PSYCHOMETRIKA—VOL. 58, NO. 4, 525–543
DECEMBER 1993

MEASUREMENT INVARIANCE, FACTOR ANALYSIS AND FACTORIAL INVARIANCE

WILLIAM MEREDITH

UNIVERSITY OF CALIFORNIA AT BERKELEY

Several concepts are introduced and defined: measurement invariance, structural bias, weak measurement invariance, strong factorial invariance, and strict factorial invariance. It is shown that factorial invariance has implications for (weak) measurement invariance. Definitions of fairness in employment/admissions testing and salary equity are provided and it is argued that strict factorial invariance is required for fairness/equity to exist. Implications for item and test bias are developed and it is argued that item or test bias probably depends on the existence of latent variables that are irrelevant to the primary goal of test constructers.

Key words: measurement invariance, test bias, item bias, factor analysis, factorial invariance, selection, group differences, fairness, equity.

Introduction

The results of a factor analysis of 12 cognitive tests are presented in Table 1. The data were taken from the archives of the Institute of Human Development at the University of California at Berkeley and consist of the scores of 86 female and 71 male participants in the longitudinal studies carried out at the Institute. These subjects are quite bright, mean adolescent IQ equal 119, and well educated. Further details on IHD longitudinal studies and participant characteristics can be found in Sands, Terry and Meredith (1989). The age of the subjects at the time these particular data were collected was approximately 53 years. The variables chosen for this example were, with one exception, taken from the WAIS-R (Wechsler, 1981) and the ETS Kit (French, Ekstrom, & Price, 1963). The WAIS-R subtests are Information (INFO), Vocabulary (VOCY), Comprehension (COMP), Digit Symbol (DSBL), Block Design (BDSN) and Object Assembly (OBSL). The ETS Kit tests are Word Beginnings and Endings (BGEN), Number Comparisons (NMCP), Subtraction and Multiplication (STML), Hidden Patterns (HDPT) and Card Rotation (CROT). The twelfth test is a highly speeded Letter Series test (LSER) developed by John L. Horn (personal communication, 1981).

The analysis presented in Table 1 is based on data that, for ease of interpretation, were standardized employing the grand means and pooled variances over the two groups. The analysis uses maximum likelihood in LISREL 7 (Jöreskog & Sörbom, 1988) and the scale-free properties of maximum likelihood ensures that the standardization is of no consequence in this analysis. The factor pattern matrix, common factor dispersion matrices and common factor means are similarly standardized.

The results presented in Table 1 are quite elegant. The structure in the factor pattern matrix is based on Horn's (1985, 1986) Gf, Gc, Gs theory, although Factor 3 is more nearly Horn's Gv than Gf. The sex differences in unique means for STML and CROT are consistent with findings in the literature. The sex difference for INFO is almost surely due to the fact that these men are more highly educated than the women.

Presidential address delivered at the Annual Meeting of the Psychometric Society in Berkeley, California, June 18–20, 1993.

Requests for reprints should be addressed to William Meredith, Department of Psychology, University of California, 3210 Tolman Hall, Berkeley, CA 94720.

0033-3123/93/1200-pa93$00.75/0

Table 1

	INVARANT FACTOR PATTERN MATRIX			UNIQUE VARIANCES FEMALES	MALES	UNIQUE MEANS FEMALES	MALES
INFO	0.687	0	0	0.527	0.527	*-0.207*	*0.255*
VOCY	0.858	0	0	0.256	0.256	0	0
COMP	0.727	0	0	0.480	0.480	0	0
BGEN	0.430	0.428	0	0.574	0.574	0	0
NMCP	0	0.693	0	0.502	0.502	0	0
STML	0	0.695	0	0.517	0.517	*-0.319*	*0.387*
DSBL	0	0.530	0.282	*0.621*	*0.389*	0	0
HOPT	0	0	0.779	0.405	0.405	0	0
CROT	0	0	0.646	0.584	0.584	*-0.176*	*0.214*
BOSN	0	0	0.782	0.380	0.380	0	0
OBSL	0	0	0.622	*0.762*	*0.424*	*0.126*	*-0.151*
LSER	0.241	0.305	0.345	0.500	0.500	0	0

FACTOR DISPERSION MATRICES FEMALES			MALES			FACTOR MEANS FEMALES	MALES
0.662			1.410			0.107	-0.129
0.104	1.015		0.298	0.982		0.385	-0.469
0.556	0.286	0.917	0.812	0.575	1.101	-0.137	0.164

CHI SQUARE = 135.23 DF = 127 PROB = 0.292

The sex differences in common factor means appear to be consistent with past findings in the literature. The differences in unique variances are a puzzle, as is the sex difference for the unique means of OBSL.

Some alternative analysis yielded the following results.

1. If sex differences in unique variances are suppressed the chi square becomes 143.27 (df = 129, p = .184). The subtractive chi square is 8.04 (df = 2, p = .018). If sex differences in unique means are suppressed, forcing all sex differences to be conveyed through the common factors, and differences in unique variances are also suppressed, the chi square statistic rises to 191.05 (df = 137, p = .002). Without simple structure the chi square is 158.59 (df = 123, p = .017) and letting everything (pattern, unique variances) be free over groups except the unique means yield a chi square of 129.23 (df = 84, p = .001). Thus we may conclude that a fully satisfactory fit is obtained only when differences in unique means and variances are introduced.

2. If the two groups are combined and sex differences are ignored, an identified 3-factor solution yields a chi square of 51.67 (df = 33, p = .020). Adding simple structure constraints increases the chi square to 75.52 (df = 47, p = .005). An identified 4-factor model gives a chi square of 32.79 (df = 24, p = .109) and adding simple structure constraints raises the chi square to 51.54 (df = 38, p = .070). The fourth factor is essentially uninterpretable unless one knows that it represents sex differences.

3. If individuals' scores are represented as deviations from their same-sex mean

and the data are subsequently combined into a single sample, a 3-factor model yields a chi square of 45.26 (df = 33, p = .076). Adding simple structure constraints increases the chi square to 71.04 (df = 47, p = .013).

Thus the only approach that provides a fully satisfactory fit with theory-driven simple structure is that embodied in Table 1. Nevertheless, we propose that there may be something deeply troubling about that solution. The problem has to do with the sex differences in unique means and variances and is the topic of the rest of this paper.

We remark that in the analysis presented, the Word Beginnings and Endings score was replaced by its natural logarithm and the Card Rotation and Subtraction and Multiplication scores were replaced by their square roots. Without these transformations the fit is not as good and more differences in unique variances are found. The justification for these transforms is partly based on the plots of squared differences of two separately timed sections against their sum and is partly theory-driven (Kearns, 1971; Meredith, 1971; Silney & Meredith, 1973).

Formulation

Let X denote an n dimensional manifest random variable with realization x. Let W denote the p dimensional latent variable, with realization w, that X "measures" or, alternatively, that "underlies" X. Consider some parent population of individuals for which "measurement" of W by utilization of X is deemed appropriate. Let V denote an m dimensional random variable, with realization v, that furnishes a basis for selection of a subpopulation from the parent by application of a selection function $\mathrm{s}(V)$, $0 \leq \mathrm{s}(v) \leq 1$. We would ordinarily suppose the coordinates of V to be representations of such attributes as race or ethnicity, sex, age, etcetera, but the precise nature of the coordinates of V is immaterial and in some applications could be themselves latent variables. The selection function $\mathrm{s}(v)$ gives the conditional probability of an individual's being in the subpopulation given $V = v$.

To avoid trivialities, the following assumptions are made throughout:

1. for every realization of W, the conditional distribution of X given $W = w$ is not degenerate;
2. W and V are not independent;
3. all first and second moments, first and second order conditional moments and other expectations employed in this paper exist.

