Multivariate Experimental Clinical Research, Volume 12, No. 1, 2000, 29-41.

A COMPARISON OF THREE APPROACHES TO CONSTRUCTING ITEM PARCELS TO IMPROVE SUBJECT-TO-PARAMETER RATIOS IN CONFIRMATORY FACTOR ANALYSIS

John R. Schallow
University of Manitoba
Winnipeg, Manitoba, Canada

ABSTRACT

A confirmatory factor analysis (CFA) of the 100 items of the Revised Eysenck Personality Questionnaire (EPQ-R) on a sample of 205 South African refugees failed to satisfactorily fit the EPQ-R's intended 4-factor structure. The lack of fit is possibly due to the low subject-to-parameter (n:k) ratio and the dichotomous item responses. Three procedures were examined for combining items into parcels to alleviate these problems. Within each of the Psychoticism, Extraversion, Neuroticism, and Lie scales, typically three items were parceled by (a) judged content, (b) highest intercorrelation, and (c) randomly. CFAs of the three methods for constructing item parcels showed equivalent, if modest, fits of the 4-factor model to the data. The Psychoticism factor was problematic, but CFAs eliminating Psychoticism indicated that the other three factors appeared to reasonably measure their intended constructs. Arguments related to simplicity and the a priori explication of meaning are used to advocate the application of content parcels.

INTRODUCTION

Researchers often cannot achieve the sample sizes of rare, remote or dear subjects to adequately analyze — according to available wisdom — the number of measures that can be readily gathered on these subjects. With exploratory factor analysis (EFA), for example, published recommendations for the ratio of observations to variables (*n:k*) have ranged from unity (Alcamoni, 1976) to in excess of 10:1 (Kunce, Cook, & Miller, 1975; Nunnally, 1978), but many texts addressing EFA sample size (e.g., Gorsuch, 1983; Tabachnick & Fidell, 1989) seem to concur on a minimal 5:1 ratio of subjects to measures.[1] Nonetheless, clinical, social, or personality researchers are often in situations where even this recommended minimum 5:1 ratio is unattainable. An example of this situation is our attempt to examine the appropriateness of the suggested four factor scoring of the 100-item revised Eysenck Personality Questionnaire (EPQ-R; Eysenck, Eysenck, & Barrett, 1985) on a sample of 205 South African refugees in Tanzania before the recent political changes in South Africa.

The main developmental work on the 4-factor (Neuroticism, Extraversion, Psychoticism, and Lie scales) structure of the EPQ-R was conducted on British samples (Eysenck et al., 1985). These South African refugees, examined in a study of the possible effects of political detention, torture, and the experience of being a refugee (Mpumlawana, 1991), seemed to represent a patently different population than that on which the EPQ-R was developed. After several months in Tanzania, data from 205 South African refugees were obtained.

As the goal was to test the ability of the previously derived 4-factor structure of the EPQ-R to "fit" the data from the South African refugee sample, confirmatory factor analysis (CFA) rather than EFA seemed appropriate (Bernstein & Teng, 1989; Briggs & Cheek, 1986; Mulaik, 1987; Watkins, 1989). As with EFA, there are no agreed-upon guidelines for determining minimum sample sizes for CFA, although there is a growing literature on this issue (e.g., Anderson & Gerbing, 1984; Balderjahn, 1985; Bearden, Sharma, & Teel, 1982; Boomsma, 1982; Gallini & Mandeville, 1984; Gerbing & Anderson, 1985; Geweke & Singleton, 1980; Huba & Bentler, 1983; Muth,n & Kaplan, 1985, 1992; Tanaka, 1987). Tanaka (1987) states that "there is some agreement that sample-size appropriateness is tied to the ratio of the number of subjects to the number of parameters estimated [which] differs somewhat from the usual concern with the ratio of numbers of subjects to number of variables" (p. 137). A guideline that Bentler (1993) acknowledges as "over-simplified" is that the ratio of sample size to the number of freely estimated parameters may be able to go as low as a minimum of 5:1.

Even by these minimal standards, a CFA of the 100-item EPQ-R on 205 South African refugees looked questionable. This article examines three approaches to improve the *n:k* ratio which involve combining several items into "parcels" or "miniscales," approaches which root back to the early days of factor analysis. Cattell introduced parceled item factoring into questionnaire research back in 1956 (a & b) because of a lack of confidence in the stability of single items and also because of the incapability of early computers to handle large matrices. Some researchers grouped items into parcels based upon apparent content (e.g., Comrey, 1970), but Cattell and Burdsal (1975) felt that this "...could be rejected as being too subjective and as depending on possibly half-unconscious and almost certainly insufficiently informed stereotypes of a particular experimenter" (p. 167). Instead, Cattell and Burdsal (1975) favored empirical approaches that parcel on the basis of items having a minimum mutual correlation of say, 0.3, although they advocated a more sophisticated extension of this that they called radial parceling. Factoring parcels rather than items generated some controversy (Cattell & Burdsal, 1975; Comrey & Duffy, 1968; Eysenck & Eysenck, 1969; Howarth & Browne, 1971).

