

The measurement-statistics controversy: Factor analysis and subinterval data

LESLIE ATKINSON

Early Intervention and Developmental Evaluation Services, Toronto, Ontario, Canada

(John Gaito, *Sponsor*)

For over 40 years now the relationship between measurement scales and statistical procedures has been debated, with some theoreticians arguing that parametric statistics require interval or ratio data, as opposed to nominal or ordinal data. This debate has occurred simultaneously but independently in the univariate and factor-analytic literatures. It is argued in this paper that subinterval data are amenable to factor analysis when correlation matrices are based on product-moment coefficients (ρ , point-biserial, ϕ). This is demonstrated with data of known factor structure. Data were transformed by downgrading and mixing scale properties, and skewing and mixing distributions. Despite increasingly tenuous variable interdependence, factor structures remained highly robust across scales, distributions, and populations. The findings offer further support for the thesis that measurement scales are irrelevant in statistical analysis.

The origins of the measurement-statistics controversy are generally traced to Stevens's (1946) postulation that measurement scale determines the appropriateness of statistical treatment. Thus, nonparametric tests are applicable to data measured with nominal and ordinal scales, whereas parametric procedures require the use of interval or ratio scales. Stevens's view was recently reiterated and expanded by Townsend and Ashby (1984) and challenged by Gaito (1980, 1986; Gaito & Yokubynas, 1986).

It is interesting to note that Stevens's (1946) argument also emerged independently among factor analysts, specifically with regard to the use of dichotomous variables. Thus Carroll (1945), Cattell (1952), Ferguson (1941), Horst (1965), and Wherry and Gaylord (1944) maintained that the use of binary data would yield artifactual dimensions known as "difficulty factors." The reasoning was that as marginal splits in dichotomous data deviate from .50: .50 proportions (i.e., as p deviates from q), the range of the ϕ coefficient (-1.00 to $+1.00$ when $p=q$) becomes increasingly restricted. Therefore, the correlation between two dichotomous variables is affected not only by the actual relationship between them, but by the degree to which p deviates from q . Hence, the extent of the relationship is necessarily misrepresented.

A formal mathematical justification of this position was never offered, and its validity has been questioned (Comrey & Levonian, 1958; McDonald, 1985). By factor-analyzing dichotomous items (from the Minnesota Multiphasic Personality Inventory) with a variety of coefficients developed for binary data, Comrey and Levonian (1958) demonstrated the consistency of factor-analytic findings. Furthermore, the ϕ coefficient analysis yielded meaning-

ful factors from the standpoint of content, and reasonable communalities. Nevertheless, the "difficulty factor" allegation resulted in the virtual abandonment of ϕ in factor-analytic work (Comrey & Levonian, 1958). Moreover, the argument was generalized to all forms of subinterval data, and one frequently sees this assumption that one cannot use factor analysis with subinterval data espoused in the literature without supporting documentation or references (e.g., Roth & Ingram, 1985; Thorley, 1987).

It is true that under most circumstances continuous data analyzed with Pearson product-moment coefficients form the ideal basis for factor analysis. As the scale is downgraded, information is lost and correlations become more tenuous and vulnerable to distortion of the distribution. The purpose of this paper is to demonstrate that nominal and ordinal scales are amenable to factoring nevertheless. This is because the underlying probability distribution, and not the measurement scale, is important to the statistical analysis itself, and the normal distribution provides an excellent approximation of the exact probabilities given even by the binomial distribution (Gaito, 1980, 1986). Moreover, the ϕ coefficient may offer a good estimate of correlation even when the distribution is distorted with a split of 90%:10% (Rummel, 1970). Thus, ϕ (for binomial data), point-biserial (for one binomial and one continuous variable), and ρ (for rank-ordered data), all of which are product-moment coefficients (e.g., Harman, 1976; Nunnally, 1978; Rummel, 1970), can be used to measure correlation and provide a sound, albeit attenuated, basis for factor analysis.

