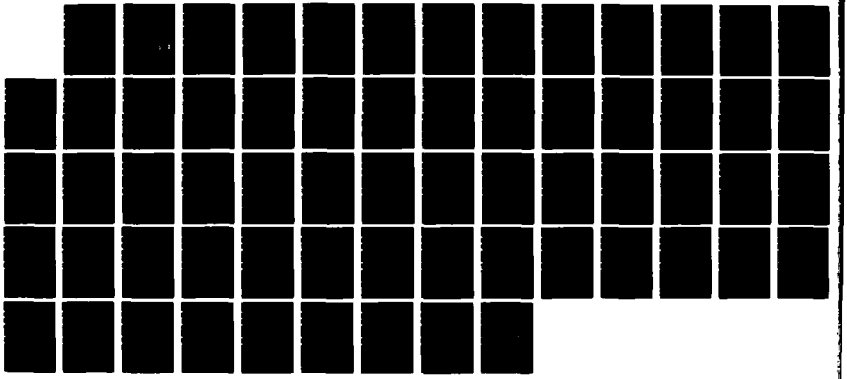
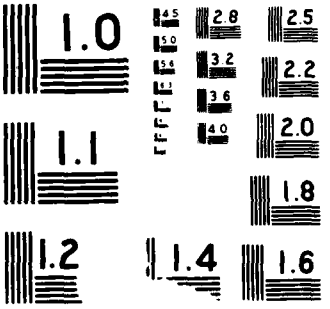
AD-A194 212 A NONPARAMETRIC MULTIDIMENSIONAL IRT APPROACH WITH APPLICATIONS TO ABILIT (U) ILLINOIS UNIV AT URBANA PSYCHOMETRIC GROUP W STOUT APR 88 88-1-ONR 1/1

UNCLASSIFIED N00014-87-K-0277 F/G 12/3 NL




1.0
1.1
1.2
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0

# A Nonparametric Multidimensional IRT Approach with Applications to Ability Estimation and Test Bias

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

April 1988



This research was sponsored by the Cognitive Science Program, Psychological Sciences Division, Office of Naval Research, under Contract N00014-87-K-0277, 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 | 1b RESTRICTIVE MARKINGS A193 214 |
| 2a SECURITY CLASSIFICATION AUTHORITY | 3 DISTRIBUTION / AVAILABILITY OF REPORT |
| 2b DECLASSIFICATION / DOWNGRADING SCHEDULE | Approved for public release; distribution unlimited |
| 4 PERFORMING ORGANIZATION REPORT NUMBER(S) | 5 MONITORING ORGANIZATION REPORT NUMBER(S) |
| ONR 88-1 | |
| 6a NAME OF PERFORMING ORGANIZATION: Psychometric Group, Department of Statistics - Univ. of Ill. / 6b OFFICE SYMBOL (If applicable) | 7a NAME OF MONITORING ORGANIZATION |
| 6c ADDRESS (City, State, and ZIP Code): Department of Statistics, 101 Illini Hall, 725 South Wright Street, Champaign, IL 61820 | 7b ADDRESS (City, State, and ZIP Code): Office of Naval Research, 800 North Quincy Street, Arlington, VA 22217-0277 |
| 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 | RR04204 | RR0420401 | 4421-548 |

11 TITLE (Include Security Classification)

A Nonparametric Multidimensional IRT Approach with Applications to Ability Estimation and Test Bias

12 PERSONAL AUTHOR(S)

Stout, William

| 13a TYPE OF REPORT | 13b TIME COVERED | 14 DATE OF REPORT (Year, Month, Day) | 15 PAGE COUNT |
|---|---|---|---|
| Technical Report | FROM _____ TO _____ | April 1988 | 52 |

16 SUPPLEMENTARY NOTATION

Condensed version accepted for publication

| 17 COSATI CODES | | | 18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number) |
|---|---|---|---|
| FIELD | GROUP | SUB-GROUP | Local independence, essential independence, essential trait, intrinsic ability scale, marginal item response function, latent dimensionality, multidimensionality. |
| 05 | 10 | | |
| | | | |

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

A determined case is made for the use of a nonparametric multidimensional monotonic IRT modeling framework with local independence replaced by the less restrictive assumption of essential independence. The concept of essential dimensionality is then introduced to count the number of dominant latent dimensions. Consequences of this more general approach include the consistent estimation of ability on a common scale using a natural class of estimators, uniqueness of the latent ability when essential unidimensionality holds, a theoretical treatment of test bias, an IRT based notion of validity, and a reassessment of the importance of the concept of item parameter invariance.

| 20 DISTRIBUTION / AVAILABILITY OF ABSTRACT | 21 ABSTRACT SECURITY CLASSIFICATION |
|---|---|
| ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS | unclassified |

| 22a NAME OF RESPONSIBLE INDIVIDUAL | 22b TELEPHONE (Include Area Code) | 22c OFFICE SYMBOL |
|---|---|---|
| Dr. James Lester | (202) 696-4503 | ONR 1142CS |

DD Form 1473, JUN 86 *Previous editions are obsolete* SECURITY CLASSIFICATION OF THIS PAGE

18 Subject terms continued:

essential dimensionality, essential unidimensionality, item response theory, latent trait theory, ability estimation, consistent estimation, item paramenter invariance, validity, linear formula scoring, nonparametric.

ABSTRACT

A determined case is made for the use of a nonparametric multidimensional monotonic IRT modeling framework with local independence replaced by the less restrictive assumption of essential independence. The concept of essential dimensionality is then introduced to count the number of dominant latent dimensions. Consequences of this more general approach include the consistent estimation of ability on a common scale using a natural class of estimators, uniqueness of the latent ability when essential unidimensionality holds, a theoretical treatment of test bias, an IRT based notion of validity, and a reassessment of the importance of the concept of item parameter invariance.

Key words: Local independence, essential independence, essential trait, intrinsic ability scale, marginal item response function, latent dimensionality, multidimensionality, essential dimensionality, essential unidimensionality, item response theory, latent trait theory, ability estimation, consistent estimation, item parameter invariance, validity, linear formula scoring, nonparametric.

# A Nonparameteric Multidimensional IRT Approach with Applications to Ability Estimation and Test Bias

Introduction. The central thesis of this paper is that a successful approach to such fundamental topics as bias, consistent estimation of the ability intended to be measured, and item calibration requires a nonparametric multidimensional item response theory (IRT) modeling approach with an infinite item pool assumed. Until recently, most theoretical and applied 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 fixed finite number of items. (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 the use of a non-parametric multidimensional monotonic IRT modeling framework with local independence replaced by a less restrictive and, we claim, psychometrically more appropriate assumption, namely essential independence. In the spirit of factor analysis, essential independence together with essential dimensionality provide a conceptual basis for establishing the number of major latent dimensions even in the presence of multiple minor dimensions. Essential unidimensionality, the existence of exactly one major dimension, provides a conceptual basis for carrying out IRT based statistical analyses that require unidimensionality. It is our position that a standard unidimensional IRT modeling approach should only be used subsequent to a careful multivariate statistical analysis of unidimensionality based on a more general nonparametric multidimensional approach like the one herein. To use uncritically 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.

Consequences of our more general multidimensional modeling approach include the establishment of consistent estimation of ability on a common ability scale even when different examinees have taken different tests, and the existence of a "unique" (appropriately defined) latent ability provided essential unidimensionality holds. As a vital part of our proposed multidimensional IRT framework, the concept of the intrinsic ability scale of a test is also presented. Further, our approach leads to a re-examination of test bias from a multidimensional perspective.

This paper continues the work of Stout (1987), where essential unidimensionality was first defined and a statistical test of essential undimensionality was presented and explored.

The paper is organized as follows: Section 1 reviews the traditional multidimensional IRT model. Section 2 defines essential dimensionality and studies some of its basic properties. Section 3 considers the consistent estimation of ability in the single-test single-population setting. The uniqueness of the latent ability is considered. Section 4 cautions against the overreliance on item parameter invariance and presents its relationship to essential unidimensionality. An IRT based definition of validity is proposed. Section 5 proposes a new definition of test bias and studies test bias from a multidimensional modeling prospective. Section 6 considers the consistent estimation of ability using any of a large class of linear formula scores including proportion correct. Section 7 considers the consistent estimation of ability in multiple-test multiple-population settings. Section 8 briefly discusses and summarizes the results of the paper.

1. Multidimensional Modeling. According to the latent trait viewpoint, each examinee is indexed by a possibly vector valued and not necessarily distinct variable $\underline{\theta}$. Associated with each item $i$ is an item response function (IRF) $P_i(\underline{\theta})$ that denotes the probability that a randomly chosen examinee from those examinees with ability $\underline{\theta}$ will get the item right. Random sampling of examinees from a specified population induces a distribution $F$ on $\underline{\theta}$, and hence, a distribution on the test responses,

$$\underline{U}_N \equiv (U_1, \cdots, U_N).$$

Here $U_i = 1$ denotes a correct response and $U_i = 0$ an incorrect response to item $i$ by a randomly chosen examinee. Note that $P_i(\underline{\theta}) = P[U_i = 1 | \underline{\Theta} = \underline{\theta}] = E[U_i | \underline{\Theta} = \underline{\theta}]$ for all $i, \underline{\theta}$. It is important to stress that a model $\underline{U}_N$ can have many latent model representations $(\underline{U}_N, \underline{\Theta})$. That is, there are many choices of $\underline{\Theta}$ such that, for all $\underline{u}_N$,

$$P[\underline{U}_N = \underline{u}_N] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P[\underline{U}_N = \underline{u}_N | \underline{\Theta} = \underline{\theta}] dF(\underline{\theta}) \tag{1.1}$$

Three characteristics of latent representations are of considerable importance:

(i) The model $(\underline{U}_N, \underline{\Theta})$ is said to be a monotone model if $P_i(\underline{\theta})$ is nondecreasing in $\underline{\theta}$ for each $i$ (here $\underline{\theta}_1 \leq \underline{\theta}_2$ if and only if $\theta_{1i} \leq \theta_{2i}$ for each coordinate $i$). $M$ will denote such a monotone model.

(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 a locally independent (LI) model if

$$P[U_1 = u_1, \cdots, U_N = u_N | \underline{\Theta} = \underline{\theta}] = \prod_{i=1}^{N} P[U_i = u_i | \underline{\Theta} = \underline{\theta}] \tag{1.2}$$

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. Usually for models when M, d = 1 holds, the IRFs will be strictly monotone.

2. A New Conceptualization of Test Dimensionality. For later use, we first establish that an LI, M latent model with latent dimensionality ≤ N is always possible for a given length N test $\underline{U}_N$.

Theorem 2.1. For any test $\underline{U}_N$ there exists an LI, M latent model representation $(\underline{U}_N, \underline{\theta})$ with dim $(\underline{\theta}) \leq N$.

Proof of Theorem 2.1. Assume that the test $\underline{U}_N$ has a particular distribution. Define

$$\theta_i = U_i \tag{2.1}$$

for $i \leq N$. Then the range of each $\theta_i$ is $\{0,1\}$ and for all $i$,

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

The intuitive interpretation of (2.1) is that $\underline{\theta}$ represents an examinee's state of knowledge of the N items, in the sense that $\theta_i = 1$ or 0 depending on whether the examinee knows the answer to item $i$ or not. Clearly, monotonicity holds because $P_i(\underline{\theta})$ is monotone in $\underline{\theta}$ for each $i$. In order to verify LI, note that for all $\underline{u}$, $\underline{\theta}$ that

$$P[\underline{U}_N = \underline{u}|\underline{\theta} = \underline{\theta}] = P[\underline{U}_N = \underline{u}|\underline{U}_N = \underline{\theta}] = \begin{cases} 1 & \text{if } \underline{u} = \underline{\theta} \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

and that (letting $0^0 = 1$ for convenience)

$$\prod_{i=1}^{N} P_i(\underline{\theta})^{u_i} [1 - P_i(\underline{\theta})]^{1-u_i} = \prod_{i=1}^{N} \theta_i^{u_i}(1 - \theta_i)^{1-u_i} = \begin{cases} 1 & \text{if } \underline{u} = \underline{\theta} \\ 0 & \text{otherwise} \end{cases}.$$

Thus LI holds. Finally, it must be verified that $(\underline{U}_N, \underline{\theta})$ is a latent representation in the sense that (1.1) holds: Note that (2.1) implies that $P[\underline{\theta} = \underline{U}_N] = P[\underline{U}_N = \underline{u}_N]$. Thus, applying (2.2) to the right hand side of (1.1) with $\underline{u} = \underline{u}_N$ shows that (1.1) holds. □

This mathematically trivial theorem provides a certain insight: For a test of length N, the assumption of an LI,M model is a totally nonrestrictive assumption provided $\underline{\theta}$ is allowed to have an N dimensional distribution. Thus, given a sequence of tests $\{\underline{U}_N, N \geq 1\}$, each $\underline{U}_N$ can be given an LI,M representation $(\underline{U}_N, \underline{\Theta}_N)$ such that $\dim(\underline{\Theta}_N) \leq N$. This fact will be used below.