Suppose that $\mathrm{F}(\cdot)$ denotes the (cumulative) distribution function of the argument in the parent population. Then $\mathrm{dF}(u)$ denotes $\mathrm{f}(u)\mathrm{d}u$, where $\mathrm{f}(u)$ is a density function if U is continuous, $\mathrm{dF}(u)$ denotes $\mathrm{prob}(U = u)$ if U is discrete, or $\mathrm{dF}(u)$ denotes a combination thereof in situations involving combinations of discrete and continuous variables. Integrals are to be interpreted in the sense of Stieltjes. We introduce the notation $\int_U \mathrm{g}(u)\, \mathrm{dF}(u)$ to indicate integration and/or summation of the product of a function $\mathrm{g}(u)$ and $\mathrm{dF}(u)$ over all possible values of U in the sample space or a sample subspace. It will be clear from the context whether the total or a subsample space is involved. Note that the integral may denote a multiple integral. Evaluated conditional distribution functions such as $\mathrm{F}(X = x|W = w)$ will be written as $\mathrm{F}(x|w)$. We adopt the convention that vectors are rows. We define the total sample space in the parent population to be the set of $n + p + m$ dimensional real vectors (x, w, v) with the property that $\mathrm{dF}(x, w, v) > 0$. This avoids problems that may otherwise arise with conditioning. Observe that the total sample space is not necessarily the same as the Cartesian product of the sample spaces of X, W and V, but may be a proper subset thereof. We want to consider, for example, cases in which the range of W is dependent on V and/or the range of X is dependent on W. We note that according to this definition a subsample space typically will be a proper subset of the total sample space after selection on V. We note that the marginal distribution of V may become degenerate (i.e., V a constant).

Finally, $\mathscr{E}$ will denote the expectation operator and the expectation of vectors or matrices will consist of the vector or matrix of expectations of the elements.

Measurement Invariance

Definition 1. The random variable X is said to be measurement invariant with respect to selection on V if $\mathrm{F}(x|w, v) = \mathrm{F}(x|w)$ for all (x, w, v) in the sample space. This definition was introduced by Mellenburgh (1989). We shall show that this definition implies the definition employed by Meredith and Millsap (1992) and Lord's (1980) definition of lack of bias.

Theorem 1. The random variable X is measurement invariant with respect to selection on V if and only if X and V are locally independent when conditioned on w for all w in the sample space.

Proof. Suppose local independence. It follows immediately from the factorization of $\mathrm{dF}(x, w, v)$ that $\mathrm{dF}(x|w, v) = \mathrm{dF}(x|w)$. Suppose $\mathrm{dF}(x|w, v) = \mathrm{dF}(x|w)$. Then $\mathrm{dF}(x, w, v) = \mathrm{dF}(x|w, v)\, \mathrm{dF}(v|w)\, \mathrm{dF}(w) = \mathrm{dF}(x|w)\, \mathrm{dF}(v|w)\, \mathrm{dF}(w)$. $\square$

We introduce selection on V. The joint conditional distribution function of X, W and V in a selected population can be obtained by integrating

$$s\,(v)\mathrm{dF}(x, w, v)/\int_V \mathrm{s}(v)\, \mathrm{dF}(v)$$

where $\mathrm{s}(V)$ is the selection function that determines the subpopulation. The integral in the denominator is over all values of V in the induced sample space, i.e., those values of V for which $\mathrm{s}(v) > 0$. In the event that V is continuous we assume that $\mathrm{s}(V)$ is such that the denominator is > 0. Denote the altered distribution function by $\mathrm{F_s}(x, w, v)$. Note that the sample space, according to our definition, typically will be altered to a proper subset.

Theorem 2. If X is measurement invariant with respect to selection on V then $\mathrm{F_s}(x|w, v) = \mathrm{F}(x|w)$ for all selection functions and all (x, w, v) in the induced sample spaces.

Proof. Given measurement invariance $\mathrm{dF}(x, w, v)$ factors into $\mathrm{dF}(x|w)\, \mathrm{dF}(w, v)$ for every (x, w, v) in the total sample space. It follows immediately that $\mathrm{dF_s}(x, w, v) = \mathrm{dF}(x|w)\mathrm{dF_s}(w, v)$ for every (x, w, v) in the induced sample space, hence $\mathrm{dF_s}(x|w, v) = \mathrm{dF}(x|w)$. $\square$

Corollary 1. If X is measurement invariant with respect to selection on V then $\mathrm{F_s}(x|w) = \mathrm{F}(x|w)$ for all selection functions and all (x, w) in the induced sample space.

Proof. It was established in the proof of Theorem 2 that $\mathrm{dF_s}(x, w, v) = \mathrm{dF}(x|w)\, \mathrm{dF_s}(w, v)$ from which it follows that $\mathrm{dF_s}(x, w) = \mathrm{dF}(x|w)\, \mathrm{dF_s}(w)$, hence $\mathrm{dF_s}(x|w) = \mathrm{dF}(x|w)$. $\square$

Corollary 2. If X is measurement invariant with respect to selection on V, given any two selection functions $\mathrm{s}(V)$ and $\mathrm{t}(V)$, $\mathrm{F_s}(x|w) = \mathrm{F_t}(x|w)$ for all (x, w) in the intersection of the induced sample spaces. Note that the sample spaces after selection need not be identical in various special cases, such as the range of X dependent on W, the range of W dependent on V, and ordinarily occur only when the selection functions have $s(v) = 0$ and/or $t(v) = 0$ for some (different) values of V.

Proof. A direct consequence of Corollary 1. □

Theorem 2 and Corollary 1 establish that the definition of measurement invariance employed in this paper implies that $F_s(x|w)$ is invariant in subpopulations derived by selection on V and is equal to $F(x|w)$ in the parent which is the definition of measurement invariance used by Meredith and Millsap (1992). In that paper it was shown that for the case of a discrete and finite V, measurement invariance as defined therein implied local independence of X and V when conditioned on w in the parent population. Then from Theorem 1 local independence of X and V implies the definition of measurement invariance used here, namely $F(x|w, v) = F(x|w)$ in the parent population. The extension to continuous or countable V seems straightforward but will not be attempted here inasmuch as it is tangential to our main concern. Corollary 2 is a generalization of Lord's (1980) definition of lack of bias.

Consider a set of selection functions, $s_1(V), s_2(V), \ldots, s_k(V), \ldots, s_q(V)$. The question arises as to whether the invariance of $F_k(x|w)$ over k is diagnostic of measurement invariance. Generally this is not true. The following Theorem provides conditions that ensure that invariance of the $F_k(x|w)$ implies measurement invariance.

Theorem 3. Suppose:

1. V is discrete and finite, taking on values $v_1, v_2, \ldots, v_k, \ldots, v_q$;
2. q selection functions $s_1(V), \ldots, s_q(V)$ which assign $s_k(v_k) = 1$; 0 otherwise, $k = 1, \ldots, q$;
3. for each pair of selection functions $s_j(V)$ and $s_k(V)$, $F_j(x|w) = F_k(x|w)$ for every (x, w) in the intersection of the induced sample of spaces.

Then X is measurement invariant with respect to selection on V.

Proof. Let $G(x|w) = F_j(x|w) = F_k(x|w)$ for every value of (x, w) in the union of the intersections. In the event that a particular (x, w) occurs in only one of the induced spaces let $G(x|w) = F_k(x|w)$ for that (x, w). Then $G(x|w)$ is defined for every (x, w) in the total sample space. For a particular selection function V is a constant (its distribution is degenerate after selection), hence

$$dF_k(x, w, v_k) = dG(x|w)\, dF(w|v_k) dF(v_k)$$

for every v_k. It follows immediately that $F(x|w, v_k) = G(x|w) = F(x|w)$ for every (x, w, v) in the total sample space. □

Presumably Theorem 3 could be proven for a set of q linearly independent $s_k(V)$ with the property that the sum of the $s_k(V) = 1$ for every v_k since the set chosen in the theorem is a vector basis for any such set.