METHOD

The argument that subinterval data can be validly factor analyzed is best demonstrated using data with known factor structure. Since its introduction, there has been substantial interest in the factor structure of the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler,

The author's address is Early Intervention and Developmental Evaluation Services, Surrey Place Centre, 2 Surrey Place, Toronto, Ontario M5S 2C2, Canada.

1981). A three-factor solution, consisting of verbal comprehension (VC), perceptual organization (PO), and freedom-from-distractibility dimensions, has not proven entirely consistent (Gutkin, Reynolds, & Galvin, 1984; Parker, 1983; Silverstein, 1982). However, a two-factor solution, consisting of VC and PO factors, has been extracted across the normative sample (see Table 1; Gutkin et al., 1984; Parker, 1983; Silverstein, 1982) and a large variety of clinic samples (for reviews see Hill, Reddon, & Jackson, 1985; Leckliter, Matarazzo, & Silverstein, 1986). The VC factor has its highest loadings on the WAIS-R verbal subtests (Information, Digit Span, Vocabulary, Arithmetic, Comprehension, Similarities), whereas Performance subtests (Picture Completion, Picture Arrangement, Block Design, Object Assembly, Digit Symbol) load highly on the PO factor (see Table 1, normative matrix).

As part of the attempt to validate the WAIS-R factor structure, Atkinson and Cyr (in press) factor-analyzed the WAIS-R subtests of 136 individuals with IQs between 50 and 80 (see Table 1). These data were adopted for the present study using the original interval scale as well as a number of transformations: (1) subtest scores were rank ordered; (2) half the subtest scores were dichotomized, such that $p, q \approx .5$, and half the scores were left on the interval scale; (3) all subtest scores were dichotomized, such that $p, q \approx .5$; (4) all subtests were dichotomized, such that for half the subtests $p, q \approx .5$, and for the remaining subtests $p \approx .9, q \approx .1$; (5) all subtests were dichotomized, such that $p \approx .9, q \approx .1$. Thus, scale properties were downgraded and/or mixed and, in some cases, the distributions were skewed.

In order to test the effects of these manipulations, the interrelatedness of the variables within the correlation matrices was then tested with Bartlett's (1950) test of sphericity, the Kaiser-Meyer-Olkin measure of sampling adequacy (MSA; Kaiser, 1970), and inspection of the off-diagonal elements of the anti-image covariance matrix (AIC; Kaiser, 1963). Each of these measures provides an estimate of the degree to which the components of a correlation matrix can be considered to share common variance. As scales are downgraded and distributions distorted, we would expect these measures to reflect the declining psychometric adequacy of the matrix.

Following these procedures, each matrix was analyzed, with two factors specified a priori, using two factor extraction techniques (principal components, unweighted least squares) and two types of rotation (varimax, oblique). Two extraction and rotation techniques were used to ensure robustness of solution. All factor solutions were then compared statistically (Armenakis, Field, & Wilmoth, 1977) with factor structures based on the untransformed sample data as well as the untransformed normative data (derived from Wechsler, 1981, Table 16).

RESULTS AND DISCUSSION

Table 1 shows the measures of psychometric adequacy for all correlation matrices. The normative matrix consists of highly interdependent variables, with very high χ^2 and MSA values, and with few AIC values $> .09$. Although less impressive, values for the untransformed low-IQ data still reveal that variables share much com-

mon variance. However, the adverse effect of scale and distribution transformations on variable intercorrelations is indicated by declining χ^2 and MSA values and increasing AIC values as these are applied to the transformed matrices. Although AIC values suggest that in every transformed matrix the unique variance is excessive (see Dziuban & Shirkey, 1974, for decision rules), the more liberal Bartlett's and MSA measures indicate that all matrices remain appropriate for factor analysis.