It is important to note that, although mathematically helpful, the latent trait $\underline{\theta}$ of Theorem 2.1 is totally uninteresting from the modeling viewpoint. For, unlike the $\underline{\theta}$ of Theorem 2.1, a valid latent trait should surely be something more abstract and indeed more "latent" than an examinee's state of knowledge concerning a particular finite set of items. Surely no meaningful cognitive construct (e.g., reading comprehension) is reducible to which of a finite set of items an examinee can correctly answer. Further, it is clearly inappropriate from a modeling viewpoint for the dimensionality of the test to equal the test length.

The $\underline{\theta}$ of Theorem 2.1 is uninteresting from another viewpoint as well. Holland and Rosenbaum (1987) strongly make the point that an assumption true for all models is "vacuous" and is neither a mathematical assumption (because it is always satisfied) nor a scientific hypothesis (because it places no testable restrictions on the behavior of observable data). The $\underline{\theta}$ guaranteed to exist by Theorem 2.1 is clearly of this vacuous character. What is interesting is that for many tests $\underline{U}_N$ there <u>do</u> exist lower dimensional (than N) latent trait representations. Indeed, the psychometrician's goal is to construct a test that validly measures the construct of interest using items sufficiently "homogeneous" that the test can be well modeled by a low dimensional or hopefully even unidimensional model.

Let us recall the <u>traditional IRT definition of test dimensionality</u>:

<u>Definition 2.1</u> The dimensionality d of a test $\underline{U}_N$ is the minimal

dimensionality required for $\underline{\theta}$ to make the latent model representation ($\underline{U}_N$, $\underline{\theta}$) an LI.M representation.

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 over-simplified fashion. It should be obvious that unidimensionality can only be approximated. Even in highly homogeneous tests the mean correlation between paired items is quite small. 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. Test theory must cope with these small correlations. We start with the assumption that responses to items have many causes or determinants.

Humphreys (1984) asserts that dominant attributes (dimensions) result from overlapping attributes common to many items. Attributes unique to individual items or common to relatively few items are unavoidable and indeed are not detrimental to the measurement of dominant dimensions. In his writings, Humphreys stresses that the low item intercorrelations researchers have observed argue strongly for viewing items as multiply determined. Although the existence of multiply determined items is rarely stressed 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 implicitly 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 factor and unique factors. Tucker has specialized this model to binary item tests in work yet to be published.

Unfortunately, the traditional definition (Definition 2.1), with its insistence on the achievement of local independence, makes no distinction between dominant and minor factors. 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, from both a psychometric and a data analytic viewpoint, 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 $d = d_0 > 1$ ($d_0$, thus assigned, possibly quite large in fact) in settings where it would be desirable to assign $d = 1$. The following hypothetical example illustrates the multidimensional nature of items in tests that should be considered unidimensional.

<u>Example</u>. Consider a multiple item "probability" test in which item 1 measures ability in probability but is influenced by the examinees' knowledge of an ordinary deck of playing cards, item 2 measures ability in probability but is influenced by the examinees' understanding of elementary physics, item 3 measures ability in probability but is influenced by the examinees' knowledge of elementary Mendelian genetics, item 4 measures ability in probability, but... .

One is clearly forced to label such a test as multidimensional according to the traditional conceptualization of dimensionality described above. Indeed, it

is clear that $d \geq 3$: with the dimensions including ability in probability the context of bridge knowledge, ability in probability in the context of elementary physics knowledge, and ability in probability in the context of knowledge of genetics.

The multicontextual nature of this example is deliberate. It seems undesirable to construct, perhaps under the guise of eliminating biased ite context free probability test (even if possible), for it would probably not measure what should be measured, namely the ability to solve probability problems in a variety of contexts. Hence, whether the multiple determinant prominent as in the above example or more subtle, a probability exam would necessity comprise multiply determined items. Moreover, testing that compr multiply determined items is necessarily widespread and is in no way restri only to tests in probability. Clearly, it would be useful to have a concep test dimensionality that would allow such tests as the above to be consider unidimensional. Such a conceptualization is provided by the essential dimensionality of a set of items, defined below. This definition is design count the number of dominant dimensions only, uninflated by the incidental multidimensionality of items.

In order to present our definition of essential unidimensionality and study the asymptotic theory of ability estimation, it is necessary to view test $U_N$ as embedded in a sequence of tests $U_1, \cdots, U_N, \cdots$, each obtainec the previous one by the addition of one more item. Two justifications for realism of this shift in modeling perspective can be given: (1) If an acti Item Banking scheme is being used to construct the test, then our embeddin scheme is totally realistic. Indeed, random sampling of items is commonly for criterion referenced tests constructed from item banks, according to

Hambleton and Swamanathan (1985, Chapter 12). (2) Even when there is no actual sampling of items from a population of items, $\underline{U}_N$ certainly can be and we think should be viewed as a representative sample of the infinite population that would be constructed by continuing to generate items by whatever test construction process has been used to generate the first $N$ items $\underline{U}_N$.

It will be assumed throughout the remainder of the paper that $\underline{U}_N$ is embedded in a sequence $\underline{U}_1, \cdots, \underline{U}_N, \cdots$. This will be referred to as the item pool formulation of IRT. This embedding is analogous to the mathematical statistician's study of the estimation of a population mean, say, by a sequence of estimators $\bar{X}_N = \Sigma_1^N X_i/N$ resulting from a sequence of random samples, each of which is obtained from the preceeding one by the selection of one more observation. As with Justification (2) above, such a sampling model is often used when the "population" being "sampled" from is only conceptual rather than an actual population.

We now define a weaker type of independence than local independence.

Definition 2.2. The latent model $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$ is said to be essentially independent (EI) if the conditional distribution of $\underline{U}_N$ given $\underline{\Theta}$ in (1.1) satisfies for each $\underline{\theta}$ in the range of $\underline{\Theta}$,

$$(2.3) \qquad D_N(\underline{\theta}) \equiv \frac{\sum_{1 \leq i < j \leq N} |\mathrm{Cov}(U_i, U_j \mid \underline{\Theta} = \underline{\theta})|}{\binom{N}{2}} \to 0 \text{ as } N \to \infty.$$

Remarks. Mathematically, the notation $\{\underline{U}_N, \underline{\Theta}, N \geq 1\}$ in Definition 2.2 means that the $\underline{U}_N$ are random vectors and $\underline{\Theta}$ a random vector defined on a common probability space. This corresponds to the intuitive notion of an infinite item pool and a fixed examinee population from which the sample is drawn. In this paper, issues of rigor when using measure-theoretic probability, although always surmountable, are suppressed in the interest of clarity.

It is informative to contrast the definition of essential independence with the traditional latent trait conceptualization of local independence given in (1.2). LI implies pairwise independence of all pairs $(U_i, U_j)$, $i \neq j$, given $\underline{\theta}$, which is equivalent to $\text{cov}(U_i, U_j | \underline{\theta} = \underline{\theta}) = 0$ for all $\underline{\theta}$, $i \neq j$. By contrast, EI only requires that for each fixed $\underline{\theta}$, $\text{cov}(U_i, U_j | \underline{\theta} = \underline{\theta})$ is small on average as the test length $N$ grows.

Now essential dimensionality can be defined.

Definition 2.3. The essential dimensionality $d_E$ of a family of tests $\{\underline{U}_N\}$ is the minimal dimensionality required for a latent trait $\underline{\theta}$ to make the latent model representation $\{\underline{U}_N, \underline{\theta}, N \geq 1\}$ an EI, M representation. When $d_E = 1$, essential unidimensionality is said to hold. If essential $d_E$ dimensionality holds using ability $\underline{\theta}$, then $\{\underline{U}_N\}$ is said to be essentially $d_E$ dimensional with respect to ability $\underline{\theta}$. An essential trait $\underline{\theta}$ is any latent trait $\underline{\theta}$ for which $\{\underline{U}_N, \underline{\theta}, N \geq 1\}$ is an EI, M representation with the essential dimensionality of $\underline{\theta}$ the minimum possible.

Remarks. Although $d_E = 0$ is theoretically possible, it is psychometrically uninteresting. Thus, to avoid irrelevant trivialities it is assumed that $d_E \geq 1$ for all latent representations considered in this paper.

The following theorem makes precise one way that essential unidimensionality might occur.

Theorem 2.2. Suppose there is a random variable $\theta$ such that for each $\theta$

$$\sup_{|i-j| \geq N} \left| \text{cov}(U_i, U_j | \theta = \theta) \right| \to 0 \quad \text{as} \quad N \to \infty.$$

Then essential unidimensionality holds.

Proof. Fix $\theta$. Fix $\epsilon > 0$. Choose $N_0$ such that

$$\sup_{|i-j| \geq N_0} \left| \text{cov}(U_i, U_j | \theta = \theta) \right| \leq \epsilon.$$

Then, for $N > N_0$,

$$\Sigma_{1\le i<j\le N}\left|\mathrm{cov}(U_i, U_j|\theta = \theta)\right| \le \epsilon N^2 + N N_0.$$

Then $D_N(\theta) \le 3\epsilon$ for N large, thus establishing essential unidimensionality. □

The following example illustrates the difference between essential dimensionality (Definition 2.3) and the traditional definition (Definition 2.1) of dimensionality.

Example 2.1. Consider the construction of a paragragh 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 $\theta$, where for convenience we think of $\theta$ as reading ability. Suppose that $\mathrm{cov}(U_i, U_j|\theta = \theta) > 0$ for all $U_i, U_j$ for the same paragraph (as should be the case) and that the IRFs are monotone. Note, using $|\mathrm{cov}(U_i, U_j|\theta = \theta)| \le 1$ for all $i, j, \theta$,

$$D_N(\theta) \le \frac{n\binom{5}{2}}{\binom{5n}{2}} \le \frac{\text{const}}{N-1} \longrightarrow 0 \quad \text{as} \quad N \longrightarrow \infty.$$

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

The example 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.

Example 2.1 suggests an interesting sufficient condition for essential unidimensionality: If there is one trait common to many items, if the other traits are "orthogonal" to one another given this "dominant" trait, and if each of the other traits influences only a bounded finite number of items, then

essential unidimensionality should hold.

Theorem 2.3. Let $\underline{U} = \{U_i,\ i \geq 1\}$ be an item pool with $\underline{U}$ partitioned as $\underline{U} = (\underline{U}^{(1)}, \underline{U}^{(2)}, \cdots)$. Suppose that the number of items in each $\underline{U}^{(i)}$ is uniformly bounded in $i$ and that local independence holds with respect to $\underline{\theta} = (\theta, \theta_1, \theta_2, \cdots)$. Suppose that the item response functions for each item of $\underline{U}^{(i)}$ depend on $\theta, \theta_i$ only, and that $\theta_i$ and $\theta_{i'}$ are conditionally independent given $\theta$ for all $i \neq i'$. Suppose for all $i$ that

$$P_j(\theta) \stackrel{\text{def}}{=} \int P_{ji}(\theta, \theta_i)\, dP(\theta_i|\theta) \tag{2.4}$$

is monotone nondecreasing in $\theta$, where $P_{ji}(\theta, \theta_i)$ is the item response function of the jth item of $\underline{U}_i$ and $P(\theta_i|\theta)$ is the conditional distribution of $\theta_i$ given $\theta = \theta$. Then $\{\underline{U}_N,\ N \geq 1\}$ is essentially unidimensional with respect to ability $\theta$.

Remarks. In the statement of Theorem 2.3 above, $\theta$ is the essential trait. The hypotheses of Theorem 2 3 can easily be modified to allow each individual item to depend on more than one nuisance parameter, or to allow the number of items in each category to grow slowly.