An obvious implication of Theorem 3 is, for example, if V is univariate and dichotomous (e.g., sex) and $F_1(x|w) = F_2(x|w)$ then X is measurement invariant with respect to selection on V. This cannot be taken to mean, however, that a further breakdown, say dividing the sexes into age groups, would yield measurement invariance for X with respect to the new V implied by the breakdown.

The failure of measurement invariance can occur when X is directly dependent on

V, when the functional form of $F(X|w, v)$ is dependent on V, or when the parameters (regression function, scedastic function) of $F(X|w, v)$ are dependent on V. We introduce the concept of *structural bias* to indicate the third form of measurement invariance failure, namely, parametric differences in $F(X|w, v)$ that depend on V. It would seem to be the case that given a particular form of $F(X|w, v)$ in which w is a Bayes sufficient statistic for x, measurement invariance automatically holds. This need not be the case inasmuch as structural bias can still occur.

Weak Measurement Invariance

The definition of measurement invariance as $F(x|w, v) = F(x|w)$ in the parent population clearly has implications that are consistent with our primitive notions of the conditions that measurement invariance should satisfy. In this section we define a weaker form of measurement invariance that does not require that the conditional distribution of X given w and v be solely dependent on w, and turns out to have consequences similar to the consequences of measurement invariance.

Definition 2. The random variable X is said to be weakly measurement invariant with respect to selection on V if $\mathscr{E}(X|w, v) = \mathscr{E}(X|w)$ and $\Sigma(X|w, v) = \mathscr{E}[(X - \mathscr{E}(X|w))'(X - \mathscr{E}(X|w))|w, v] = \Sigma(X|w)$ for all (w, v) in the sample space.

In many situations weak measurement invariance would imply measurement invariance. Suppose, for example, that the i-th component of X, X_i is conditionally binomial $[N_i, g_i(w)]$, $i = 1, \ldots, n$, where $0 < g_i(w) < 1$ is a function of W and that the components are locally independent. Then we would have weak measurement invariance and also measurement invariance. One can construct compound binomials, however, that are weakly invariant but not invariant. The difference lies principally in the fact that higher order conditional moments may be dependent on V.

One could define weak measurement invariance in terms of the conditional expectation of X alone. But consider the following. Suppose X provides a basis for choosing employees or students from an applicant pool. Clearly, differences in $\Sigma(X|w, v)$ introduces an element of unfairness into the situation. This is especially true if the conditional variances are systematically larger in one group vis-a-vis another. The implication is that individuals with the same qualifications would have different likelihoods of being chosen depending on group membership.

Lemma 1. Suppose a function of X, $g(X)$, such that $\mathscr{E}(g(X)|w, v) = \mathscr{E}(g(X)|w)$ for all (w, v) in the sample space. Suppose further a function of V, $h(V)$. Then the conditional covariance (correlation) of $g(X)$ and $h(V)$, given $W = w$, is zero for all w.

Proof. $\mathscr{E}(g(X)h(V)|w) = \int_V \int_X g(x)\, dF(x|w, v)h(v)\, dF(v|w) = \int_v \mathscr{E}(g(X)|w)h(v)\, dF(v|w) = \mathscr{E}(g(X)|w)\mathscr{E}(h(V)|w)$ □

The converse of Lemma 1 is not true. Note that Lemma 1 is true for all possible functions of V that meet the conditions of the Lemma; in particular $h(V) = V_i$ where V_i is the i-th component of V, and $h(V) = s(V)$ where $s(V)$ is a selection function.

Lemma 2. Suppose a function of X, $g(X)$, such that $\mathscr{E}(g(X)|w, v) = \mathscr{E}(G(X)|w)$ for all (w, v) in the sample space. Consider any selection function $s(V)$. Then $\mathscr{E}_s(g(X)|w, v) = \mathscr{E}(g(X)|w)$ for all (w, v) in the sample space induced by $s(V)$.

Proof. The joint distribution of (x, w, v) after selection on V is obtained from $s(v)\, dF(x, w, v)/\int_V s(v)\, dF(v)$ for all values of (x, w, v) in the induced sample space (i.e., $s(v) > 0$). Consequently $F_s(x|w, v) = F(x|w, v)$ for all values of (x, w, v) in the induced space after selection. Hence $\mathscr{E}_s(g(X)|w, v) = \mathscr{E}(g(X)|w)$. □

Let $\mathrm{Vec}(X'X)$ denote the $n(n + 1)/2$ random vector whose elements are $\mathrm{Vec}(X'X) = \{X_1^2, X_1X_2, X_2^2, X_1X_3, \ldots, X_n^2\}$.

Theorem 4. If the random variable X is weakly measurement invariant with respect to selection on V, then the components of X and $\mathrm{Vec}(X'X)$ are all conditionally uncorrelated, given $W = w$, with any function, $h(V)$, of V, for all w in the sample space.

Proof. Note that $\Sigma(X|w, v) = \mathscr{E}(X'X|w, v) - \mathscr{E}(X'|w, v)\mathscr{E}(X|w, v)$ so that given weak measurement invariance $\mathscr{E}(X'X|w, v) = \mathscr{E}(X'X|w)$. Apply Lemma 1 to the functions $g_1(X) = X_1$, $g_2(X) = X_2, \ldots, g_n(X) = X_n$, $g_{n+1}(X) = X_1^2$, $g_{n+2}(X) = X_1X_2$, etcetera in turn. □

Theorem 5. If X is weakly measurement invariant with respect to selection on V then $\mathscr{E}_s(X|w, v) = \mathscr{E}(X|w)$ and $\Sigma_s(X)|w, v) = \Sigma(X|w)$ for all selection functions and all (w, v) in the sample space induced by selection on V.

Proof. Analogous to the proof of Theorem 4 with Lemma 2 applied instead of Lemma 1. □

Corollary 3. If X is weakly measurement invariant with respect to selection on V then $\mathscr{E}_s(X|w) = \mathscr{E}(X|w)$ and $\Sigma_s(X|w) = \Sigma(X|w)$ for all selection functions and all w in the induced sample space.

Proof. Established by integrating $\mathscr{E}_s(g(X)|w, v)\, dF_s(w, v) = \mathscr{E}(g(X)|w)\, dF_s(w, v)$ with respect to V. Then let $g_1(X) = X_1$, etcetera.

It follows from Corollary 3 that for every pair of selection functions, $s(V)$ and $t(V)$, $\mathscr{E}_s(X|w) = \mathscr{E}_t(X|w)$ and $\Sigma_s(X|w) = \Sigma_t(X|w)$ if X is weakly measurement invariant with respect to V. □

Theorem 4 is the weak measurement invariance analogue of Theorem 1 although unlike Theorem 1 it is not an if-and-only-if theorem. It establishes that if X is weakly measurement invariant with respect to V, then every component of X and $\mathrm{vec}(X'X)$ is conditionally uncorrelated with every component of V when conditioned on w. A converse to Theorem 4 might be developed by imposing more restrictive conditions (e.g., monotonicity). Theorem 5 and Corollary 3 are the weak measurement analogues of Theorem 2 and Corollary 1 and establish the invariance of $\mathscr{E}_s(X|w)$ and $\Sigma_s(X|w)$ over all possible selection functions, or, put another way, that the regression of X on W and the conditional dispersion of X given w is unaffected by selection if weak measurement invariance holds.

