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