Proof of Theorem 2.3. Choose $U_j$, $U_{j'}$ from different partitions - say, $i$ and $i'$. Then, denoting the joint density of $\theta_i$ and $\theta_{i'}$, given $\theta = \theta$ by $f(\theta_i, \theta_{i'}|\theta)$,

$$\begin{aligned} E(U_j U_{j'}|\theta = \theta) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_{ji}(\theta, \theta_i)\, P_{j'i'}(\theta, \theta_{i'})\, f(\theta_i, \theta_{i'}|\theta)\, d\theta_i d\theta_{i'} \\ &= \int_{-\infty}^{\infty} P_{j'i'}(\theta, \theta_{i'}) \left[\int_{-\infty}^{\infty} P_{ji}(\theta, \theta_i)\ f(\theta_i|\theta)\, d\theta_i\right] f(\theta_{i'}|\theta)\, d\theta_{i'} \\ &= E[U_j|\theta = \theta]\, E[U_{j'}|\theta = \theta]. \end{aligned}$$

Thus, given $\theta$, $U_j$ and $U_{j'}$ are therefore conditionally independent. Thus $\mathrm{cov}(U_j, U_{j'}|\theta = \theta) = 0$. Now let the number of items $K_i$ in each partition set $\underline{U}^{(i)}$ be bounded by $K$. Then, $D_N(\theta)$ satisfies for all $\theta$

$$D_N(\theta) \le \frac{\sum_{i=1}^{n_N} \binom{K_i}{2}}{\binom{N}{2}}$$

where $n_N$ = the number of partitions into which the first $N$ items are split. Thus, noting that $\sum_{i=1}^{n_N} K_i = N$, for all $\theta$

$$D_N(\theta) \le \frac{\sum_{i=1}^{n_N} K_i^2}{N(N-1)} \le \frac{KN}{N(N-1)} \le \frac{K}{N-1} \longrightarrow 0 \quad \text{as} \quad N \longrightarrow \infty,$$

which establishes the result since M holds for the $P_j(\theta)$ of (2.4) by hypothesis.

□

3. Application to Consistent Estimation of Ability. We turn now to the problem of estimating a particular latent ability $\theta$ in the presence of other ("nuisance") abilities. In order to illuminate certain theoretical issues most clearly, ability estimation will be considered in its simplest setting: we consider a fixed test for a single examinee population with no consideration of scaling/equating issues such as the need to find a common ability scale when using more than one test. Later in Sections 4, 6, and 7 we will address some of the practical problems raised when the rather strict one-test one-population assumptions are relaxed.

We suppose that $\{\underline{U}_N, \underline{\theta}, N \ge 1\}$ is either essential $d_E$ dimensional or traditional d dimensional with respect to $\underline{\theta}$ for some $d_E \ge 1$ or $d \ge 1$ respectively. The item response functions for $\{\underline{U}_N, \underline{\theta}, N \ge 1\}$ are denoted by $P_i(\underline{\theta})$. Let $\theta$ be the ability desired to be estimated, and suppose that $\underline{\theta}$ determines $\theta$; i.e., that $\theta$ is a function of $\underline{\theta}$.

In this single population problem there is nothing unique or preferable about the $\theta$ scale. That is, any strictly increasing transformation $A(\theta)$ yields an equally acceptable scale for purposes of estimating $\theta$. Let

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

(the distinction between $P_i(\theta)$ and $P_i(\underline{\theta})$ henceforth assumed clear from context). The $P_i(\theta)$s are called the marginal item response functions with respect to ability $\theta$. Let

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

$A_N(\theta)$ is called the intrinsic ability scale for $\theta$ relative to the test $\underline{U}_N$ and to 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 $\theta$. Under the assumption of strict monotonicity, Theorem 3.1 below implies that $A_N(\theta)$ is strictly increasing in $\theta$ and hence is an acceptable scale for estimating $\theta$.

Considerable recent attention has been focused on nonmonotone unidimensional item response functions. It has been shown that attractive distractors are a source of nonmonotonicity. 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 $P_i(\underline{\theta})$ can be monotone and yet $P_i(\theta)$ nonmonotone:

Example 3.1. Let $P(\theta_1, \theta_2) = (\theta_1 + \theta_2)/17$, $1/4 \le \theta_1 \le 1$, $0 \le \theta_2 \le 16$, and $f(\theta_2|\theta_1) = \theta_1/4$ if $0 \le \theta_2 \le 4/\theta_1$; $= 0$ otherwise.

Then

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

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

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

Of course, as is intuitively clear, mild and natural regularity conditions preclude this nonmonotone behavior. Indeed, the nonmonotonicity of a projected

item response function can occur only when the multidimensional ability $\underline{\theta}$ has some sort of negative association among its components.

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

$$P[\underline{X} \geq \underline{t}] \leq P[\underline{Y} \geq \underline{t}] \tag{3.3}$$

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 non-negative nondecreasing real valued function. Then

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

Theorem 3.1. Let $(\underline{U}_N, \underline{\theta})$ be a monotone representation. Let $\underline{\theta} \equiv (\theta, \underline{\theta}_2)$. Suppose that, for every $\theta_1 < \theta_2$ pair that the distribution of $\underline{\theta}_2$ given $\theta = \theta_2$ is stochastically larger than the distribution of $\underline{\theta}_2$ given $\theta = \theta_1$. Then, for $i \leq N$, each marginal item response function $P_i(\theta)$ of (3.1) is nondecreasing in $\theta$.

Proof. It must be shown that

$$\int P_i(\theta, \underline{\theta}_2) dP(\underline{\theta}_2 | \theta)$$

is nondecreasing in $\theta$ where $P(\underline{\theta}_2 | \theta)$ denotes the distribution of $\underline{\theta}_2$, given $\Theta = \theta$. Fix $\theta' < \theta''$. By Lemma 3.1, noting that $P_i(\theta, \underline{\theta}_2)$ is for fixed $\theta$ a nondecreasing function of $\underline{\theta}_2$ by the assumption of monotonicity,

$$\int P_i(\theta', \underline{\theta}_2) dP(\underline{\theta}_2 | \theta') \leq \int P_i(\theta', \underline{\theta}_2) dP(\underline{\theta}_2 | \theta'') \tag{3.4}$$

But, because $P_i(\underline{\theta}) \equiv P_i(\theta, \underline{\theta}_2)$ is nondecreasing in $\theta$ for each fixed $\underline{\theta}_2$,

$$\int P_i(\theta', \underline{\theta}_2) dP(\underline{\theta}_2 | \theta'') \leq \int P_i(\theta'', \underline{\theta}_2) dP(\underline{\theta}_2 | \theta'') \tag{3.5}$$

The combination of (3.4) and (3.5) yields the desired result. □

We now turn directly to the ability estimation problem. As we shall see, essential unidimensionality characterizes the consistent estimation of some unidimensional latent ability; moreover, it implies that, in a certain sense,

the latent ability is unique. It is in this spirit that an essential trait (recall Definition 2.3) 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. The main estimation tool is proportion correct $\bar{U}_N$. Later in Sections 6 and 7 generalizations based on "linear formula scores" are explored.

Before stating Theorem 3.2, we must carefully consider what it means to consistently estimate $\theta$ using an item pool formation. Recall our viewpoint that any strictly monotone transformation of $\theta$ -- for example $A_N(\theta)$, which is strictly monotone when the IRFs are -- is an acceptable scale on which to estimate $\theta$. Clearly $\bar{U}_N$ is a natural estimator of $A_N(\theta)$.

Let N(C) denote the cardinality of a finite set C. For $\epsilon>0$, let $\mathscr{C}_N(\epsilon)$ denote the collection of all subsets C of $\{1,\ldots,N\}$ such that $N(C)/N \geq \epsilon$. For example, $\mathscr{C}_N(1/2)$ consists of all subsets containing at least half the integers between 1 and N. We will call a sequence $\{C_N,\ N \geq 1\}$ of integer subsets nonsparse provided there exists $\epsilon > 0$ such that $C_N \in \mathscr{C}_N(\epsilon)$ for every N. Let $C \in \mathscr{C}_N(\epsilon)$. Define

$$\bar{U}_C = \sum_{i \in C} U_i/N(C).$$

Definition 3.2. It is said that $\theta$ may be consistently estimated (in probability) using the sequence $\{U_N,\ N \geq 1\}$ of items if for every nonsparse sequence $\{C_N,\ N \geq 1\}$ there exists a sequence $\{g_{C_N}(\theta),\ N \geq 1\}$ of functions of $\theta$ such that for each $\theta$, given $\Theta = \theta$,

$$\bar{U}_{C_N} - g_{C_N}(\theta) \to 0 \tag{3.6}$$

in probability as $N \to \infty$.

Remarks. The intuitive idea of the definition is that any nonsparse subsequence of items should be usable to estimate $\theta$ in the sense of (3.6). Also, not all sparse subsequences of items need necessarily be usable to estimate $\theta$. For example if every $2^k$th item in a "mathematics" test were a "verbal" item, we would still be able to consistently estimate $\theta$ (mathematics) since the bad sequence of estimators $\{\sum_{k=1}^{M} U_{2^k}/M,\ M \geq 1\}$ is formed from too sparse a subsequence of the items $\{U_i,\ i \geq 1\}$. However, Definition 3.2 does for example require $\{\sum_{i=1}^{M} U_{10i}/M,\ M \geq 1\}$ to estimate $\theta$ in the sense of (3.6). Thus, if every 10th item were a verbal item, the sequence of items $\{U_N,\ N \geq 1\}$ would not be able to consistently estimate $\theta$.

A reasonable question to ask is why our definition of consistent estimation should require (3.6) rather than merely requiring the existence of functions $\{g_N(\theta)\}$ such that for each given $\theta$

$$\bar{U}_N - g_N(\theta) \to 0 \tag{3.7}$$

in probability as $N \to \infty$. One reason is that if a test is formed by sampling items, as in item banking or computerized adaptive testing, then clearly $\theta$ must be estimable using any reasonable sequence $C_N$ in (3.6). A second reason that requiring (3.7) is inappropriate is that it is vacuous in the sense that every test $\underline{U}_N$ of fixed length $N$ can be viewed as embedded in an essentially $d_E$ dimensional sequence of tests $\{\underline{U}_N,\ \underline{\theta},\ N \geq 1\}$ such that (3.7) holds for a judicious choice of $\theta$. The following example 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 family of tests where (3.7) is satisfied for a mathematically judicious choice of $\theta$, $\theta$ being some function of the dimensions $(\theta_1, \theta_2)$. However, (3.6), which postulates consistent estimation of $\theta$ by all nonsparse subsequences of items, is seen to fail in the sense that

two nonsparse subsequences of items can be selected that estimate $\theta_1$, $\theta_2$ respectively. 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 (3.6) instead of (3.7) as a definition of consistency reflects these considerations.

Example 3.2. Let $(\underline{U}_{2K}, \underline{\Theta})$ be a LI, M family of latent variable models with $\underline{\theta} = (\theta_1, \theta_2)$ and for $1 \le i \le K$,

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

where the distribution of $\underline{\Theta}$ is given by $\Theta_1, \Theta_2$ independent identically distributed with $\Theta_1$ uniformly distributed on [0,1]. Let $\Theta = \Theta_1 + \Theta_2$. Fix $\theta$. Then, standard multivariable calculus yields for $1 \le i,j \le K$, $1 \le k \le 2K$

$$E(U_k|\Theta = \theta) = \frac{\theta}{2}, \; \operatorname{cov}(U_{2i-1}, U_{2j}|\Theta = \theta) = -\frac{\theta^2}{12} \tag{3.8}$$

and, if $i \neq j$,

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

Thus, using (3.8),

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

Further, $E\left[\bar{U}_{2K}|\Theta = \theta\right] = \theta/2$. Thus $P\left[|\bar{U}_{2K} - \frac{\theta}{2}| > \epsilon \;|\Theta = \theta\right] \le \operatorname{Var}(\bar{U}_{2K})/\epsilon^2 \to 0$ as $K \to \infty$. Hence, for each $\theta$, 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 (3.7) does hold in what is clearly an essentially two dimensional sequence of tests. However, it is intuitively clear that (3.6) fails since the even items can be used to consistently estimate $\theta_1$ and odd items to consistently estimate $\theta_2$. □

Now Theorem 3.2 can be stated and proved.

Theorem 3.2. Let $\{\underline{U}_N\}$ be essentially unidimensional with respect to ability $\Theta$. Then, $\theta$ may be consistently estimated. In particular, for each given $\theta = \theta$, (3.7) holds.