With this assurance, the matrices were factor analyzed. Table 2 shows the normative and untransformed low-IQ sample factor matrices, as well as the rotated factor matrices based on the transformed low-IQ sample data. Visual inspection of the matrices reveals the clear extraction of PO and VC factors in every case—across samples, measurement scales, measurement scale mixes, and distortions of the distribution. Although there are minor discrepancies in factor structure based on the transformed data (i.e., Digit Span and Arithmetic did not always load as highly as expected on the VC factor), it is interesting to note that these two subtests are the lowest loading of all VC subtests in both the normative and untransformed low-IQ sample data. Digit Span and Arithmetic generally load on the freedom-from-distractibility factor, when this third factor is extracted. It is because factor analysis based on subinterval data gives “adequate but not ‘best’” (McDonald, 1985, p. 201) estimates of factor structure that McDonald recommended the term “heuristic factor analysis.” In any case, when factor matrices extracted from transformed data bases are compared to those derived from the untransformed low-IQ sample data, congruence coefficients range from .95 to .99 (see Table 3), accounting for between 90% and 99% of the variance. When the factor matrices based on transformed data are compared with a second sample, represented by the normative matrix, congruence coefficients range from .88 to .96, explaining between 78% and 92% of the variance. Although not reproduced here for lack of space, results were almost identical when alternative factor extraction (unweighted least squares) and rotation (oblique) procedures were used. This is impressive evidence indeed for the use of dichotomous and ordinal data in factor analysis. As Rummel (1970) argued, “The recognition of . . . scales helps one to understand the flexibility of factor anal-

Table 1
Adequacy of Correlation Matrices

Measure	Low-IQ Matrix*						
	Normative Matrix	Untransformed	Rank Ordered	Dichotomized ($p \approx q$) and Continuous	Dichotomized ($p, q \approx .5$)	Dichotomized ($p \approx .5$ or $.9, q \approx .5$ or $.1$)	Dichotomized ($p \approx .9, q \approx .1$)
Bartlett's	$\chi^2 = 7441$ $p < .00001$	$\chi^2 = 603$ $p < .00001$	$\chi^2 = 201$ $p < .00005$	$\chi^2 = 545$ $p < .00005$	$\chi^2 = 402$ $p < .00005$	$\chi^2 = 306$ $p < .00005$	$\chi^2 = 309$ $p < .00005$
MSA	.94	.86	.61	.84	.82	.79	.73
AIC (% $> .09$)	9.1	25.5	41.8	29.1	29.1	40.0	45.5

Note—Bartlett's = Bartlett's test of sphericity; MSA = measure of sampling adequacy; AIC = inspection of the off-diagonal elements of the anti-image covariance matrix. *These data were adapted from Atkinson and Cyr's (1987) factor analysis of the Wechsler Adult Intelligence Scale-Revised subtest scores of 136 individuals with IQs between 50 and 80.

Table 2
Factor Structure of the WAIS-R (Principal Components, Varimax Rotation)

Subtest	Low-IQ Matrix													
	Normative Matrix		Untransformed		Rank Ordered		Dichotomized ($p \approx q$) and Continuous		Dichotomized ($p, q \approx .5$)		Dichotomized ($p \approx .5$ or $.9, q \approx .5$ or $.1$)		Dichotomized ($p \approx .9, q \approx .1$)	
	PO	VC	PO	VC	PO	VC	PO	VC	PO	VC	PO	VC	PO	VC
Information	.25	.84	.22	.77	-.06	.67	.28	.69	.16	.67	.01	.81	.01	.79
Digit Span	.30	.59	.39	.40	.13	-.07	.41	.35	.38	.24	.19	.49	.00	.26
Vocabulary	.27	.87	-.01	.85	-.32	.72	.04	.84	.12	.80	.11	.67	.07	.68
Arithmetic	.40	.67	.55	.36	.45	.07	.68	.15	.74	.04	.42	.21	.29	.35
Comprehension	.26	.81	.42	.64	.25	.57	.48	.58	.54	.38	.41	.50	.52	.44
Similarities	.34	.76	.23	.61	-.03	.53	.18	.72	.17	.72	.32	.50	.19	.65
Picture Completion	.64	.43	.47	.52	.16	.55	.56	.37	.58	.36	.62	.33	.56	.45
Picture Arrangement	.49	.48	.72	.28	.59	.24	.70	.19	.68	.21	.72	.05	.70	.15
Block Design	.79	.34	.85	.20	.78	-.03	.81	.22	.69	.13	.75	.14	.77	.02
Object Assembly	.87	.14	.81	.04	.66	-.11	.76	.07	.68	.04	.61	.16	.67	-.10
Digit Symbol	.49	.43	.74	.23	.52	.06	.68	.23	.66	.23	.62	.15	.63	.13
Eigenvalue	.96	6.09	4.86	1.34	2.09	1.95	4.66	1.27	4.04	1.24	3.59	1.17	3.34	1.46
% Variance	25.8	38.4	30.89	27.7	19.0	17.7	31.6	22.4	29.3	18.7	24.7	18.8	24.0	19.5