It is customary to assume that latent variables "explain" the covariation of manifest variables (i.e., that $\Sigma(X|w)$ is diagonal for all w in the sample space. The following corollary shows that the diagonality holds in selected subpopulations.

Corollary 4. If X is weakly measurement invariant with respect to selection on V and if $\Sigma(X|w) = \Theta(w)$, diagonal, for all w, then $\Sigma_s(X|w) = \Theta(w)$ for all selection functions and all w in the corresponding induced sample space.

Proof. Follows directly from Corollary 3. □

It could be an interesting exercise to attempt to introduce Stout's (1990) concepts of essential dimensionality, perhaps dramatically reducing p as a consequence, and essential independence instead of $\Theta(w)$ diagonal into this sort of development.

We state without proof the weak measurement analogue of Theorem 3.

Corollary 5. Suppose conditions (1) and (2) of Theorem 3 hold and that for every pair of selection functions $s_j(V)$ and $s_k(V)$, $\mathscr{E}_j(X|w) = \mathscr{E}_k(X|w)$ and $\Sigma_j(X|w) = \Sigma_k(X|w)$ for every w in the intersection of the induced sample spaces. Then X is weakly measurement invariant with respect to selection on V.

Now consider that W is a random variable. Taking expectations with respect to W, standard theorems yield $\mathscr{E}(X) = \mathscr{E}[\mathscr{E}(X|W]$ and $\Sigma(X) = \Sigma[\mathscr{E}(X|W)] + \mathscr{E}[\Sigma(X|W)]$. This holds true in any population, with double expectations both occurring in the particular population. The following corollary follows immediately:

Corollary 6. If X is weakly measurement invariant with respect to selection on V then $\mathscr{E}_s(X) = \mathscr{E}_s[\mathscr{E}(X|W)]$ and $\Sigma_s(X) = \Sigma_s[\mathscr{E}(X|W)] + \mathscr{E}_s\Sigma(X|W)$ for every subpopulation generated by selection on V. Furthermore, if $\Sigma(X|w) = \Theta(w)$, diagonal, for all w in the total sample space then $\mathscr{E}_s\ \Sigma(X|W) = \mathscr{E}_s\Theta(W)$ is diagonal in every subpopulation.

Proof. Follows directly from Corollary 3. □

We remark that obviously every theorem and corollary proven for the case of weak measurement invariance holds for measurement invariance.

Factorial Invariance

Partition W into $W = (Z, U)$ where Z is of dimension $r < n$ and U has dimension n. Let Λ denote an $n \times r$ matrix of full column rank and α a vector of length n.

Definition 3. The factor analysis model holds in a population if the following conditions are met:

(i) There exist random variables Z and U such that $x = \alpha + z\Lambda' + u$ for every (x, z, u) in the sample space,
(ii) Z and U are uncorrelated in the population,
(iii) the components of U are mutually uncorrelated in the population.

The random variable U consists of latent factors specific to each individual manifest variable plus measurement error. So u is in some sense indeterminate. This infelicity will be corrected in the next section. Combining specific and error random variables makes the notation and proofs simpler in this section. The usual assumptions about errors of measurement are implicit but generally are not needed in this development. When the factor analysis model holds in a population we shall say X is factorial in the population.

In the population, let μ and Σ denote the mean vector and dispersion matrix of X, ξ and Φ the mean vector and dispersion matrix of Z, η and Ψ, diagonal, the mean vector and dispersion matrix of U. As in the discussion of measurement invariance, we suppose a selection variable V and a selection function $s(V)$ that produces a selected subpopulation from the parent. The corresponding symbols subscripted with s will denote the equivalent vectors and matrices after selection. We remark that Condition

(i) of Definition 3 implies that $x = \alpha + z\Lambda' + u$ for every (x, z, u) in the induced sample space after selection on V.

The usual factor analytic consequences follow from Definition 3;

$$\mu = \alpha + \xi\Lambda' + \eta, \tag{1}$$

and

$$\Sigma = \Lambda\Phi\Lambda' + \Psi. \tag{2}$$

Without loss of generality we may take ξ and η to be null vectors. In the sequel we will make frequent use of the fact that the dispersion matrix of a random variable is the sum of the dispersion matrix of conditional expectations with respect to another random variable plus the expectation with respect to that other variable of the conditional dispersion matrices. The selection variable will be the conditioning variable and (z, u) the variables of interest.

The fact that X is factorial in the population does not imply that Z and U will be uncorrelated, nor that the elements of U will be mutually uncorrelated in a selected subpopulation. Nor need ξ_s and η_s be null vectors. Let Γ_s denote the matrix of covariances of Z and U, and Ω_s the dispersion matrix of U, in a selected subpopulation. Corollary 7 follows immediately from (i) of Definition 3.

Corollary 7. If X is factorial in the population, then after selection on V

$$\mu_s = \alpha + \xi_s\Lambda' + \eta_s, \tag{3}$$

and

$$\Sigma_s = \Lambda\Phi_s\Lambda' + \Gamma_s\Lambda' + \Lambda\Gamma_s' + \Omega_s. \tag{4}$$

Equation (4) was derived by Bloxom (1972) as his Case Ia.

Theorem 6. Suppose that X is factorial in the population, and that for some selection variable, and for every v in the sample space, $\mathscr{E}(U|v) = \eta = 0$ and the conditional dispersion matrix of (Z, U) given $V = v$ takes the form

$$\begin{bmatrix} \Phi(v) & \Gamma'(v) \\ \Gamma(v) & \Psi(v) \end{bmatrix} \quad \text{with } \Psi(v) \text{ diagonal. Then}$$

in every selected subpopulation

$$\mu_s = \alpha + \xi_s\Lambda', \tag{5}$$

and

$$\Sigma_s = \Lambda\Phi_s\Lambda' + \Gamma_s\Lambda' + \Lambda\Gamma_s' + \Psi_s, \tag{6}$$

with Ψ_s diagonal.

Proof. Established by taking expectations of the elements of the conditional mean vector and conditional dispersion matrix with respect to $F_s(z, u|v)$ and noting that since $\mathscr{E}(U|v) = 0$, η_s and the dispersion matrix of $\mathscr{E}(U|V)$ are null, as is the covariance matrix of $\mathscr{E}(Z|V)\mathscr{E}(U|V)$, in every subpopulation. Equation (6) is essentially Bloxom's case Ib. Note that Ψ_s, diagonal, can vary over subpopulations. Also observe that somewhat weaker assumptions have been made. □

If we add to the assumptions of Theorem 6 the assumption that $\Gamma(v)$, the matrix of conditional covariances of Z and U, is null we obtain the following corollary:

Corollary 8. Given the assumptions of Theorem 6 and $\Gamma(v) = 0$ for every v in the sample space, then for every selected subpopulation equation (5) holds and

$$\Sigma_s = \Lambda\Phi_s\Lambda' + \Psi_s. \tag{7}$$

Definition 4. X is said to be strongly factorial invariant with respect to selection on V if equations (5) and (7) hold for every subpopulation derived by selection on V.

Further results can be obtained by assuming that $\Psi(v) = \Psi$.

Corollary 9. Given the assumptions of Theorem 6, $\Gamma(v) = 0$, and $\Psi(v) = \Psi$ for every v in the sample space, then for every selected subpopulation equation (5) holds and

$$\Sigma_s = \Lambda\Phi_s\Lambda' + \Psi. \tag{8}$$

Definition 5. X is said to be strictly factorial invariant with respect to selection on V if equations (5) and (8) hold for every subpopulation derived by selection on V.