Conversely, if for some monotone latent model $\{\underline{U}_N, \Theta\}$ the undimensional $\theta$ may be consistently estimated, then $\{\underline{U}_N, \Theta\}$ is an EI, M representation and hence essential unidimensionality holds.

Proof. Assume essential unidimensionality. Fix $\epsilon > 0$ and $\theta$.

$$P[|\bar{U}_N - A_N(\theta)| > \epsilon | \Theta = \theta] \le \mathrm{Var}(\bar{U}_N | \Theta = \theta)/\epsilon^2 \tag{3.9}$$

since $E[\bar{U}_N | \Theta = \theta] = A_N(\theta)$.

But, noting that

$$\sum_{i=1}^{N} \mathrm{Var}(U_i | \Theta = \theta) \le N/4, \tag{3.10}$$

it follows that

$$\mathrm{Var}(\bar{U}_N | \Theta = \theta) \le \frac{1}{4N} + 2 \sum_{1 \le i < j \le N} \frac{\mathrm{Cov}(U_i, U_j | \Theta = \theta)}{N^2} \to 0 \quad \text{as} \quad N \to \infty$$

by the assumption of essential unidimensionality with respect to $\theta$. Thus (3.7) follows with $g_N(\theta) = A_N(\theta)$. Now consider any nonsparse sequence $\{C_N, N \ge 1\}$. It follows from essential unidimensionality that

$$\frac{1}{\binom{N(C_N)}{2}} \sum_{i,j \in C_N, i \ne j} |\mathrm{cov}(U_i, U_j | \Theta = \theta)| \to 0$$

since for some $\epsilon > 0$, $N(C_N)/N \ge \epsilon$ for all N. Thus the same argument that

established (3.7) yields (3.6); i.e., consistent estimation of $\theta$. Here $g_{C_N}(\theta) = \sum_{i \in C_N} P_i(\theta) \cdot N(C_N)$.

Conversely, suppose $\theta$ may be consistently estimated. Consider first the case of $C_N = \{1, \ldots, N\}$. Let $\{g_N(\theta),\ N \geq 1\}$ denote the centering functions for $U_N$ guaranteed to exist by Definition 3.2. Suppose without loss of generality that $0 \leq g_N(\theta) \leq 1$ for all $N, \theta$. Note that

$$(3.11) \qquad 0 \leq \mathrm{Var}\left(\sum_{i=1}^{N} U_i \,|\, \Theta = \theta\right) = \sum_{i=1}^{N} \mathrm{Var}(U_i \,|\, \Theta = \theta) + 2 \sum_{1 \leq i < j \leq N} \mathrm{Cov}(U_i, U_j \,|\, \Theta = \theta),$$

which implies by (3.10) that

$$(3.12) \qquad \inf_{\theta} \sum_{1 \leq i < j \leq N} \mathrm{Cov}(U_i, U_j \,|\, \Theta = \theta) \geq -\frac{N}{8}$$

Therefore

$$\frac{1}{\binom{N}{2}} \sum_{1 \leq i \neq j \leq N} \mathrm{Cov}(U_i, U_j \,|\, \Theta = \theta)$$

cannot have any negative limit points.

For any bounded random variable $X$, denoting the bound by a (i.e., $|X| \leq a$),

$$(3.13) \qquad P[|X| \geq \epsilon] \geq \frac{EX^2 - \epsilon^2}{a^2}$$

Thus, letting $X = \bar{U}_N - g_N(\theta)$, for $\epsilon > 0$,

$$(3.14) \quad P[|\bar{U}_N - g_N(\theta)| > \epsilon \,|\, \Theta = \theta] \geq E|\bar{U}_N - g_N(\theta)|^2 - \epsilon^2 \geq \mathrm{Var}(\bar{U}_N \,|\, \Theta = \theta) - \epsilon^2.$$

By the consistent estimation of $\theta$, $\bar{U}_N - g_N(\theta) \to 0$ in probability as $N \to \infty$. Thus

$$P[|\bar{U}_N - g_N(\theta)| > \epsilon \,|\, \Theta = \theta] \to 0 \quad \text{as} \quad N \to \infty.$$

Thus, using (3.11) and (3.14), for each $\epsilon > 0$,

$$\frac{\sum_{i=1}^{N} \mathrm{Var}(U_i \,|\, \Theta = \theta)}{N^2} + \frac{2}{N^2} \sum_{1 \leq i < j \leq N} \mathrm{Cov}(U_i, U_j \,|\, \Theta = \theta) - \epsilon^2$$

has no positive limit points. But

$$\sum_{i=1}^{N} \mathrm{Var}(U_i|\theta = \theta)/N^2 \longrightarrow 0$$

for all $\theta$. Thus, for each $\theta$

$$\frac{2}{N^2} \sum_{1\le i<j\le N} \mathrm{Cov}(U_i, U_j|\theta = \theta) \tag{3.15}$$

has no positive limit points. But when (3.12) is used for each $\theta$, the expression in (3.15) has no negative limit points.

Thus

$$\frac{1}{\binom{N}{2}} \sum_{1\le i<j\le N} \mathrm{cov}(U_i, U_j|\theta = \theta) \to 0 \tag{3.16}$$

as $N \to \infty$. But, this same argument implies for fixed $\epsilon > 0$ and <u>any</u> nonsparse sequence $\{C_N, N \ge 1\}$ that

$$\frac{1}{\binom{N(C_N)}{2}} \sum_{i,j\in C_N,\ i\neq j} \mathrm{cov}(U_i, U_j|\theta = \theta) \to 0 \tag{3.17}$$

as $N \to \infty$.

Now suppose that $D_N(\theta) \to 0$ as $N \to \infty$. Thus it is easily seen that for some $\epsilon > 0$ there exists a subsequence $N'$ of the positive integers and subsets $C_{N'} \in \mathcal{C}_{N'}(\epsilon)$ and $\epsilon' > 0$ such that

$$\frac{1}{\binom{N(C_{N'})}{2}} \left| \sum_{i,j\in C_{N'},\ i\neq j} \mathrm{cov}(U_i, U_j|\theta = \theta) \right| > \epsilon', \tag{3.18}$$

for all $N'$ and such that every summand is the same sign. But this clearly contradicts (3.17). Thus $D_N(\theta) \to 0$ as $N \to \infty$, establishing EI for $\{\underline{U}_N, \theta\}$ and hence the desired essential unidimensionality. □

<u>Remarks</u>. It is interesting to note that Theorem 3.2 allows the consistent estimation of ability even if the IRFs are unknown to the practitioner. That is, use of $\bar{U}_N$ to estimate $A_N(\theta)$ <u>does not</u> require knowledge of the form of $A_N(\theta)$. As long as no attempt is being made to estabish a standardized ability

scale across tests (e.g., as a precurser to equating tests) knowledge of t IRFs is not required. Moreover, consistent estimation of ability with unk IRFs is possible in several populations being administered the same test - Section 5. Also note that the proof of Theorem 3.2 makes clear that when consistently estimated that $g_{C_N}(\theta) = A_{C_N}(\theta)$ always works in (3.6).

It is a foundationally relevant fact that essential unidimensionality implies under a mild and natural regularity condition for $\{\underline{U}_N\}$ that the ability is, in a certain sense, unique, as Theorem 3.3 below asserts.

Definition 3.3. Let a sequence of tests $\{\underline{U}_N\}$ be essentially unidimensio with respect to ability $\Theta$. Suppose for every fixed $\theta_1$ such that $\theta_1$ i the range R of $\Theta$ (i.e., $P[\Theta \subset R] = 1$ with R "minimal") that there ex $\epsilon_{\theta_1} > 0$ and an open neighborhood $H_{\theta_1}$ of $\theta_1$ such that for all $\theta_2 \in H_{\theta_1}$ range R of $\Theta$ that

$$(3.19) \qquad \frac{1}{N} \; \frac{\sum_{i=1}^{N} P_i(\theta_2) - P_i(\theta_1)}{\theta_2 - \theta_1} \geq \epsilon_{\theta_1} > 0 \quad \text{for all } N.$$

Then $\{\underline{U}_N, \Theta\}$ is said to be locally asymptotically discriminating (LAD) wi respect to $\Theta$.

Remark. What LAD really supposes is that $\sum_{i=1}^{N} P_i(\theta)/N$ is increasing faster some positive-slope linear function in some neighborhood of $\theta$ for every independent of N.

Theorem 3.3. Suppose $\{\underline{U}_N\}$ is essentially unidimensional with respect tc $\Theta$ and $\Theta'$. Let the corresponding marginal item response functions be denot

$$P_i(\theta) = E(U_i|\Theta = \theta), \; P_i'(\theta) = E(U_i|\Theta' = \theta)$$

for all $\theta$. Suppose $\{\underline{U}_N, \Theta\}$ is LAD with respect to $\Theta$. There then exists function g defined on the range R' of $\Theta'$ such that

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

and the range of g is R.

Remarks. Since a d=1, M, LI model is also an EI model, note that Theorem 3.4 holds for d=1, M, LI models as well. Thus Theorem 3.4 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 $\theta$ and $\theta'$

$$\bar{U}_N - A_N(\theta) \to 0 \tag{3.20}$$

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

$$\bar{U}_N - A'_N(\theta') \to 0 \tag{3.21}$$

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 $\theta$, $\theta'$. Then, for each $\theta$, $\theta'$ such that $G_{\theta,\theta'} \neq \phi$, (3.20) and (3.21) imply on $G_{\theta,\theta'}$ that

$$A_N(\theta) - A'_N(\theta') \to 0 \tag{3.22}$$

Fix $\theta' \in R'$ and let, denoting the empty set by $\phi$,

$$B_{\theta'} = \{\theta | G_{\theta,\theta'} \neq \phi\}$$

Note that $B_{\theta'} \neq \phi$ for all $\theta' \in R$ since each examinee has an ability value for both $\Theta$ and $\Theta'$. 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 (3.22) implies that

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

That is,

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

contradicting (3.19). Thus $B_{\theta'}$ consists of a unique $\theta$ for each $\theta'$: i.e., a function g is defined:

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

Choose $\theta'_2 > \theta'_1$ with $\theta'_1 \in R'$, $\theta'_2 \in R$. 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 $\{\underline{U}_N, \Theta'\}$ is M. By the definition of g, recalling (3.22), 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 the $P_i(\theta)$s. That is, g is monotone nondecreasing and well defined for all $\theta' \in R'$.

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

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

and

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

it follows that the range of g is R. □

Remarks. (1) Note that the theorem does not claim that g is strictly increasing. That is, the rescaling given by g could assign many $\theta'$ to the same $\theta$. Because no assumption analogous to (3.19) was made for $\Theta'$, this is of course expected, for the $\Theta'$ scale could produce a finer partition than needed to achieve essential unidimensionality. Thus the collapsing of distinct $\theta'$ into a single $\theta$ cannot be ruled out. The essential point is that, if for the $\Theta$ scale there exists an interval [a,b] such that,

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

then the $\Theta'$ scale should be rescaled so that all $\theta' \in [a,b]$ should be collapsed to a single point, say $\theta'_a$. However, assuming (3.19) for $\theta'$ as well does imply a strictly increasing g.

(2) In a private communication, Brian Junker has pointed out that an alternate proof of Theorem 3.4 can be given that produces g explicitly. It seems worthwhile to describe this construction: By the Helly Selection Theorem

for uniformly bounded increasing functions (Billingsley, [1968], p. 227) one can exhibit an integer subsequence $N_k$ and functions $A(\theta)$, $A'(\theta)$ such that for all $\theta$

$$A(\theta) = \lim_{k\to\infty} A_{N_k}(\theta), \quad A'(\theta) = \lim_{k\to\infty} A'_{N_k}(\theta).$$

By (3.19), $A(\theta)$ can be shown to be invertible. Junker's proof then shows that $g$ can be defined by

$$g(\theta') = A^{-1}(A'(\theta')) \tag{3.23}$$

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

It is an 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. In this context, the following example shows that it is possible to construct a sequence of tests $\{\underline{U}_N, N \geq 1\}$ that is not essentially unidimensional. Thus, the concept of essential unidimensionality is <u>not</u> mathematically vacuous.

