
800 N. Quincy Street
Arlington, VA 22217-5000

Assistant for MPT Research,
Development and Studies
OP 01B7
Washington, DC 20370

Dr. Judith Orasanu
Basic Research Office
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Dr. Randolph Park
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

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

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

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

Department of Operations Research,
Naval Postgraduate School
Monterey, CA 93940

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

Dr. Malcolm Ree
AFHRL/MOA
Brooks AFB, TX 78235

Dr. Barry Riegelhaupt
HumRRO
1100 South Washington Street
Alexandria, VA 22314

Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg.
Knoxville, TN 37916-0900

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152-6800

Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Dr. Mary Schratz
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Dan Segall
Navy Personnel R&D Center
San Diego, CA 92152

Dr. W. Steve Sellman
OASD(MRA&L)
2B269 The Pentagon
Washington, DC 20301

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

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

Dr. H. Wallace Sinaiko
Manpower Research
and Advisory Services
Smithsonian Institution
801 North Pitt Street, Suite 120
Alexandria, VA 22314-1713

Dr. Richard E. Snow
School of Education
Stanford University
Stanford, CA 94305

Dr. Richard C. Sorensen
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Paul Speckman
University of Missouri
Department of Statistics
Columbia, MO 65201

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. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Brad Sympson
Navy Personnel R&D Center
Code-62
San Diego, CA 92152-6800

Dr. John Tangney
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Kikumi Tatsuoka
CERL
252 Engineering Research
Laboratory
103 S. Mathews Avenue
Urbana, IL 61801

Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

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

Dr. Robert Tsutakawa
University of Missouri
Department of Statistics
222 Math. Sciences Bldg.
Columbia, MO 65211

Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E. Street, NW
Washington, DC 20415

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 440
St. Paul, MN 55114

Dr. Frank L. Vicino
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Howard Wainer
Educational Testing Service
Princeton, NJ 08541

Dr. Ming-Mei Wang
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Thomas A. Warm
FAA Academy AAC934D
P.O. Box 25082
Oklahoma City, OK 73125

Dr. Brian Waters
HumRRO
12908 Argyle Circle
Alexandria, VA 22314

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

Dr. Ronald A. Weitzman
Box 146
Carmel, CA 93921

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Douglas Wetzel
Code 51
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90089-1061

German Military Representative
ATTN: Wolfgang Wildgrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington, DC 20016

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

Dr. Hilda Wing
NRC MH-176
2101 Constitution Ave.
Washington, DC 20418

Dr. Martin F. Wiskoff
Defense Manpower Data Center
550 Camino El Estero
Suite 200
Monterey, CA 93943-3231

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

Dr. George Wong
Biostatistics Laboratory
Memorial Sloan-Kettering
Cancer Center
1275 York Avenue
New York, NY 10021

Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
Code 51
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
03-T
Educational Testing Service
Rosedale Road
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

Mr. Anthony R. Zara
National Council of State
Boards of Nursing, Inc.
625 North Michigan Avenue
Suite 1544
Chicago, IL 60611

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