Note—WAIS-R = Wechsler Adult Intelligence Scale-Revised; PO = perceptual organization factor; VC = verbal comprehension factor.

ysis ... and to apply it within substantive domains that have yet to develop interval or ratio scales'' (p. 223). In this regard, Rummel argued that sometimes subinterval data may be superior to interval data for factor analysis. For instance, in the case where measurement is unreliable, it is better to rank order observations or dichotomize them rather than enter substantial error variance into the factor analysis. Excessive error variance may result in spurious but seemingly interpretable factors (Armstrong & Soelberg, 1968; Horn, 1967). Or again, in the case where two distributions are of vastly different shapes, the Pearson product-moment coefficient underestimates the degree of association between the variable involved. In this case, converting the data to rank order form may actually prove more powerful than factor analysis of the interval data. This is not to say that the Pearson product-moment coefficient with continuous data is not generally the best (i.e., most powerful and most accurate) coeffi-

cient to use. However, this is not the only correlation coefficient upon which factor analytic work can or should be based.

It is important to note that all factor analyses in the present paper were based on product-moment correlations—phi, rho, point-biserial, or Pearson. Other forms of correlation have proven less reliable. Phi-over-phi-max, a coefficient specifically designed to overcome the "difficulty factor" problem, tends to overestimate the degree of correlation and produce inflated factor loadings (Comrey & Levonian, 1958). Biserial and tetrachoric coefficients also exaggerate the degree of association. In addition, they are mathematically impoverished (Nunnally, 1978), and the latter have very stringent normality assumptions (Christoffersson, 1975; Rummel, 1970).

In sum, I argue that subinterval data are appropriate for factor analysis. This was demonstrated with a number of measurement scales, scale mixes, and distortions

Table 3
Factor Solution Congruence Coefficients

	Low-IQ Matrix											
	Untransformed		Rank Ordered		Dichotomized ($p \approx q$) and Continuous		Dichotomized ($p, q \approx .5$)		Dichotomized ($p \approx .5$ or $.9, q \approx .5$ or $.1$)		Dichotomized ($p \approx .9, q \approx .1$)	
	PO	VC	PO	VC	PO	VC	PO	VC	PO	VC	PO	VC
Untransformed Low-IQ Matrix												
PO	—		.95		.99		.98		.96		.95	
VC		—	.95		.98		.95		.97		.98	
Normative Matrix												
PO	.96		.92		.95		.94		.95		.93	
VC		.98	.88		.95		.91		.94		.94	

Note—PO = perceptual organization factor; VC = verbal comprehension factor.

of the underlying distributions. The present results offer further support for the thesis that measurement scales are irrelevant in statistical analysis.