An additional problem questionnaire item responses pose for factor analysis, besides their stability, is that rating scales are not normally distributed (Drasgow & Kanfer, 1985) and, of course, this is particularly true of dichotomous responses. Bernstein and Teng (1989) suggest most researchers are unaware or disregard the seriousness of the problems posed by categorical data (Bock & Aitken, 1981; Muth,n, 1978, 1984, 1987; Takane & de Leeuw, 1987). Essentially, parceling has also been recommended as a solution for these considerations (Bernstein & Teng, 1989; Drasgow & Kanfer, 1985; Skinner, 1984). Cook, Dorans and Eignor (1988)

caution, in fact, that the radial parceling procedure recommended by Cattell and Burdsal (1975) involves all the problems inherent in the factor analysis of item data.

Notwithstanding Cattell and Burdsal's (1975) negative opinion of experimenters making judgments about apparent content and meaning in parceling items, much would seem to commend this simple procedure. By whatever process formed, any researchers' parcels are public, replicable, and testable — at least some of the "stuff" of science. Attempts to articulate in advance what a small parcel of items means seems eminently consistent with the theory driven nature of CFA. Moreover, interpretation of what a collection of items means cannot be sidestepped; factor analysis may indicate that a group of items belong together but the latent meaning of that group of items remains the fallible act of hypothetical construction. As Mulaik (1987) has put it, "inductive procedures do not automatically give us meanings" (p. 301). In any event, almost any manner of grouping items together that are all supposedly suffused by the same common factor may not, in the final analysis, make much difference. In this article, I examine these issues by comparing CFA results produced by three different methods for constructing item parcels: judged content, high mutual intercorrelations, and random assignment. Specifically, I compare CFA results produced by these three item-parceling methods to evaluate the fit of the 4-factor EPQ-R structure on the sample of the 205 South African refugees.

METHOD

SUBJECTS

Two hundred ten South African refugees residing in two small communities in Tanzania sponsored by the African National Congress — Mazimbu and Dakawa — completed an extensive battery of questionnaires intended to assess the psychological effects of political detention, torture under detention, and being a refugee. Although versions of the battery were available in Xhosa and Zulu (which had been independently back translated to assure their accuracy of meaning), the vast majority of respondents chose to complete the English language version. Completed batteries were obtained from 205 exiles (148 males and 57 females) ranging in age from 13 to 47 years old (M = 25 years). The average length of time spent in exile was 3.5 years.

THE REVISED EYSENCK PERSONALITY QUESTIONNAIRE

In response to criticisms of the Psychoticism (P) scale of the original Eysenck Personality Questionnaire (EPQ: Eysenck & Eysenck, 1975), Eysenck, Eysenck, and Barrett (1985) attempted to improve the P scale by adding new items. The new P scale has 32 items, seven more than the original EPQ P scale. Very slight changes were made to the Extraversion (E) and Neuroticism (N) scales and the Lie (L) scale was unchanged from the original.

Eysenck et al. (1985) reported satisfactory reliabilities (Cronbach's alpha) for all four scales ($.78 \leq \alpha \leq .90$). Hypothesized as measuring orthogonal factors, the P, E, and N scales appeared only very modestly correlated ($-.07 \leq r \leq .23$). As would be anticipated, the L scale was somewhat negatively correlated with the personality scales ($-.16 \leq r \leq -.34$). In addition to improved reliability, the

distribution of the revised P scale was closer to normal and had a higher mean score than the original scale.

PARCEL CONSTRUCTION

Content parcels. Based upon a preliminary examination of the items and the goal of achieving a $n{:}k > 5$, it was decided to generally combine items into triads and, to keep the parcels roughly the same size, to allow no parcels to be smaller than two items or greater than four items. Within each of the four identified scales of the EPQ-R, items were combined into parcels on the perceived similarities of the contents of items with the vast majority of parcels (28/33) being triads. The items and the labels given these content parcels are shown in Table 1, along with some descriptive statistics. While content parcels were constructed with no attention to item intercorrelations, all but one of the content parcels had mean inter-item correlations that exceeded those of the items generally in the scales they were representing.