Equation (8) is Bloxom's Case IIb and was originally derived by Meredith (1964). In those derivations the conditions are stronger, to wit, independence of U and V and linearity and homoscedasticity of the regression of Z on V.

Observe that our version of Bloxom's Case Ib (Corollary 7) and the definitions of strong and strict factorial invariance all require that (5) hold, i.e., that mean differences in X between selected subpopulations all be conveyed through mean differences in the common factor Z between subpopulations. *Clearly, then, the evaluation of strong or strict factorial invariance requires modeling mean vectors as well as dispersion matrices.* Modeling mean vectors in multiple group factor analysis was introduced by Sörbom (1974).

The question arises as to whether or not a selection variable can exist such that X is factorial in the population with Equations (3) and (7) or (8) holding. That is the topic of the next Theorem and Corollary.

Theorem 7. Given that X is factorial in the population, without loss of generality, $\eta = 0$, and $\mathscr{E}(U|v) \neq 0$ for some subset of V with probability measure > 0. Then $\mu_s \neq \alpha + \xi_s\Lambda'$ for some subpopulations derived by selection on V.

Proof. Suppose $\mathscr{E}(U_j|v) \neq 0$ for some component of U and subset of V. Given the conditions of the theorem, we can always determine a selection function that assigns $s(v) > 0$ to those values of V for which $\mathscr{E}(U_j|v) > 0$ and $s(v) = 0$ to those values of V for which $\mathscr{E}(U_j|v) < 0$, or vice versa. Note that such values of v must exist since $\eta = 0$. For such a selection function $\mu_s = \alpha + \xi_s + \eta_s$. □

Corollary 10. If X is factorial in the population, X is strongly factorial invariant with respect to selection on V if and only if

(i) the conditional expectation of U given v is null;
(ii) the conditional covariances of Z and U given v are all zero;
(iii) the conditional dispersion matrix of U given v is diagonal

for all v in the sample space (except a set of measure zero).

Proof. Follows directly from Theorems 6 and 7, and Corollary 8. A similar result holds for strict factorial invariance. □

Consider some selection variable such that X is *not* strongly (or strictly) factorial invariant with respect to selection on it. Then Corollary 10 establishes that (3) and (7) (or (8)) cannot hold for *every* subpopulation generated by selection on that selection variable. Corollary 10 does not rule out the possibility that (3) and (7) (or (8)) hold for some selection function. The subsequent development addresses subpopulation differences in η.

Consider, for example, the following. Suppose V is discrete and finite, taking on values v_k, $k = 1, \ldots, q$. Suppose further that $\mu_k = \alpha + \xi_k \Lambda' + \eta_k$ and $\Sigma_k = \Lambda \Phi_k \Lambda' + \Psi_k$ for $k = 1, \ldots, q$. Then any selection function such that $s(v_k) = 1$; 0 otherwise, yields (3) and (7) or (8) in the selected subpopulation. And if for some subset of k, η_k = constant, more general selection functions could yield (3) and (7) or (8) for selected subpopulations.

There is a problem here. Suppose that it has been established that $\mu_k = \alpha + \xi_k \Lambda' + \eta_k$ and $\Sigma_k = \Lambda \Phi_k \Lambda' + \Psi_k$ for $k = 1, \ldots, q$ disjoint populations with the η_k all distinct. Does this imply that a factor model with $x = \alpha + z\Lambda' + u$ holds in the population that is the union of these disjoint populations?

Theorem 8. Suppose that for q disjoint populations $\mu_k = \alpha + \xi_k \Lambda' + \eta_k$, $\Sigma_k = \Lambda \Phi_k \Lambda' + \Psi_k$, diagonal, the η_k are all distinct, $\sum \nu_k \xi_k = 0$, and $\sum \nu_k \eta_k = 0$, where ν_k is the relative proportion of members of the kth population in the union of these populations. Then a factor analytic model holds, with $x = \alpha + z\Lambda' + u$, in the union of populations if and only if $\sum \nu_k \xi'_k \eta_k = 0$ and $\sum \nu_k \eta'_k \eta_k$ is a diagonal matrix.

Proof. The conditions of the theorem imply the existence of random variables Z and U in each population such that $x = \alpha + z\Lambda' + u$ with $\mathscr{E}_k Z = \xi_k$, $\mathscr{E}_k U = \eta_k$, dispersion matrices Φ_k for Z, Ψ_k for U and $\mathscr{E}_k(Z - \xi_k)'(U - \eta_k) = 0$. (In fact, one of the problems of factor analysis, factorial indeterminacy, is that "too many such variables exist", Guttman, 1955). It follows that in the union we may write $x = \alpha + z\Lambda' + u$ and that $\mathscr{E}Z = \sum \nu_k \xi_k = 0$; similarly, $\mathscr{E}U = 0$.

It can readily be shown that $\Phi = \sum \nu_k \Phi_k + \sum \nu_k \xi'_k \xi_k \mathscr{E}(Z'U) = \sum \nu_k \mathscr{E}_k (Z - \xi_k)'(U - \eta_k) + \sum \nu_k \xi'_k \eta_k = \sum \nu_k \xi'_k \eta_k$ and $\mathscr{E}(U'U) = \sum \nu_k \Psi_k + \sum \nu_k \eta'_k \eta_k$. Then Z and U are uncorrelated in the union if and only if $\sum \nu_k \xi'_k \eta_k$ is null and since Ψ_k is diagonal, $k = 1, \ldots, q$, $\mathscr{E}(U'U)$ = a diagonal matrix if and only if $\sum \nu_k \eta'_k \eta_k$ is diagonal. □

Note that if $\eta_k = 0$, hence $\mu_k = \alpha + \xi_k \Lambda'$ for all k, Theorem 8 goes through trivially. Given the freedom of choice for ξ_k and η_k, that is, a choice of identification, these restrictions seem relatively weak. But consider the following theorem.

Theorem 9. Let A denote the $q \times r$ matrix whose k-th row is ξ_k, B denote the $q \times n$ matrix whose k-th row is η_k, N a diagonal matrix whose kk-th element is ν_k and let D denote a diagonal matrix some of whose diagonal elements may be zero. Now suppose rank (A) $< q \leq n$. Then the conditions, (i) $\sum \nu_k \xi'_k \eta_k = 0$ and, (ii), $\sum \nu_k \eta'_k \eta_k$ = D hold if and only if the number of nonzero diagonal elements of D is less than or equal to q minus the rank of A.

Proof. Condition (ii) requires that B'NB = D. If $n > q$, $n - q$ columns of B must be null for this to be true, since the rank of B is equal at most to q. Furthermore, condition (i) requires A'NB = 0, which is true if and only if the non-null columns of B lie in the orthogonal compliment of the column space of A. This, in turn, implies that the rank of the nonnull columns is less than or equal to q minus the rank of A. Condition (ii) still must hold so the maximum number of nonnull columns of B is q minus the rank of A. □

So the restrictions on A′NB and B′NB turn out to be more than trivial unless $q > n + r$ and remark that B′NB = D is never trivial. In particular we note that only a relatively small number of variables can have nonzero elements in the η_k for Theorem 9 to hold. We also note that if rank (A) = q then we must have $\eta_k = 0$ for all k, in order for the factor model represented by Λ to hold in the union of populations. Referring back to the example in the introduction, we see that the three-factor simple structure solution presented there cannot hold in the population that is the union of males and females.

Combining disjoint populations into a parent population which is their union implies the existence of a discrete, finite, selection variable which was employed to generate the subpopulation. The next corollary ensues.

Corollary 11. Suppose the conditions of Theorem 8 hold with $\eta_k = 0$ for all k. Then X is strongly factorial invariant with respect to selection on the implied V.