<u>Example 3.3</u>. Let for each $r > 1$ $U_i = U_j$ for $2^{(r-1)^2} < i, j \leq 2^{(r^2)}$ and $U_i$ independent of $U_j$ otherwise. Let $P[U_i = 1] = P[U_i = 0] = 1/2$ for all $i$ define the marginal distributions. Then, letting

$$\bar{U}^{(r^2)} = \frac{\sum_{i=2^{(r-1)^2}+1}^{2^{(r^2)}} U_i}{2^{r^2} - 2^{(r-1)^2}},$$

it follows that

(3.24) $$P[\bar{U}^{(r^2)} = 1] = P[\bar{U}^{(r^2)} = 0] = 1/2,$$

that $\{\bar{U}^{(r^2)}, r \geq 1\}$ are independent, and that with probability one,

(3.25) $$\bar{U}_{r^2} \geq 1 - \frac{1}{2^{2r-1}} \quad \text{or} \quad \bar{U}_{r^2} \leq \frac{1}{2^{2r-1}} .$$

Suppose essential unidimensionality for $\{\underline{U}_N, \theta\}$. Then by Theorem 3.1, given $\Theta = \theta$,

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

in probability as $N \to \infty$.

Thus, given $\Theta = \theta$,

$$\bar{U}_{r^2} - A_{r^2}(\theta) \longrightarrow 0$$

as $r \to \infty$. Thus, by a standard probability argument, there exists a subsequence $r_i^2$ such that given $\Theta = \theta$

$$\bar{U}_{r_i^2} - A_{r_i^2}(\theta) \to 0$$

with probability one as $i \to \infty$. Thus, given $\Theta = \theta$,

(3.26) $$\bar{U}^{(r_i^2)} - A_{r_i^2}(\theta) \longrightarrow 0$$

with probability one as $i \to \infty$. Because $M$ is part of the assumption of essential unidimensionality, $A_{r_i^2}(\theta)$ is nondecreasing in $\theta$. A contradiction of this is now obtained. By (3.24), $\{\bar{U}^{(r_i^2)}, i \geq 1\}$ is a sequence independent identically distubuted random variables with marginal distribution $p(0) = p(1) = 1/2$. Let, for $i \geq 1$ and fixed $1/16 \geq \epsilon > 0$

$$A_{i,0} = [A_{r_i^2}(\theta) \leq \epsilon], \qquad A_{i,1} = [A_{r_i^2}(\theta) \geq 1 - \epsilon]$$

By (3.25), (3.26), and the distribution of $\{\bar{U}^{(r_i^2)}, i \geq 1\}$, we choose $i'$ so large that

$$(3.27) \qquad P[A_{i',0}] \geq 1/2 - \epsilon, \quad P[A_{i',1}] \geq 1/2 - \epsilon,$$
$$P[A_{i'+1,0}] \geq 1/2 - \epsilon, \quad P[A_{i'+1,1}] \geq 1/2 - \epsilon$$

and further that

$$(3.28) \qquad P[A_{i',0} \cap A_{i'+1,1}] \geq 1/4 - \epsilon, \quad P[A_{i',1} \cap A_{i'+1,0}] \geq 1/4 - \epsilon.$$

Because the events in (3.28) each have positive probability, they are each nonempty events of the probability space, $\Omega$ say. This will produce a contradiction as follows: Let $\omega_0$ and $\omega_1$ be outcomes of $\Omega$ such that

$$\omega_0 \in A_{i',0} \cap A_{i'+1,1}, \qquad \omega_1 \in A_{i',1} \cap A_{i'+1,0}$$

Then obviously $\omega_0 \in A_{i',0}$, $\omega_1 \in A_{i',1}$, so that $\theta(\omega_0) < \theta(\omega_1)$ by the monotonicity of $A_{r_i^2}(\theta)$. But, similarly, $\omega_1 \in A_{i'+1,0}$, $\omega_0 \in A_{i'+1,1}$ so that $\theta(\omega_1) < \theta(\omega_0)$, a contradiction. □

Theorems 3.2 has an interesting multidimensional analogue. Let for a latent model $(U_N, \underline{\theta})$ with item response functions $\{P_i(\underline{\theta}), 1 \leq i \leq n\}$,

$$(3.29) \qquad A_N(\underline{\theta}) = \sum_{i=1}^{N} P_i(\underline{\theta})/N.$$

(the distinction between $A_N(\underline{\theta})$ and $A_N(\theta)$ henceforth assumed clear from context).

Theorem 3.4. Suppose essential $d_E$ dimensionality with respect to ability $\underline{\theta}$. Then, $\underline{\theta}$ is able to be consistently estimated in probability in the sense of (3.6) with $g_{C_N}(\theta)$ replaced by $g_{C_N}(\underline{\theta})$ in (3.6).

Suppose that essential $d_E$ dimensionality fails for $\{U_N\}$. Then there do not exist a $d_E$ dimensional $\underline{\theta}$ and accompanying functions $g_{C_N}(\underline{\theta})$ such that $(U_N, \underline{\theta})$ is monotone and for each given $\underline{\theta} = \underline{\theta}$

given, into an operational definition with only the "observed" data given, that is with the distribution of the test $\underline{U}_N$ given. Of course, in any statistical application the "given" distribution of $\underline{U}_N$ is observed only with error because only a finite amount of data is ever available. With only the distribution of $\underline{U}_N$ given, then the definition must contend with the essential nonuniqueness of the IRFs and accompanying scale for $\theta$.

Definition 4.1. Let $\underline{U}_N^{(B)}$ and $\underline{U}_N^{(C)}$ represent a test administered to two populations, B and C. Then d-dimensional invariance holds provided each test administration has a d-dimensional M, LI representation using the same item response functions. That is, for each $\underline{u} \equiv (u_1,\ldots,u_N)$ with $\underline{\theta}$ and $\underline{\theta}'$ d-dimensional,

$$(4.1) \qquad P[\underline{U}_N^{(B)} = \underline{u}] = \int \prod_{i=1}^{N} \left[P_i(\underline{\theta})\right]^{u_i} \left[1 - P_i(\underline{\theta})\right]^{1-u_i} dP^{(B)}(\underline{\theta})$$

and

$$(4.2) \qquad P[\underline{U}_N^{(C)} = \underline{u}] = \int \prod_{i=1}^{N} \left[P_i(\underline{\theta}')\right]^{u_i} \left[1 - P_i(\underline{\theta}')\right]^{1-u_i} dP^{(C)}(\underline{\theta}')$$

where $P^{(B)}$ and $P^{(C)}$ are arbitrary distributions on $R^d$, d dimensional Euclidean space.

Remarks on Definition 4.1. (1) A key point to note is that the ability distributions $P^{(B)}$ and $P^{(C)}$ are arbitrary and in no way required to be related to one another. This amounts to allowing an arbitrary choice of ability metric for each population in an effort to obtain the same item response functions $\{P_i(\theta)\}$ in (4.1) and (4.2). The two metrics need not be the same in any mathematical or psychological sense. Nevertheless, once statistical evidence is given that (4.1) and (4.2) hold, it is standard IRT practice to declare that a common ability metric has been found.

(2) In applications, because the latent ability is usually assumed to be unidimensional, "invariance" usually means unidimensional invariance. For example, when invariance is used to justify a technique for identifying biased items, then the practitioner surely has unidimensional invariance in mind (see Lord (1980), Chapter 14, for example).

(3) Of course, once the IRFs for a model are specified, then invariance holds for all subpopulations of $\underline{\theta}$. For, an IRT model, once specified, by its very structure assigns to each examinee a fixed $\underline{\theta}$. Thus altering the distribution of $\underline{\theta}$ by choosing a subpopulation of examinees cannot change the IRFs. The distribution of $\underline{U}_N$ for the subpopulation is then derived from (4.2) with $P^{(C)}(\underline{\theta})$ the subpopulation distribution and the IRFs identical to those in (4.1), the expression for the entire population. Thus, the Lord viewpoint of fixing the latent variable $\underline{\theta}$ is appropriate when focusing on a subpopulation <u>after</u> the IRT model has been specified.

(4) Note that (4.1) and (4.2) really state that populations B and C being administered the test $\underline{U}_N$ each can be modeled by a M d=1 LI model.

The following idealized example, in the author's opinion, illustrates a fundamental flaw in the uncritical application of invariance.

Example 4.1. Consider two populations of examinees, males and females say. Let $\theta$ denote the unidimensional ability intended to be measured. Let $P_i(\theta)$, $1 \leq i \leq N$ denote a family of item response functions that satisfies (4.1) for males. Suppose that the items are <u>uniformly</u> biased against females in the sense that

$$P[U_i = 1 | \text{female of ability } \theta] = P_i(\theta - 1) \quad \text{for all } i, \theta \text{ and}$$

$$P[U_i = 1 | \text{male of ability } \theta] = P_i(\theta) \text{ for all } i, \theta.$$

Thus, for females, for all $\underline{u}$, with $P^F$ denoting the distribution of ability for females,

$$P[\underline{U}_N = \underline{u}] = \int \prod_{i=1}^{n} \left\{ \left[P_i(\theta - 1)\right]^{u_i} \left[1 - P_i(\theta - 1)\right]^{1-u_i} \right\} dP^F(\theta).$$

But, a simple change of variable $\theta' = \theta - 1$ yields

$$P[\underline{U}_N = \underline{u}] = \int \prod_{i=1}^{N} \left[P_i(\theta')\right]^{u_i} \left[1 - P_i(\theta')\right]^{1-u_i} dP^F(\theta' + 1)$$

Thus (4.2) holds with $P^F(\theta' + 1)$ the new ability distribution. Unidimensional invariance therefore holds, in spite of the pervasive (and uniform) sex bias in the test. □

The example is certainly idealized. For example, some items would surely be more biased than others in any actual application. But it represents a serious practical problem, in the author's view. Unidimensional invariance is no guarantee against failure to identify pervasive bias. What is really going on is that if administered simultaneously to males and females, the test is driven by a two-dimensional latent variable $(\theta_1, \theta_2)$, where $\theta_1$ is the ability to be measured and $\theta_2$ (= -1 for females, = 0 for males say) measures the degree of bias. For example, $\theta_1$ could be mathematical ability and $\theta_2$ could be familiarity with computers. However, the above example is easily seen to be unidimensional in the traditional sense. For, let $\theta = \theta_1 + \theta_2$. Then

$$P[\underline{U}_N = \underline{u}] = \int \prod_{i=1}^{n} \left[P_i(\theta)\right]^{u_i} \left[1 - P_i(\theta)\right]^{1-u_i} dP(\theta) \tag{4.3}$$

where $P(\theta) = P^F(\theta+1)$. Thus, pervasive bias is possible even when traditional unidimensionality holds. In this regard the following easily proved theorem is relevant.

Theorem 4.1. Let $\underline{U}_N^{(B)}$ and $\underline{U}_N^{(C)}$ represent a test administered to two populations, B and C. Then unidimensional invariance holds if and only if traditional unidimensionality holds.

Proof. Assume traditional unidimensionality. Then (4.3) holds for monotone IRFs $P_i(\theta)$. But then (really the context of Remark (3) above)

$$P[\underset{\sim}{U}_N = \underset{\sim}{u}|B] = \int \prod_{i=1}^{n} \left[P_i(\theta)\right]^{u_i} \left[1 - P_i(\theta)\right]^{1-u_i} dP_B(\theta)$$

follows trivially where $P_B(\theta) = P(\theta|B)$. The same holds for C; thus, unidimensional invariance holds.

Assume unidimensional invariance; i.e., (4.1) and (4.2) with $\theta$, $\theta'$ real valued. Let $\theta''$ be the $\theta$ of (4.1) for all Population B examinees and $\theta''$ be the $\theta'$ of (4.2) for all Population C examinees. Thus, each examinee of B U C is assigned a unidimensional $\theta''$. Then, for some $0 < p < 1$

$$P[\underset{\sim}{U}_N = \underset{\sim}{u}] = P[\underset{\sim}{U}_N = \underset{\sim}{u}|B]\, p + P[\underset{\sim}{U}_N = \underset{\sim}{u}|C](1-p)$$

$$= \int \prod_{i=1}^{N} \left[P_i(\theta'')\right]^{u_i} \left[1 - P_i(\theta'')\right]^{1-u_i} d\left\{pP^{(B)}(\theta'') + (1-p)\, P^{(C)}(\theta'')\right\}.$$

Clearly, letting $P''(\theta'') = pP^{(B)}(\theta'') + (1-p)P^{(C)}(\theta'')$ completes the proof.