TABLE 1
DESCRIPTION OF CONTENT PARCELS WITH STATISTICS DESCRIBING SKEWNESS (Sk), KURTOSIS (Kur), MEAN INTER-ITEM CORRELATION (M_r), AND CRONBACH'S ALPHA (α)

Scale/Content Parcel	Items	*Sk*	*Kur*	M_r	α
Psychoticism					
Disobedient	18, 29, 50, 88	0.58	-0.57	.15	.46
Distrusting	34, 85, 95	0.34	-0.76	.23	.47
Intimidating	30, 73, 91	1.47	2.14	.22	.46
Callous	12, 37, 99	1.36	1.84	.22	.46
Rejecting Family	42, 68, 96	0.31	-1.12	.08	.20
Impulsive	2, 81	0.74	-0.51	.09	.15
Unconventional	48, 59, 64	0.87	-0.26	.10	.26
Financial Issues	7, 9, 75	0.10	-0.62	.05	.13
Miscellaneous	25, 56	2.32	4.77	.23	.37
Disregards Manners	14, 21, 41	2.75	0.78	.39	.66
Inconsiderate	5, 54, 79	1.80	3.23	.23	.47
Extraversion					
Outgoing	28, 36, 40, 55	0.21	-1.02	.21	.52
Lively	11, 16, 94	-0.85	-0.67	.44	.70

Table 1 continued

Scale/Content Parcel	Items	*Sk*	*Kur*	M_r	α
Sociable	20, 33, 45, 58	-0.96	0.14	.27	.60
Activity Oriented	1, 72, 90	-0.30	-1.25	.31	.57
Life of the Party	24, 51, 78	-0.08	-1.32	.34	.61
Talkative	6, 47, 63	-0.04	-1.30	.25	.50
Quick Deciding	61, 67, 69	-0.61	-0.86	.31	.57
Neuroticism					
Ennui	70, 76, 84	-0.08	-1.38	.38	.64
Moody	3, 8, 92	-0.15	-1.36	.35	.61
Irritable	17, 26, 97	-0.33	-0.94	.20	.42
Sensitive	22, 87, 100	-0.12	-1.22	.22	.46
Worrying	38, 43, 52, 74	-0.54	-0.29	.17	.56
Nervous	35, 46, 83	0.70	-1.08	.52	.66
Sleepless/Tired	60, 65	0.00	-1.22	.24	.36
Guilt/Shame	13, 31, 80	-0.54	-1.49	.24	.49
Lie					
Obedient Child	39, 53, 57	-0.03	-1.53	.45	.71
Denies Flaws I	10, 49, 93	0.24	-0.80	.19	.41
Denies Flaws II	77, 82, 89	0.02	-1.13	.12	.28
Denies Flaws III	27, 44, 66	-0.64	-0.37	.26	.51
Virtuous Habits	23, 32, 62	-0.12	-1.12	.32	.59
Responsible	4, 19, 71	0.30	-0.88	.19	.41
Moral Rectitude	15, 86, 98	-0.46	-0.46	.22	.46

Correlated parcels. Another set of parcels was constructed by successively selecting triads of items within the suggested EPQ-R scales that were the most highly intercorrelated of the remaining items. In the case of Psychoticism and Extraversion (32 and 23 items, respectively) the last remaining items formed dyads rather than triads. Although 13 of the correlated parcels partially overlapped (sharing two items) with the conceptual parcels, only two correlated parcels were identical to content parcels: Nervous (Neuroticism Scale) and Moral Rectitude (Lie Scale).

Random parcels. The last set of parcels was constructed by randomly selecting triads of items from within the four suggested EPQ-R scales. Again in the cases of Psychoticism and Extraversion the last remaining items to be parceled were dyads rather than triads. Five random parcels partially overlapped (sharing two

items) with content parcels; one additional random parcel was identical to a content parcel: Ennui (Neuroticism). Eight random parcels partially overlapped with correlated parcels.

TABLE 2
EPQ-R SCALE STATISTICS FOR SCALES BASED ON RAW ITEMS, CONTENT PARCELS, CORRELATED PARCELS, AND RANDOM PARCELS

		Parceling Method		
EPQ-R Scale	Raw Items	Content	Correlated	Random
Psychoticism (M = 8.1, SD = 3.7)				
N	32	11	11	11
M_r	.06	.12	.11	.15
α	.63	.60	.56	.67
Extraversion (M = 12.9, SD = 4.1)				
N	23	7	8	8
M_r	.10	.20	.19	.25
α	.72	.67	.65	.73
Neuroticism (M = 13.3, SD = 4.6)				
N	24	8	8	8
M_r	.13	.28	.24	.30
α	.78	.75	.73	.78
Lie (M = 11.9, SD = 3.9)				
N	21	7	7	7
M_r	.11	.23	.22	.26
α	.72	.68	.67	.71

Note: M_r refers to the mean inter-item correlation for raw items and mean inter-parcel correlation for the parcels constructed according to each of the three methods.

RESULTS

SCALE STATISTICS

EPQ-R scale statistics on the raw items, content parcels, correlated parcels, and random parcels are displayed in Table 2. While the mean inter-item correlations were somewhat higher among the parcels than among the raw items, internal consistency reliability estimates for the each type of parcel were quite similar to those produced by the individual items.