Proof. Follows directly from the fact that X is factorial in the population and Corollary 10. □

The following theorem is of some interest because of the light it sheds on the nature of possible selection variables.

Theorem 10. If X is strongly factorial invariant with respect to selection on V, then U and V are uncorrelated.

Proof. Strong factorial invariance implies $\mathscr{E}(U|v) = 0$ for all v, hence $\mathscr{E}(V'U) = \mathscr{E}(V'\mathscr{E}(U|V)) = 0$. □

Clearly this rules out any function of X as a candidate for a selection variable yielding strong or strict factorial invariance.

Muthén (Muthén, 1989) has addressed some of the issues discussed in this section.

Factorial and Measurement Invariance

We now separate measurement errors and specific factors, retaining U to denote the specific or unique factor random variable. The first condition of Definition 3 is altered to

(i′) $\mathscr{E}(X|z, u) = \alpha + z\Lambda' + u$ for every (z, u) in the sample space.

Since fixing an individual fixes (z, u), $\mathscr{E}(X|z, u)$ is a weak true score random variable (Lord & Novick, 1968). It follows that measurement errors are uncorrelated with true scores. We will also assume that for every *individual*, the dispersion matrix of measurement errors is diagonal (e.g., experimentally independent measures).

Denote the expected dispersion matrix of errors over individuals with same value of (z, u, v), where V is a selection variable, by $\Theta(z, u, v)$. Observe that in the development of weak measurement invariance the condition $\mathscr{E}(X|w) = \mathscr{E}(X|w, v)$ implies that $\mathscr{E}(X|w)$ is an average true score, the averaging taking place over individuals with the same value of v.

It would appear that if a factor model holds for true scores in some population and if $\Theta(z, u, v) = \Theta(z, u)$ for every (z, u, v) in the sample space, X would be weakly measurement invariant with respect to selection on V as a consequence. The ensuing discussion shows that more stringent conditions are needed.

In what follows we shall assume that V is discrete and finite and let Δ denote the dispersion matrix of the specific factors and Θ the dispersion matrix of measurement errors in the population. Then $\Psi = \Delta + \Theta$.

What are the implications of fitting factor analytic models for weak measurement invariance? Suppose that for q subpopulations defined by $v_1, v_2, \ldots, v_q$ it has been established that $\mu_k = \alpha + \xi_k\Lambda' + \eta_k$ and $\Sigma_k = \Lambda_k\Phi_k\Lambda_k + \Psi_k$ and that there exists no rotation that will yield identical factor pattern matrices over the k subpopulations. Obviously X is not weakly measurement invariant (is structurally biased) with respect to selection on V. Even if $\Lambda_k = \Lambda$, $k = 1, \ldots, q$, X cannot be weakly measurement invariant unless the η_k and ξ_k satisfy the conditions of Theorem 8, that is, that a factor model holds in the union of subpopulations.

It turns out that there is still a problem. Suppose $\mathrm{A}(V)$ is a vector function of V and $\mathrm{D}(V)$ is a diagonal matrix function of V ($\mathrm{A}(V)$ bears no obvious relation to the matrix used in Theorem 9). Suppose $\mathscr{E}(X|z, u, v) = A(v) + z\Lambda' + u\mathrm{D}(v)$. Clearly X is structurally biased in this case. Write A_k for $\mathrm{A}(v_k)$, etcetera. It follows immediately that $\mu_k = \mathrm{A}_k + \xi_k\Lambda' + \eta_k\mathrm{D}_k$ and $\Sigma_k = \Lambda\Phi_k\Lambda' + \mathrm{D}_k^2\Delta_k + \Theta_k$. Letting $\alpha = \sum\nu_k\mathrm{A}_k$, $\Psi_k = \mathrm{D}_k^2\Delta_k + \Theta_k$ and substituting η_k for $\eta_k\mathrm{D}_k + \mathrm{A}_k - \alpha$ we may write $\mu_k = \alpha + \xi_k\Lambda + \eta_k$ and $\Sigma_k = \Lambda\Phi_k\Lambda' + \Psi_k$. Obviously conditions on the A_k vector can be found that satisfy the conditions of Theorem 8. Consequently structural bias is not inconsistent with factorial invariance of the sort exemplified by Theorem 8. Furthermore, if $\eta_k = 0$ for all k we have $\mathscr{E}(X|z, u, v) = \alpha + z\Lambda' + u\mathrm{D}(v)$ and X is strongly factorial invariant with respect to V but also structurally biased with respect to V. In other words, structural bias is indistinguishable from natural group differences in unique variances and (minimal) differences in specific factor means. The foregoing discussion leads to the following Theorem.

Theorem 11. If the true score random variable, $\mathscr{E}(X|z, u)$, underlying X is strictly factorial invariant with respect to selection on some selection variable, V, and if $\Theta(z, u, v) = \Theta(z, u)$ for every (z, u, v) in the sample space, then X is weakly measurement invariant with respect to selection on V.

Proof. Follows directly from conditions required for strict factorial invariance and the definition of weak measurement invariance. □

Note that Theorem 11 is not an if-and-only theorem. Weak measurement invariance can hold without factorial invariance.

The conditions of Theorem 11 imply that in every selected subpopulation $\mu_s = \alpha + \xi_s\Lambda'$ and

$$\Sigma_s = \Lambda\Phi_s\Lambda' + \Delta + \Theta_s \tag{9}$$

(which is a special form of strong factorial invariance). Practical evaluation requires knowledge of average measurement error variances, Θ_s, and could be accomplished by employing tau equivalent measures (Lord & Novick, 1968), or beta equivalent measures (Meredith, 1971), or linearly equivalent (congeneric) measures (Jöreskog, 1971).

Nevertheless, successful fitting of (5) and (9) to a set of disjoint subpopulations does not necessarily imply that X is weakly measurement invariant with respect to selection on the implied selection variables. The second condition $\Theta(z, u, v) = \Theta(z, u)$ for all (z, u, v) in the sample space remains to be evaluated and requires either a strong modeling (e.g., strong true score theory) approach or perhaps approaches along the lines being developed by Mazzeo and Chang (1993) and Chang and Mazzeo (1993).

Now if measurement errors are assumed homogeneous, $\Theta(z, u, v) = \Theta$, and if X is strictly factorial invariant, it must be the case that $\Delta_k = \Delta$, $k = 1, \ldots, q$. We make the following assertion. *If X is strictly factorial invariant with respect to selection on V, X is almost certainly weakly measurement invariant with respect to V.* We make this assertion on the grounds that $\Delta_k + \Theta_k = \Psi$, $k = 1, \ldots, q$ is an extremely unlikely event, not as a proven assertion. It is possible for $\Theta_k = \Theta$, for all k, even if Condition (ii) of weak measurement invariance does not hold; Θ_k is an average measurement error in a group.

Item Bias

The developments in this section were stimulated by the work of Stout (1990) and Shealy and Stout (1993a, 1993b) and their remarks on bias. Also see Muthén (1989) and Muthén and Lehman (1985) for an approach similar to the one taken here. We now suppose that X is a latent variable, that for some population, we observe Y of dimension n, dichotomous, and assume conditions similar to those employed by Lord (1952) in his derivation of the normal ogive model. We state the following assumption.

1. If the i-th component, X_i, is greater than equal some constant c_i, then $Y_i = 1$; $Y_i = 0$ otherwise.
2. X is factor analytic in the population with one common factor.
3. The conditional expectation of X given z is $\alpha + \Lambda z$ and the conditional dispersion matrix is Δ, not depending on z.
4. The conditional distribution of X given z is multivariate normal.