REFERENCES

- ARMENAKIS, A. A., FIELD, H. S., & WILMOTH, J. N. (1977). An algorithm for assessing factor structure congruence. *Educational & Psychological Measurement*, **37**, 213-214.
- ARMSTRONG, J. S., & SOELBERG, P. (1968). On the interpretation of factor analysis. *Psychological Bulletin*, **70**, 361-364.
- ATKINSON, L., & CYR, J. J. (in press). Factor structure of the WAIS-R in a sample of individuals with low I.Q. *American Journal of Mental Retardation*.
- BARTLETT, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, **3**, 77-85.
- CARROLL, J. B. (1945). The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika*, **10**, 1-19.
- CATTELL, R. B. (1952). *Factor analysis: An introduction and manual for the psychologist and social scientist*. New York: Harper & Row.
- CHRISTOFFERSSON, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, **40**, 5-32.
- COMREY, A. L., & LEVONIAN, E. (1958). A comparison of three point coefficients in factor analysis of MMPI items. *Educational & Psychological Measurement*, **18**, 739-755.
- DZIUBAN, C. D., & SHIRKEY, E. S. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, **81**, 358-361.
- FERGUSON, G. A. (1941). The factorial interpretation of test difficulty. *Psychometrika*, **6**, 323-329.
- GAITO, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin*, **87**, 564-567.
- GAITO, J. (1986). Some issues in the measurement-statistics controversy. *Canadian Psychology*, **27**, 63-68.
- GAITO, J., & YOKUBYNAS, R. (1986). An empirical basis for the statement that measurement scale properties (and meaning) are irrelevant in statistical analyses. *Bulletin of the Psychonomic Society*, **24**, 449-450.
- GUTKIN, T. B., REYNOLDS, C. R., & GALVIN, G. A. (1984). Factor analysis of the Wechsler Adult Intelligence Scale-Revised (WAIS-R): An examination of the standardization sample. *Journal of School Psychology*, **22**, 83-93.
- HARMAN, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.
- HILL, T. D., REDDON, J. R., & JACKSON, G. N. (1985). The factor structure of the Wechsler scales: A brief review. *Clinical Psychology Review*, **5**, 287-306.
- HORN, J. L. (1967). On subjectivity in factor analysis. *Educational & Psychological Measurement*, **27**, 535-538.
- HORST, P. (1965). *Factor analysis of data matrices*. New York: Holt, Rinehart & Winston.
- KAISER, H. F. (1963). Image analysis. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 156-166). Madison, WI: University of Wisconsin Press.
- KAISER, H. F. (1970). A second generation little jiffy. *Psychometrika*, **35**, 401-416.
- LECKLITER, I. N., MATARAZZO, J. D., & SILVERSTEIN, A. B. (1986). A literature review of factor analytic studies of the WAIS-R. *Journal of Clinical Psychology*, **42**, 332-342.
- MCDONALD, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum.
- NUNNALLY, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- PARKER, K. (1983). Factor analysis of the WAIS-R at nine age levels between 16 and 74 years. *Journal of Consulting & Clinical Psychology*, **51**, 302-308.
- ROTH, D. L., & INGRAM, R. E. (1985). Factors in the Self-Deception Questionnaire: Associations with depression. *Journal of Personality & Social Psychology*, **48**, 243-251.
- RUMMEL, R. J. (1970). *Applied factor analysis*. Evanston, IL: Northwestern University Press.
- SILVERSTEIN, A. B. (1982). Factor structure of the Wechsler Adult Intelligence Scale-Revised. *Journal of Consulting & Clinical Psychology*, **50**, 661-664.
- STEVENS, S. S. (1946). On the theory of scales of measurement. *Science*, **103**, 677-680.
- THORLEY, G. (1987). Factor study of a psychiatric child rating scale based on ratings made by clinicians on child and adolescent clinic attenders. *British Journal of Psychiatry*, **150**, 49-59.
- TOWNSEND, J. T., & ASHBY, F. G. (1984). Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin*, **96**, 394-401.
- WECHSLER, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised*. New York: The Psychological Corporation.
- WHERRY, R. J., & GAYLORD, R. H. (1944). Factor pattern of test items and test as a function of the correlation coefficient: Content, difficulty, and constant error factors. *Psychometrika*, **9**, 237-244.

(Manuscript received for publication December 28, 1987.)