□

Remark. The theorem shows us that (unidimensional) invariance is simply traditional unidimensionality.

Theorem 4.1 and the results concerning essential unidimensionality suggest that unidimensional invariance be redefined so that it dovetails with essential unidimensionality.

Definition 4.2. Let $\{\underset{\sim}{U}_N^{(B)},\ N \geq 1\}$ and $\{\underset{\sim}{U}_N^{(C)},\ N \geq 1\}$ represent a sequence of tests administered to two populations, B and C. Then essential d-dimensional invariance holds provided each test administered has an essential d-dimensional representation using the same latent model representation $P[\underset{\sim}{u}_N|\underset{\sim}{\theta} = \underset{\sim}{\theta}]$. That is, for each $\underset{\sim}{u} = (u_1, \cdots, u_N)$ with $\underset{\sim}{\theta}$ and $\underset{\sim}{\theta}'$ d-dimensional,

$$P\left[\underset{\sim}{U}_N^{(B)} = \underset{\sim}{u}\right] = \int P\left[\underset{\sim}{u}\middle|\underset{\sim}{\theta} = \underset{\sim}{\theta}\right] dP^{(B)}(\underset{\sim}{\theta}) \tag{4.4}$$

and

$$(4.5) \qquad P\left[\underline{U}_N^{(C)} = \underline{u}\right] = \int P\left[\underline{u} \middle| \underline{\theta}' = \underline{\theta}'\right] dP^{(C)}(\underline{\theta}')$$

where $P^{(B)}$ and $P^{(C)}$ are arbitary distributions on $\underline{R}^d$, d dimensional Euclidean space and (4.4) and (4.5) each define essential d-dimensional models.

The analogue of Theorem 4.1 is trival to state and prove.

Theorem 4.2. Let $\underline{U}_N^{(B)}$ and $\underline{U}_N^{(C)}$ represent a test administered to two populations, B and C. Then essential unidimensional invariance holds if and only if essential unidimensionality holds.

Proof. Same as that of Theorem 4.1, except for minor details. □

Example 4.1 compels us to be cautious concerning the centrality of the concept of invariance in IRT modeling. For, unidimensional invariance (whether essential or traditional) clearly does not preclude the inappropriate assignment of a common metric to the underlying ability of interest in a single test, two population problem. Can something be substituted for unidimensional invariance that will rule out such faulty applications? We suggest that the central property that must hold in such single-test, multiple-population applications is essential unidimensionality together with the conclusion that the underlying essentially unidimensional ability $\theta$ is the ability intended to be measured. In the above example, $\theta_1$ was the ability intended to be measured rather than $\theta = \theta_1 + \theta_2$. This suggests the following definition.

Definition 4.3. A test sequence $\{\underline{U}_N, N \geq 1\}$ is said to be valid provided (i) it is essentially unidimensional with respect to $\theta$ and (ii) $\theta$ is the ability desired to be measured.

Certain results in Section 3 support the appropriateness of this definition. First, Theorem 3.3 states that, under mild regularity conditions, essential unidimensionality with respect to $\theta$ guarantees that, up to monotone

transformations, the $\theta$ of the model is unique. That is, essential unidimensionality makes the measurement of $\theta$ well defined in the sense that $\theta$ itself is "unique" and hence well defined. Second, Theorem 3.2 shows that $\underset{\sim}{U}_N$, through computation of the statistic $\tilde{U}_N$, can be used to consistently estimate $\theta$ by use of the rescaling $A_N(\theta)$. That is, the data can be used operationally to obtain $\theta$, as one would expect a "valid" test to be able to do.

Of course, IRT validity as defined above requires both essential unidimensionality and that the underlying latent ability $\theta$ is the ability intended to be measured. Statistical analysis of data from the administration of a test cannot in the absence of additional data concerning other valid tests, external criteria, etc. be used to ascertain whether the latent ability being measured is the one intended. However, statistical analysis of data from the administration of a test <u>can</u> be used to assess whether the prerequisite essential unidimensionality holds. Moreover, as remarked above, the author's (Stout, 1987) statistical test of unidimensionality is designed to address precisely this question of whether essential unidimensionality holds.

One final point must be emphasized. If essential unidimensionality holds for a combined multiple-population test, then it is purely a matter of taste and convenience which transformation of the the underlying ability $\theta$ is used for the ability scale. In Section 3, $A_N(\theta)$ is used for the one population case because it makes the basic estimation results especially easy to formulate. Clearly, if one wishes to use a common metric for two or more populations being administered the same test, then the $A_N(\theta)$ of the combined superpopulation is totally appropriate. That is, the theory of Section 3 easily extends to the fixed test multiple population setting. This is developed in Section 5.

5. Two Group Test Bias. In this section, we will apply the theory of Section 3 to the situation in which the test population is assumed to consist of two groups of examinees, B and C. The main objective is to assess whether the estimation of ability is somehow "unfair" to Group B as compared with group C, or vice versa. Let $\theta$ be the unidimensional ability intended to be measured in the combined population. Let $\theta^B$ and $\bar{U}_N^B$ denote respectively the ability intended to measured and the test scores of a randomly chosen examinee from Group B. Define $\theta^C$ and $\bar{U}_N^C$ similarly. Note that by definition $\theta = \theta^B$ for each Group B examinee and $\theta = \theta^C$ for each Group C examinee.

The results of Section 3 suggest that essential unidimensionality implies consistent estimation of $\theta$ in each group.

Theorem 5.1. If essential unidimensionality holds for $\theta$ in the combined population consisting of Group B and Group C examinees, then $\theta$ is able to be consistently estimated in each population using the $A_N(\theta) \equiv E[\bar{U}_N|\theta]$ scale computed from the combined population.

Proof. Fix $\theta$. By Theorem 3.2, given $\theta = \theta$,

$$\bar{U}_N - A_N(\theta) \to 0 \tag{5.1}$$

in probability as $N \to \infty$. Let $\mathcal{B}$ denote the event that a randomly sampled examinee (according to the distribution of $\theta$) is a Group B examinee. Fix $\epsilon > 0$. Let $G_N = \left[|\bar{U}_N - A_N(\theta)| > \epsilon\right]$. It is an elementary fact of probability that $P[G_N] \to 1$, and $P[\mathcal{B}] > 0$ implies that $P[G_N|\mathcal{B}] \to 1$. Thus given $\theta^B = \theta$, it follows that

$$\bar{U}_N^B - A_N(\theta) \to 0 \tag{5.2}$$

in probability as $N \to \infty$. The argument is the same for any nonsparse sequence $\{C_N, N \geq 1\}$. Thus, the result is proved for Group B. The argument is the same for Group C. □

Since essential unidimensionality guarantees consistent estimation of $\theta$ in each group, this suggests that a natural setting for the study of test bias is under the assumption that $\{U_N, \underline{\theta}\}$ is essentially $d_E$ dimensional for some $d_E > 1$. Essential $d_E$ dimensionality for some $d_E > 1$ is assumed throughout the remainder of Section 5. We suppose throughout Section 5 that $\underline{\theta}$ determines $\theta$; i.e., $\theta$ is a function of $\underline{\theta}$. Then, without loss of generality, we assume that the first component of $\underline{\theta}$ is the ability $\theta$ intended to be measured. Thus $\underline{\theta} = (\theta, \underline{\theta}_2)$ where $\underline{\theta}_2$ consists of "nuisance" abilities. We first note that essential $d_E$ dimensionality guarantees that the proportion correct consistently estimates $A_N(\underline{\theta})$ among all examinees of ability $\underline{\theta}$ regardless of their group membership.

Theorem 5.2. Let $\underline{\Theta}^B$ and $\bar{U}_N^B$ denote respectively the ability and the test score of a randomly chosen examinee from Group B. Define $\underline{\Theta}^C$ and $\bar{U}_N^C$ analogously. Then, for each nonsparse sequence $\{C_N, N \geq 1\}$ and for each $\underline{\theta}$, given $\underline{\Theta}^B = \underline{\theta}$,

$$\bar{U}_{C_N}^B - A_{C_N}(\underline{\theta}) \to 0 \tag{5.3}$$

in probability as $N \to \infty$; and for each $\underline{\theta}$, given $\underline{\Theta}^C = \underline{\theta}$,

$$\bar{U}_{C_N}^C - A_{C_N}(\underline{\theta}) \to 0 \tag{5.4}$$

in probability as $N \to \infty$.

Proof. Essentially the same as that of Theorem 5.1. □

We propose the following definition of test bias.

Definition 5.1. We say that there is no test bias in the estimation of $\Theta$ if, for each $\theta$, the distribution of $\underline{\Theta}^B$ given $\Theta^B = \theta$ is equal to that of $\underline{\Theta}^C$ given $\Theta^C = \theta$.

Remark. It is essential to note that this definition of test bias places r restrictions on the distributions of the ability to be measured for Group E Group C. The existence of test bias rests in the conditional distributions $\underline{\Theta}_2^B$ given $\underline{\Theta}^B$ and $\underline{\Theta}_2^C$ given $\underline{\Theta}^C$ and not in the marginal distributions of and $\underline{\Theta}^C$. The point is that bias is not to be mistaken for a genuine differ between the two groups in the ability $\Theta$ to be measured. Rather, it rests group ability differences in other attributes also influencing the test in essential way among examinees with the same $\Theta$ ability.

It is useful to illustrate the role of Definition 5.1 with a simple example.

Example 5.1. Suppose essential 2-dimensionality for $\underline{\Theta}$. Let $\underline{\theta} = (\theta, \theta_2)$ fix $\Theta = \theta$. Let $\Theta_2$ be a discrete random variable with range {0,1}. Supp for Group B that, given $\Theta = 0$, then $\Theta_2 = 0$ with probability 3/4 and th for Group C, given $\Theta = 0$, then $\Theta_2 = 1$ with probability 3/4. Recall that $A_N(\underline{\theta}) = E[\bar{U}_N|\underline{\theta}] = \sum_{i=1}^{N} P_i(\underline{\theta})/N$. Suppose the item response functions are su that $A_N(0, 0) = 1\ 8$, $A_N(0, 1) = 7/8$ for all N. Then, according to Theorem given $\underline{\Theta} = (0, 0)$,

$$\bar{U}_N^B - \frac{1}{8} \to 0 \quad \text{and} \quad \bar{U}_N^C - \frac{1}{8} \to 0,$$

each in probability as $N \to \infty$. Also, given $\underline{\Theta} = (0,1)$,

$$\bar{U}_N^B - \frac{7}{8} \to 0 \quad \text{and} \quad \bar{U}_N^C - \frac{7}{8} \to 0,$$

each in probability as $N \to \infty$. Note that

$$P[\Theta_2^B = 1 \mid \Theta^B = 0] = \frac{1}{4}, \; P[\Theta_2^C = 1 \mid \Theta^C = 0] = \frac{3}{4}.$$

Thus test bias in the sense of Definition 5.2 exists. How does this test b affect the asymptotic behavior of $\bar{U}_N^B$ and $\bar{U}_N^C$ ? For Group B,

$$P[\bar{U}_N^B \quad \frac{1}{8} \to 0 \mid \Theta^B = 0] = \frac{3}{4}, \; P[\bar{U}_N^B - \frac{7}{8} \to 0 \mid \Theta^B = 0] = \frac{1}{4};$$

while for Group C,

$$P[\bar{U}_N^C - \frac{1}{8} \to 0 \mid \theta^C = 0] = \frac{1}{4}, \quad P[\bar{U}_N^C - \frac{7}{8} \to 0 \mid \theta^C = 0] = \frac{3}{4}.$$

Thus, even though (5.3) and (5.4) hold, the behaviors of $\bar{U}_N^B$ and $\bar{U}_N^C$ in their attempts to estimate $\theta$ (on some scale) clearly favor Group C over Group B for examinees of ability $\theta = 0$. Here the asymptotic distribution of $\bar{U}_N^C$ given $\theta^C = 0$ is stochastically larger than the distribution of $\bar{U}_N^B$ given $\theta^B = 0$. □

Note that the marginal distributions of the ability to be measured in Group B and in Group C played no role in the example.

Recall that essential $d_E$ dimensionality for $d_E > 1$ implies by Theorem 3.7 that there do not exist functions $g_N(\theta)$ such that

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

in probability as $N \to \infty$. This precludes consistent estimation of $\theta$. However, if there is no test bias (in the sense of Definition 5.1), then using proportion correct to score the test is guaranteed not to favor either group over the other asymptotically in any way whatsoever, as the following theorem makes precise.