These conditions result in normal ogive item response functions for the set of dichotomous variables, Y. Again, consider a selection variable V. We assert the following. *Given Conditions 1, 2, 3 and 4, the same normal ogive model holds in every subpopulation if and only if X is strictly factorial invariant with respect to V.* We will not prove this assertion, but claim it is fairly obvious.

What is item bias and where does it come from? First, it is clear that if the selection variable is such that X is not strictly factorial invariant, then apparent item bias will occur. This will happen when either Bloxom's case Ia or Ib holds, when group differences in means consistent with Theorem 8 hold, and when strong factorial invariance holds with natural group differences in unique variances. Thus a normal ogive model may characterize a parent population and items can appear to be biased. Note that Bloxom's Ib implies a two-factor model in subpopulations.

A second possibility is that X actually has multiple common factors with group differences in common factor means and dispersion matrices. We argue that fitting a normal ogive model to selected subgroups would be tantamount to extracting the dominant factor with each group with different pattern elements in the different subgroups yielding apparent item bias. In this case a further source of apparent bias would arise from mean differences in the "ignored" common factors. In particular, then, we note that group differences in relatively unimportant factors, both unique and common, will give rise to evidence of item bias.

Suppose we have q disjoint subpopulations and that a one-factor model holds for the latent variable X in every subpopulation with $\mu_k = \alpha + \mathscr{E}_k\Lambda'_k + \eta_k$ and $\Sigma_k = \Lambda_k\Phi_k\Lambda'_k + \Delta_k$. Then imagine that we fit a normal ogive item response function model for Y in each subpopulation and, using a common metric for the one latent variable that results, look for differences in the item response functions. Now we clearly have structural bias and would get different item response functions in different groups. Even if there were no differences in the factor pattern vectors, bias would be revealed as a consequence of additive and multiplicative structural bias of the sort introduced in the previous section. We remind the reader that natural differences in specific factor variances cannot be distinguished from multiplicative structural bias and natural differences in specific factor means may not be distinguishable from additive structural bias.

The ideas developed here clearly apply to models with guessing parameters and have obvious implications for other IRT models such as the logistic.

Fairness and Equity

Suppose again that X is manifest and that the first component of X, X_1, is a measure of job performance, academic success or "salary." Let the remaining components of X, $X_2 \ldots, X_n$, be measures of qualification such as test scores, GPA's and educational level if X_1 is job performance or academic success; or measures of merit and/or qualification if X_1 is salary. We give the following definitions of fairness and equity. It is presumed that $X_2, \ldots, X_n$ furnish the basis for choosing an applicant from a pool; or the basis for salary allocation.

Definition 6. When choosing among candidates from an applicant pool the situation is fair if every individual with same "true" value of X_1 (e.g., $\mathscr{E}(X_1|z, u)$) has the same likelihood of being chosen regardless of the applicant's sex, ethnicity, etcetera (i.e., identical conditional distributions of $(X_2, \ldots, X_n)$.

Definition 7. When X_1 is salary the situation is equitable if individuals with the same true merit (e.g., $\mathscr{E}(X_2, \ldots, X_n|z, u)$) have the same conditional distribution of salary, regardless of sex, race, etcetera.

Now, again, imagine $k = 1, \ldots, q$ disjoint populations defined by sex, race, etcetera, and that a factor analytic model holds in each group (subpopulation). We assert that if the following three conditions hold, then fairness (Definition 6), or equity (Definition 7), as the case may be, exists:

(i) X is strictly factorial invariant with respect to the implied selection variable.
(ii) In the first row of Λ only λ_{11} is nonzero
(iii) for all k, only the first element of ξ_k is nonzero.

Condition (ii) can always be satisfied trivially by rotation. Condition (iii) is not trivial and requires no group differences in the common factors that are irrelevant to X_1. Similarly, Condition (i) implies no group differences in specific factor means and variances. We note that differences in the Ψ_k, that is, strong factorial invariance, are problematic. A case can be made that $\Psi_k = \Delta + \Theta_k$ is satisfactory if X is weakly measurement invariant. This requires establishment of weak measurement invariance by external means and knowledge of the Θ_k.

Proofs of this assertion will not be provided here. For a more extensive discussion of these issues see Gregory (1991) and Millsap and Meredith (in press). In Gregory (1991) it is argued that multiple measures of job performance and/or success are vital for the full establishment of fairness and some weakening of the restriction to strict factorial measurement invariance is introduced.

Note that strict factorial invariance for the q subpopulations does not imply that the regression of X_1 on $X_2, \ldots, X_n$ or the regressions of $X_2, \ldots, X_n$ on X_1 are identical in the various subgroups. The finding of homogeneity of regression equations on the other hand is generally not consistent with strict factorial invariance, even with an additional assumption of homogeneity of variances about regression. So homogeneity of regression equations cannot imply fairness or equity as defined herein. (See Birnbaum, 1985, for a discussion of this result with respect to salary equity.)

We finally note that (LISREL) testable models occur when $n = 2$, that is, one "dependent" variable and one measure of qualification or merit.

Configural Invariance

It has been argued by Horn, McArdle and Mason (1983) that configural invariance is an important finding when modeling disjoint populations. We take configural invariance to mean that the same simple structure exists in the subpopulations in the sense that zero elements are found in the same locations in Λ_k in the several subpopulations studied and that homologous non-zero elements all have the same sign across subpopulations. We develop a scenario that could lead to such a result in practice. We suppose two "common" latent variables. Partition X into $X = \{X_1, X_2, X_3\}$, and U homologously. Suppose Z_1, Z_2, and U are independent and $\mathscr{E}(X_1|z, u) = \mathrm{H}_1(z_1) + u_1$, $\mathscr{E}(X_2|z, u) = \mathrm{H}_2(z_2) + u_2$, and $\mathscr{E}(X_3|z, u) = \mathrm{H}_3(z_1) + \mathrm{H}_4(z_2) + u_3$, where $\mathrm{H_j}(z_i)$ is a vector of ogival functions of Z_i. Further suppose these ogival functions have different "slopes" but are all "centrally located" in the sense that their inflection points are near the medians of Z_1 and Z_2 respectively.

Now imagine some selection variables independent of U, and selection functions that produce the following subpopulations. Population 1 is recruited from the high end of Z_1 and Z_2; Population 2 is recruited from the central parts of the distributions of Z_1 and Z_2; Population 3 is "high" on Z_1, and "central" on Z_2; Population 4 is "central" on Z_1 and "low" on Z_2; etcetera. Within each of these subpopulations $\mathrm{H}_i(z_i)$ could be well approximated by a linear function $\alpha_j + z_i\Lambda_k'$, $i = 1, 2, j = 1, 2, 3, k = 1, 2, \ldots$. Given the seeming robustness of factor analytic models to mild departures from linearity we might expect to find that the model $\Sigma_k = \Lambda_k\Phi_k\Lambda_k' + \Psi_k$ fits reasonably well, so well in fact that its inadequacy could not be detected in reasonably sized samples. Then matrices Λ_k could be found with configural invariance.

Clearly other scenarios using other kinds of functions and selection strategies could be developed.

Remarks

It should be obvious that measurement invariance, weak measurement invariance, strong and strict factorial invariance are idealizations. They are, however, enormously useful idealizations in their application to psychological theory building and evaluation. Their validity and existence in the real world of psychological measurement and research can never be finally established in practice. It would be of enormous utility if a factor model incorporating Tryon's (1958) notion of salient dimensionality (Stout's, 1990, essential dimensionality) and Stout's concept of essential independence could be developed and shown to have invariance properties analogous to those developed in this paper (see also McCallum & Tucker, 1991, in this regard.) Work appears to be progressing along these lines (Junker, 1991).