Theorem 5.3. Suppose there is no test bias. Let $\{C_N, N \geq 1\}$ be any nonsparse sequence. Then for each $\theta$, the asymptotic distribution of $\bar{U}_{C_N}^B$ given $\theta^B = \theta$ is the same as the asymptotic distribution of $\bar{U}_{C_N}^C$ given $\theta^C = \theta$.

Proof. Fix $\theta$. The argument used to prove Theorem 5.1 is easily modified to establish that, given $\theta^B = \theta$,

$$\bar{U}_{C_N}^B - A_{C_N}(\theta, \underline{\theta}_2^B) \to 0 \tag{5.5}$$

in probability as $N \to \infty$, and that, given $\theta^C = \theta$,

$$\bar{U}_{C_N}^C - A_{C_N}(\theta, \underline{\theta}_2^C) \to 0 \tag{5.6}$$

in probability as $N \to \infty$. Here $A_N(\underline{\theta}) \equiv A_N(\theta, \underline{\theta}_2)$. But absence of test bias merely means that the distribution of $\underline{\theta}_2^B$ given $\theta^B = \theta$ is the same as the distribution of $\underline{\theta}_2^C$ given $\theta^C = \theta$. The desired result then follows from (5.5) and (5.6). □

Remarks. (1) The point of the theorem is clear: If test bias does not exist when $d_E > 1$, then $\bar{U}^B_{C_N}$ and $\bar{U}^C_{C_N}$ are equally inconsistent in their respective attempts to estimate $\theta$. That is, each group is equally mistreated.

(2) At the end of Section 3, we pointed out that essential unidimensionality for a latent trait family of models $\{\underline{U}_N, \theta\}$ still allows for fixed measurement error for a finite length test administration, this caused by the presence of abilities other than $\theta$, these other abilities being inessential. Clearly, in the two population setting of Section 5, it similarly holds that this source of measurement error can favor one population over another in a finite length test administration, even when essential unidimensionality holds. From a lack of model fit perspective, the issue becomes one of whether the magnitude of the differences $A^B_N(\theta) - A^C_N(\theta)$ as $\theta$ varies are too large to be ignored. Here $A^B_N(\theta) = E\left[\frac{1}{N}\sum_{i=1}^{N} P_i(\underline{\theta}^B) | \theta^B = \theta\right]$ and $A^C_N(\theta)$ is defined similarly, where $\{\underline{U}_N, \theta, N \geq 1\}$ is assumed to have $d_E = 1$ and traditional dimensionality $\dim(\theta) > 1$. Because $\dim(\theta) > 1$, $A^B_N(\theta)$ and $A^C_N(\theta)$, the Group B and Group C intrinsic ability scales defined by (3.1) and (3.2) will in general be different.

(3) The theoretical results of this paper caution against the casual use of short tests with confidence that test bias will not occur. For, the shorter the test, the harder it is to assess essential unidimensionality. Furthermore, even if essential unidimensionality holds, the shorter the test, the more likely (3.30) (or the opposite inequality) is to hold to a damaging degree. By contrast Theorem 5.1 guarantees for a long essentially unidimensional test that (3.30) will have little to no ill effect in the consistent estimation of the essential trait in each population.

6. Essential Unidimensionality and Linear Formula Scoring. Thus far we have presented our thesis in the context of ability estimation with proportion right used as the estimator of ability. Of course, one of the major contributions of IRT has been the establishment that the use of a "linear formula score"

$$\sum_{i=1}^{N} a_{i,N} U_i \quad , \text{ all } \quad a_{iN} \geq 0$$

can be more appropriate than the use of $\bar{U}_N$. For example, with $N$ fixed in the M, LI case of two parameter logistic modeling, $\sum_{i=1}^{N} a_{i,N} U_i$ is a sufficient statistic for $\theta$ provided $a_{i,N}$ is proportional to the discrimination parameter of the ith item.

Definition 6.1. A sequence of linear formula scores with coefficients $\{a_{i,N}; 1 \leq i \leq N, N \geq 1\}$ is called admissible if

$$0 \leq a_{i,N} \leq \frac{K}{N} \quad \text{for all} \quad i, N \tag{6.1}$$

for some constant K.

Remarks. Several special cases of linear formula scores are admissible. First $a_{i,N} = 1/N$ for $i \leq N$, yielding $\{\bar{U}_N, N \geq 1\}$ is clearly admissible. Second $a_{i,N} = 1/N(C_N)$ for all $i \in C_N$ and equal to zero otherwise with $\{C_N, N \geq 1\}$ a nonsparse sequence of integer subsets clearly yields an admissible sequence of linear formula scores since by definition $N(C_N) \geq \epsilon N$ for all $N$ for some $\epsilon > 0$. Third, suppose a two parameter logistic model for $\{U_i, i \geq 1\}$ with discrimination parameters $a_i$ satisfying

$$0 < \epsilon \leq a_i \leq K < \infty \quad \text{for all } i. \tag{6.2}$$

Then, the normalized sufficient statistic

$$\frac{\sum_{i=1}^{N} a_i U_i}{\sum_{i=1}^{N} a_i} \tag{6.3}$$

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

It can be argued that most useful linear scoring methods surely satisfy (6.1) since (6.1) merely requires a scoring method where no single item or small subset of items is allowed undue influence on the overall formula score.

Note that if $\sum_{i=1}^{N} a_{i,N} \to 0$ as $N \to \infty$, then the corresponding formula score is of no practical interest since

$$\tilde{U}_N \equiv \sum_{i=1}^{N} a_{i,N} U_i \to 0$$

with probability one as $N \to \infty$ and hence cannot estimate anything consistently (other than 0). Thus each admissible linear formula score of interest should also satisfy for some $\epsilon > 0$

$$\sum_{i=1}^{N} a_{i,N} \geq \epsilon \quad \text{for all} \quad N \geq 1. \tag{6.4}$$

Definition 6.2 An admissible sequence of linear formula scores with coefficients satisfying (6.4) is called an admissible nonsparse sequence.

In this regard, note that if $a_{i,N} > 0$ only on a sparse sequence of integer subsets $C_N$ (i.e., $C_N \subset \{1,\ldots,N\}$, where $N(C_N)/N \to 0$ as $N \to \infty$), then either (6.1) or (6.4) is violated. That is, an attempt to use a "sparse" sequence of items to estimate $\theta$ will either result in an inadmissible sequence of linear formula scores or one which is sparse (in the sense that (6.4) fails) and is hence useless.

It is now possible to advance a theory very similar to that of Sections 2 - 5 that includes admissible nonsparse linear formula scoring. For example, the analogue of the sufficiency part of Theorem 3.2 is as follows:

Theorem 6.1. Let $\{U_N, \Theta\}$ be essentially unidimensional with respect to $\theta$. Then, for each given $\Theta = \theta$ and each set of nonsparse admissible linear formula scores

$$\tilde{U}_N = \sum_{i=1}^{N} a_{i,N} U_i$$

it follows that

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

in probability as $N \to \infty$ where

$$\tilde{A}_N(\theta) \equiv \sum_{i=1}^{N} a_{i,N} P_i(\theta)$$

Proof. Fix $\theta$. It suffices to show that

$$\mathrm{Var}\left[\sum_{i=1}^{N} a_{i,N} U_i \,\middle|\, \Theta = \theta\right] \to 0.$$

Now

$$\mathrm{Var}\left[\sum_{i=1}^{N} a_{i,N} U_i \,\middle|\, \Theta = \theta\right] = \sum_{i=1}^{N} a_{i,N}^2 \,\mathrm{Var}(U_i|\Theta = \theta)$$
$$+ 2 \sum_{1 \le i < j \le N} a_{i,N}\, a_{j,N}\, \mathrm{Cov}(U_i, U_j|\Theta = \theta) \to 0 \quad \text{as} \quad N \to \infty$$

by admissibility and essential unidimensionality; that is, (6.1) holds. □

The use of $\tilde{U}_N$ as estimator in the development of the theory in Sections 2 - 5 was done for simplicity and clarity of exposition and not because of necessity. Further generalization along the lines of Sections 2 - 5 using nonsparse admissible linear formula scoring is routine and is left to the reader.

## 7. Essential Unidimensionality and Consistent Estimation of $\theta$ on the $\theta$ scale.

The use of $\{\tilde{U}_N\}$ or more generally of a nonsparse admissible linear formula score $\{\Sigma_{i=1}^{N} a_{i,N} U_i\}$ as a sequence of estimators of $\theta$ on the intrinsic ability scale when essential unidimensionality holds supposes a single fixed test administered to one or more populations. Applications in this setting were developed in Sections 3, 5, and 6. Such single-test applications of IRT occur less frequently in practice than multiple-population multiple-test applications.

In multiple-test applications a standardized ability scale is usually desired, perhaps as a prerequisite to a horizontal equating of the various

tests. In many such applications, the items (or at least a common core of them) have been calibrated relative to the constructed standardized ability scale $\theta$. Then a major application is the estimation of individual abilities on the $\theta$ scale. Estimation of $\theta$ 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 $\hat{\theta}$ "consistently" estimates $\theta$ under suitable regularity conditions in the sense that, given $\Theta = \theta$, $\hat{\theta} \to \theta$ in probability as the number of items $N \to \infty$. (Here and throughout Section 7, "consistency" is used in its usual mathematical statistical sense and is not to be confused with its special usage as given by Definition 3.2.) Only rarely however, is it possible to provide an explicit formula for the MLE as a function of $\underline{U}_N$. Moreover, the MLE is usually a highly non-linear function of $\underline{U}_N$. Thus in the known IRFs case it seems desirable to seek alternatives to MLE that are based on linear formula scoring and for which explicit formulae are available. We now propose a family of such estimators, using the results of Sections 3 and 6.

Recall from Theorem 3.2 that when $\{\underline{U}_N\}$ is essentially unidimensional with respect to $\Theta$ then for each given $\Theta = \theta$,

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

in probability as $N \to \infty$. This suggests estimating $\theta$ 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 as $N \to \infty$ should hold. Moreover, recalling Theorem 6.1 and its notation, this result should generalize to nonsparse admissible linear formula scoring with, for each $\Theta = \theta$,

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

in probability as $N \to \infty$. Theorem 7.1 below states that this is true provided local asymptotic discrimination holds. Definition 7.1 is the appropriate analogue of Definition 3.3.

Definition 7.1. Let a sequence of tests be essentially unidimensional with respect to ability $\theta$. Let $\tilde{A}_N(\theta) = \Sigma_{i=1}^N a_{ni} P_i(\theta)$ be formed from a nonsparse admissible sequence of linear formula scores. Suppose for every fixed $\theta_1$ such that $\theta_1$ is in the range $R$ of $\Theta$ that there exists $\epsilon_{\theta_1} > 0$ and an open neighborhood $H_{\theta_1}$ of $\theta_1$ such that for all $\theta_2 \in H_{\theta_1}$ and in the range $R$ of $\Theta$ with $\theta_2 > \theta_1$ that

$$(7.1) \qquad \frac{\sum_{i=1}^N a_{iN} P_i(\theta_2) - \sum_{i=1}^N a_{iN} P_i(\theta_1)}{\theta_2 - \theta_1} \geq \epsilon_{\theta_1} (\theta_2 - \theta_1) \quad \text{for all } N.$$

Then $\{U_N, \Theta, \tilde{A}_N(\theta)\}$ is said to be locally asymptotically discriminating (LAD) with respect to $\Theta$.

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

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

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

Theorem 7.1. Let $\{U_N, \Theta\}$ be essentially unidimensional with respect to $\theta$. Suppose $\tilde{A}_N(\theta) = \Sigma_{i=1}^N a_{ni} P_i(\theta)$ is formed from a nonsparse admissible sequence of linear formula scores $\tilde{U}_N = \Sigma_{i=1}^N a_{ni} U_i$. Suppose $\{U_N, \Theta, \tilde{A}_N(\theta)\}$ is LAD with respect to $\Theta$. Then, for each given $\Theta = \theta$,

$$(7.2) \qquad \tilde{A}_N^{-1}(\tilde{U}_N) \to \theta$$

in probability as $N \to \infty$.

Proof. Fix $\theta$. By Theorem 6.1, given $\Theta = \theta$,

$$\tilde{U}_N - \tilde{A}_N(\theta) \to 0 \tag{7.3}$$

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

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

with probability one as $k \to \infty$. Choose $\{N(j)\}$. Then (7.3) implies that

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

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

$$\tilde{U}_{N(j(k))} - \tilde{A}_{N(j(k))}(\theta) \to 0 \tag{7.4}$$

with probability one. By (7.1) 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

$$|\tilde{A}_N^{-1}(\mu_2) - \theta| \le K_\theta \, |\mu_2 - \tilde{A}_N(\theta)| \quad \text{for all } \theta. \tag{7.5}$$

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

$$\tilde{U}_{N(j(k))} \le \inf_\theta A_{N(j(k))}(\theta). \tag{7.6}$$

By (7.4) and LAD, there exists $\epsilon_k \downarrow 0$ such that for all large k

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

Thus $\tilde{U}_{N(j(k))} > \inf_\theta A_{N(j(k))}(\theta)$ for all sufficiently large k using LAD. Hence (7.6) cannot hold for arbitrarily large k.

Thus, combining (7.5) with (7.4),

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

with probability one, as required. □

Remarks. (1) Theorem 7.1 provides a large class of sequences of estimators of $\theta$, 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.

(2) It is to be noted that Holland, Junker, and Thayer (1987) have proposed using $\{A_N^{-1}(\bar{U}_N)\}$ to estimate the distribution of $\theta$ 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.

(3) It is elementary to show that (7.2) holding for all $\theta$ implies

$$\tilde{A}_N^{-1}(\tilde{U}_N) \to \theta \tag{7.5}$$

in probability as $N \to \infty$. Given the IRT context, (7.2) is perhaps a more interesting formulation than (7.5). It does of course follow from (7.5) that $\tilde{A}_N^{-1}(\tilde{U}_N)$ can be used as a method of estimating the distribution of $\theta$.

(4) Note that (7.2) states that convergence in probability to individual ability holds regardless of which of a large class of estimators is used. That is, convergence in probability to individual ability holds for every nonsparse admissible sequence of linear formula scores.

8. Discussion and Summary of Results. The purpose of the paper is to argue that a successful approach to certain fundamental test measurement topics such as bias requires multidimensional latent modeling. 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 mesh with the empirical reality of multiply determined items, attempts to count only the dominant dimensions. Theorems 2.2 and 2.3 present conditions that guarantee that essential unidimensionality holds.

In Section 3, essential unidimensionality is shown in Theorems 3.2 and 3.3 to characterize the consistent estimation of a unidimensional latent trait. Here the "consistent estimation of $\theta$" is defined precisely in Definition 3.2. In order to facilitate this, the concepts of marginal item response functions and intrinsic ability scale are presented. This theory is applicable to single-test applications and does not require that the IRFs be known (i.e., calibrated).

Theorem 3.3 shows that essential unidimensionality guarantees, under the mild regularity condition of local asymptotic discrimination of $\{\underline{U}_N\}$, that the latent ability is unique up to monotone transformations. That is, essential unidimensionality, an empirically testable condition, guarantees that the latent trait of the model is "well defined" in the sense that it is completely specified, up to a monotone transformation. Loosely stated, the "data" (distributions of $\underline{U}_N$'s) determine the latent trait when essential unidimensionality holds. The above results, as with most of the results of the paper, requires an infinite item pool formulation. It is the author's position that such a formulation facilitates the study of many foundational IRT issues.

Example 3.3 shows that the concept of essential $d_E$ dimenionality for some $d_E < \infty$ is not vacuous by showing that test sequences exist that could be intuitively described as essentially $\infty$-dimensional. It is pointed out that

essential unidimensionality, when other (minor) dimensions besides the ability of interest $\theta$ are present, does not rule out non-random pre-asymptotic bias in the estimation of $\theta$ for short tests.

In Section 4, the uncritical acceptance of the centrality of the role of item parameter invariance is challenged. In particular, Example 4.1 shows that invariance (precisely, unidimensional invariance) can hold and yet pervasive bias against a particular group still exist. It is shown that unidimensional invariance is equivalent to traditional unidimensionality holding. Essential d-dimensional invariance is defined by replacing local independence by essential independence in the definition of invariance. Then essential unidimensional invariance is shown to be equivalent to essential unidimensionality. These results motivate a simple latent trait based definition of validity, namely that validity holds if (i) essential unidimensionality holds and (ii) the (unique by Theorem 3.4) essential trait $\theta$ is the ability intended to be measured.

Section 5 addresses the issue of test bias in a single-test two group setting from the viewpoint of consistent estimation. Essential unidimensionality is shown to guarantee consistent estimation of ability in both groups. Thus the issue of test bias can be analyzed assuming that essential $d_E$ dimensionality holds for some $d_E > 1$. The test bias problem is then characterized as the estimation of the intended to be measured $\theta$ for a two-group latent model with essential $d_E$ dimensional latent ability $(\theta, \underset{\sim}{\theta}_2)$, where the "nuisance" ability $\underset{\sim}{\theta}_2$ is $d - 1$ dimensional. Test bias for Groups B and C is defined as the conditional distributions of $\underset{\sim}{\theta}_2^B$ given $\theta^B = \theta$ and $\underset{\sim}{\theta}_2^C$ given $\theta^C = \theta$ differing for at least one value of $\theta$. It is stressed that the marginal distributions of $\theta^B$ and $\theta^C$ play no role in the definition. Example 5.1 demonstrates the role that test bias as herein defined plays in the attempts to

estimate $\theta$ in Groups B and C. Finally, Theorem 5.3 shows, recalling that $d_E > 1$ assumed, that the lack of test bias implies that $\bar{U}_N^B$ and $\bar{U}_N^C$ are equally inconsistent in their attempts to estimate $\theta$; that is, neither population B or C is favored over the other.

Section 6 demonstrates that the theory of Sections 1 - 5 need not be presented only in the context of the behavior of $\bar{U}_N$, but that admissible linear formula scoring can be used as a basis for Sections 2 - 5 with only minor alterations required. This development is largely left to the reader. Here an admissible linear formula score $\{\sum_{i=1}^{N} a_{i,N} U_i,\ N \geq 1\}$ is one where $0 \leq a_{i,N} \leq K N$ for all $i$ and some fixed $K < \infty$. It is noted that most linear formula scores of interest are admissible.

Section 7 addresses the problem of estimating $\theta$ on the $\theta$ scale in the multiple-test multiple-population problem with known IRFs assumed. Theorem 7.1 establishes that $\theta$ can be consistentely estimated on the $\theta$ scale by a large class of sequences of estimators in the sense that for each such sequence,

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

in probability as $N \to \infty$. Each such sequence is computable, has an explicit formula, and is based on an admissible linear formula score in an intuitive natural way.

Acknowledgements. I wish to thank Brian Junker for the useful comments and discussions resulting from his careful and thoughtful reading of the manuscript. I wish to thank Charles Davis for useful commentary that lead to several improvements, in particular for Section 4. I wish to thank D.R. Divgi for an insight that lead to a useful reformation of (2.3).

## References


Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

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

Harmon, H.H. (1965). Modern Factor Analysis. University of Chicago Press, Chicago.

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

Holland, P.W. and Rosenbaum, P. (1987). Conditional association and unidimensionality in monotone latent variable models. To appear Annals of Statistics.

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

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 development in item response theory. Manuscript.

Roznowski, M. (1986). The use of tests manifesting sex differences as measures of intelligence: implications for measurement bias. In press.

Segall (1983). Assessment and comparison of techniques for transforming parameters to a common metric in item response theory. Thesis, University of Illinois at Urbana-Champaign.

Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics, Wiley, New York.

Stout, W. (1987). A nonparametric approach for assessing latent trait dimensionality. Psychometrika, 52,589-617.

Tatsouka, K.K. and Tatsouka, M.M. (1986). Diagnosis of Cognitive Errors by Statistical Pattern Recognition Methods. Manuscript.

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.

University of Illinois/Stout

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
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
Educational Testing Service
Princeton, NJ 08450

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

Dr. Arthur S. Blaiwes
Code N711
Naval Training Systems Center
Orlando, FL 32813

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 Onderzoe
Rekruterings-En Selectiecentr
Kwartier Koningen Astrid
Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code N-095R
Naval Training Systems Center
Orlando, FL 32813

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

Dr. Lyle D. Broemeling
ONR Code 1111SP
800 North Quincy Street
Arlington, VA 22217

Mr. James W. Carey
Commandant (G-PTE)
U.S. Coast Guard
2100 Second Street, S.W.
Washington, DC 20593

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

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

Dr. Robert Carroll
OP 01B7
Washington, DC 20370

Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

University of Illinois/Stout

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

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 Crombag
University of Leyden
Education Research Center
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

Mr. Timothy Davey
University of Illinois
Educational Psychology
Urbana, IL 61801

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 Building
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
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
201 Fishburne Bldg.
Atlanta, GA 30322

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

Dr. Pat Federico
Code 511
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
Program
P.O. Box 168
Iowa City, IA 52240

Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Dr. Myron Fischl
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

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
Bolling AFB, DC 20332

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

Dipl. Pad. Michael W. Habon
Universitat Dusseldorf
Erziehungswissenschaftliches
Universitatsstr. 1
D-4000 Dusseldorf 1
WEST GERMANY

Dr. Ronald K. Hambleton
Prof. of Education & Psychology
University of Massachusetts
at Amherst
Hills House
Amherst, MA 01003

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

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

Dr. Paul W. Holland
Educational Testing Service
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 90010

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
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

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

Dr. Robert Jannarone
Department of Psychology
University of South Carolina
Columbia, SC 29208

Dr. Dennis E. Jennings
Department of Statistics
University of Illinois
1409 West Green Street
Urbana, IL 61801

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

Dr. Milton S. Katz
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

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
University of Texas-Austin
Measurement and Evaluation
Center
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
San Diego, CA 92152-6800

Dr. Daryll Lang
Navy Personnel R&D Center
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

University of Illinois/Stout

Dr. Charles Lewis
Educational Testing Service
Princeton, NJ 08541

Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

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. Milton Maier
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. William L. Maloy
Chief of Naval Education
and Training
Naval Air Station
Pensacola, FL 32508

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

Dr. Clessen Martin
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

Dr. James McBride
Psychological Corporation
c/o Harcourt, Brace,
Javanovich Inc.
1250 West 6th Street
San Diego, CA 92101

Dr. Clarence McCormick
HQ, MEPCOM
MEPCT-P
2500 Green Bay Road
North Chicago, IL 60064

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

Dr. Robert McKinley
Educational Testing Service
20-P
Princeton, NJ 08541

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

Dr. Barbara Means
Human Resources
Research Organization
1100 South Washington
Alexandria, VA 22314

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
Oklahoma City, OK 73069

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

Fleet Support Office,
NPRDC (Code 301)
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 Olson
WICAT, Inc.
1875 South State Street
Orem, UT 84057

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
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

Administrative Sciences Department,
Naval Postgraduate School
Monterey, CA 93940

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/MP
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

University of Illinois/Stout

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

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
3108 AustinPeay 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
Fujusawa 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
Alexandria, VA 22314

Dr. Richard E. Snow
Department of Psychology
Stanford University
Stanford, CA 94306

Dr. Richard 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. Peter Stoloff
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

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

Maj. Bill Strickland
AF/MPXOA
4E168 Pentagon
Washington, DC 20330

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
San Diego, CA 92152-6800

University of Illinois/Stout

Dr. John Tangney
AFOSR/NL
Bolling AFB, DC 20332

Dr. Kikumi Tatsuoka
CERL
252 Engineering Research
Laboratory
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 310
St. Paul, MN 55114

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

Dr. Howard Wainer
Division of Psychological Studies
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
Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Dr. Brian Waters
Program Manager
Manpower Analysis Program
HumRRO
1100 S. Washington St.
Alexandria, VA 22314

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

Dr. Ronald A. Weitzman
NPS, Code 54Wz
Monterey, CA 92152-6800

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Douglas Wetzel
Code 12
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90007

University of Illinois/Stout

German Military Representative
ATTN: Wolfgang Wildegrube
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 GF-176
2101 Constitution Ave
Washington, DC 20418

Dr. Martin F. Wiskoff
Navy Personnel R & D Center
San Diego, CA 92152-6800

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
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
Computer-based Education
Research Laboratory
University of Illinois
Urbana, IL 61801

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

Dr. Joseph L. Young
Memory & Cognitive
Processes
National Science Foundation
Washington, DC 20550

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

END
DATED
FILM
8—88
DTIC