The concept of a parent population for which measurement with X is deemed appropriate is slippery at best. If weak measurement invariance fails we would be justified in asserting that measurement with X is inappropriate. We are often unsure as to what the selection variables actually are and whether they meet the criteria that "legitimate" selection variables must meet, namely, non-degeneracy of the joint distribution of X and V. We also note that a given X may turn out to be, for example, measurement invariant with respect to selection on one particular V but not even weakly measurement invariant with respect to selection on some other V; or strongly factorial invariant for one V but not so for another, and so forth.

Our discussion of item bias from the normal ogive model point suggests strongly that the most appropriate methods of evaluating bias and DIF are those developed, and continuing to be developed, by Stout and his coworkers (Shealy & Stout, 1993a, 1993b; Stout, 1990). In particular this body of work recognizes the multifactorial nature of "ability" and the fact that apparent item bias can be a consequence of group differences on what can be viewed as relatively unimportant latent variables.

It is the purpose of this paper to provide a conceptual framework for thinking about problems of measurement invariance (or bias), to define concepts such as measurement invariance and factorial invariance, to demonstrate conditions under which invariance can occur and to furnish tools for addressing further issues such as fairness in employment testing, salary equity, and cross-sectional developmental change. Parenthetically, we remark that most of the published work on employment and salary equity is either wrong (and harmful) or inadequate.

Let us consider what can be done in actual factor analytic practice. It would appear that, setting aside the difficulties of defining a parent population, obtaining a reasonably large representative sample from it, and evaluating the fit of a factor model, fitting (well) a factor analytic model to data obtained from a parent population would establish factorial invariance with respect to selection on almost any legitimate selection variable. But such is not the case.

Consider simultaneous factor model fitting to disjoint populations which, we insist, must involve modeling means as well as dispersion matrices. The following cases can arise.

I. Different factor pattern matrices and different means and variances of the unique (specific plus error) factors over groups.
II. Different means and variances of the unique factors over groups.
III. Strong factorial invariance, i.e., different unique factor variances over groups.
IV. Strict factorial invariance.

If Case I is the result, a factor model for the union of the disjoint populations may still be found. This presupposes that either the supermatrix formed by adjoining the different factor pattern matrices has low column dimensionality or low rank (approximately in practice) and that restrictions on the unique factors analogous to those of Theorem 8 hold. (Note that Bloxom's Case Ib is a special case of Case I and is consequently problematic, although still informative.) So fitting a factor model to the union of populations is not necessarily consistent with any sort of factorial invariance.

If Case II is the result, a factor model may be found in the union with the number of common factors exceeding the common column dimensionality of the factor pattern matrices unless the conditions of Theorem 8 hold. So a factor model for the union of populations can contain factors that are solely the result of group differences. Consider the example in the introduction when the sexes are combined, yielding a four factor solution.

If either Case III or Case IV is the result, then we can be assured that $x = \alpha + z\Lambda' + u$ holds in the union of populations. But, observe that Cases III and IV require simultaneous model fitting and cannot be presumed on the basis of a factor analysis of the union.

The results of simultaneous model fitting are clearly informative no matter which case holds. Now distinguish the scientific from the practical use of simultaneous factoring. From the point of view of the scientist all four cases are meaningful and acceptable if the results fit into a substantive theoretical framework. For Case I we would surely want configural invariance to hold. For Case II we would want simple structure, or some other form of theoretical driven structure to hold, and the mean differences of the unique factors to fall in some pattern consistent with psychological theory. We argue, however, that invariance of Λ should take primacy over simple structure in this case and in Cases III and IV as well. This means, in our view, that simply identified invariant models should be fit first and simple structure specifications introduced subsequently (see McArdle & Cattell, in press). It can be shown that underrepresentation of "primary" factors, that is, fewer than three "markers", can lead to Case II. Adding manifest variables can turn unique factors into common factors with Cases III or IV resulting.

From the point of view of the scientist either Cases III or IV are the most desirable outcomes. The fact that so many of our samples are samples of convenience adds intuitive weight to this argument.

We claim that for the practical user of tests and other measures, strict factorial invariance is essential and that outcome measures ("success," salary) should regularly be included in the analyses.

References

Birnbaum, M. H. (1985). Relationships among models of salary bias. *American Psychologist, 40*, 862–6.

Bloxom, B. (1972). Alternative approaches to factorial invariance. *Psychometrika, 37*, 425–40.

Chang, H. & Mazzeo, J. (1993, June). The identifiability of expected score in conceptualizations of differential item functioning for polytomously scored items. Paper presented at Psychometric Society Annual Meeting.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.

Gregory, K. (1991). *A reconsideration of bias in employment testing from the perspective of factorial invariance*. Unpublished doctoral dissertation, University of California at Berkeley, Berkeley, CA.

Guttman, L. (1955). The determinacy of factor score matrices with implications for five basic problems in common factor theory. *British Journal of Statistical Psychology, 8*, 65–81.

Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence*. New York: Wiley.

Horn, J. L. (1986). Intellectual ability concepts. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3). Hillsdale, NJ: Erlbaum.

Horn, J. L., McArdle, J. J., Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *The Southern Psychologist, 1*, 179–88.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109–133.

Jöreskog, K. G. & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago: SPSS.

Junker, B. W. (1991). Essential independence and likelihood based ability estimation for polytomous items. *Psychometrika, 56*, 255–78.

Kearns, R. J. (1971). *Latent class models for response emission tests*. Unpublished doctoral dissertation, University of California at Berkeley, Berkeley, CA.

Lord, F. M. (1952). A theory of test scores. New York: Psychometric Society; *Psychometric Monograph*, No. 7.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Mazzeo, J. & Chang, H. (1993, June). *Detection of differential item functioning for polytomously scored items. An adaptation of Shealy and Stout's SIBTEST procedure*. Paper presented at Psychometric Society Annual Meeting.

McArdle, J. J. & Cattell, R. B. (in press). Structural equation models of factorial invariance in parallel proportion profiles and oblique confactor problems. *Multivariate Behavioral Research*.

McCallum, R. C. & Tucker, L. R. (1991). Representing sources of error in the common factor model: Implications for theory and practice. *Psychological Bulletin*, *109*, 502–11.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, *13*, 127–43.

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, *29*, 177–85.

Meredith, W. (1971). Poisson distributions of error in Mental Test Theory. *British Journal of Mathematical and Statistical Psychology*, *24*, 49–82.

Meredith, W. & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika*, *57*, 289–311.

Millsap, R. E. & Meredith, W. (in press). Statistical evidence in salary discrimination studies: Nonparametric inferential conditions. *Applied Psychological Measurement*.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557–88.

Muthén, B. & Lehman, J. (1985). Multiple group IRT modeling: Applications to item bias analysis. *Journal of Educational Statistics*, *10*, 133–42.

Sands, L., Terry, H., & Meredith, W. (1989). Change and stability in adult intellectual functioning assessed by Wechsler item responses. *Psychology and Aging*, *2*, 79–87.

Shealy, R. & Stout, W. (1993a). An item response theory model for test bias and differential test functioning. In P. W. Holland & H. Wainer. *Differential item functioning: Theory and practice*. Hillsdale, NJ: Erlbaum.

Shealy, R. & Stout, W. (1993b). A model based standardization approach that separates true bias/DIF as well as item bias/DIF. *Psychometrika*, 159–194.

Silny, F. & Meredith, W. (1973). Test of a model for response emission tests. *Proceedings, 81st Annual Convention, American Psychological Association*.

Stout, W. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, *55*, 293–325.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology*, *28*, 229–39.

Tryon, R. C. (1958). General dimensions of individual differences: Cluster analysis vs. factor analysis. *Educational & Psychological Measurement*, *18*, 447–95.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. New York: Harcourt, Brace, Jovanovich.