ITEM CFA

All CFAs were analyzed using Jöreskog and Sörbom's (1984) LISREL VI program (version 6.6). Items were given free loadings to be estimated on their respective factors and fixed zero loadings on all other factors. The latent factors were all given unit variances and were allowed to be correlated. Because of the dichotomous nature of responses, the item data was analyzed by ordinary least squares estimation of a tetrachoric correlation matrix (cf. Jöreskog & Sörbom, 1984). While two-thirds of items loaded > .30 on the factors they were supposed to be measuring, the overall goodness of fit was quite unsatisfactory (GFI = .53, AGFI = .51, RMSR = .14). Of course, the *n:k* ratio was very unsatisfactory (*n:k* < 2:1) for this analysis..

PARCEL CFAs

The three sets of parcels were also analyzed with model described above for the items. With estimates of parcel loadings on factors and six intercorrelations between factors with unitized variances, there were at most 40 parameters to be freely estimated from data on 205 subjects thereby achieving Bentler's recommended 5:1 minimum ratio.

Parcels made assumptions of normality more tenable. For items, the mean absolute value for skewness was 0.95 (70% < *Sk* = 1) and for kurtosis was 2.22 (11% < *Kur* = 1). These were substantially reduced for parcels; for example, the mean absolute value for skewness was 0.60 (85% < *Sk* = 1) and for kurtosis was 1.37 (42% < *Kur* = 1) for content parcels. Nonetheless, as a precaution parcel data were first analyzed with ordinary least squares estimations of polychoric correlation matrices (cf. Jöreskog and Sörbom, 1984) followed by maximum likelihood solutions for covariance matrices. Inasmuch as the results are nearly identical, only the maximum likelihood solutions will be reported here.

For all factors except P almost all parcels, however constructed, meaningfully load (> .30) on the scales they were meant to measure. For P, however, a minority of any type of parcels loaded meaningfully. Only 4/11 Psychoticism content parcels had loads greater than .30. Correlated parcels had only 5/11 with loading greater than .30, while random parcels had only 3/11 parcels with loadings greater than .30. The orthogonality of the P, E, and N factors seemed tenable only for E relative to both P and N; P and N were significantly ($p < .002$) correlated for all three parcel techniques' solutions. While the L factor had significant correlations negative with P, E, and N factors in each parcel approach ($p < .05$), it was particularly associated with P, their shared variance approaching 50%. The overall

goodness-of-fit indices, displayed in Table 3, showed all three parcel approaches to provide equivalent modest overall fits of the EPQ-R factor models to the data.

TABLE 3
GOODNESS-OF-FIT INDICES FOR CONFIRMATORY NON-ORTHOGONAL FACTOR ANALYSES OF THE EPQ-R BASED ON CONTENT, CORRELATED, AND RANDOM PARCELS

Parceling Method	χ^2	*df*	χ^2/df	GFI	AGFI	RMSR
Content	735.7	489	1.50	0.82	0.79	0.08
Correlated	731.3	521	1.40	0.82	0.80	0.08
Random	725.6	521	1.39	0.83	0.80	0.07

PARCEL CFAs OMITTING PSYCHOTICISM

In light of the weakness of the P factor revealed by all previous analyses, CFAs on all three parcels were repeated omitting any specification and all measures of Psychoticism. Overall goodness-of-fit indices improved (GFIs = .88-.89; AGFIs = .86; RMSRs = .05-.07) for these three factor solutions. Average parcel loadings on their respective factors were .51 (.47-.56). While pursuing changes indicated by significant modification indices did result in solutions with GFIs greater than .90, these changes simply resulted in modest secondary loadings for parcels on additional factors.

DISCUSSION

With a very low subject-to-parameter ratio (*n:k* < 2:1) and the gross non-normality of dichotomous data, a CFA of EPQ-R item responses gathered from a sample of South African refugees could not be adequately fitted to the four-factor model suggested by Eysenck et al. (1985). Parceling typically triads of items within the scales they were intended to measure by three quite different procedures — judged content similarity, highest intercorrelation, and random assignment — resulted in more satisfactory *n:k* ratios and more normal data distributions. As Table 2 presented, all three item-parcel construction techniques produced similar internal consistency statistics to the those produce by the EPQ-R scales' original items. While parcels were clearly superior to items in producing more satisfactory fits of CFAs to the Eysenck four-factor model, little discrimi-

nated statistically between the three procedures examined for parceling items. The factor loadings of all three parceling procedures raised some doubt about the adequacy of the P scale with this sample. The average loadings of all parcel approaches on P were only about .25, less than what Comrey (1973) characterizes as poor loadings. In contrast, the average loadings of parcels onto the other factors (.42 .51) are what Comrey (1973) describes as fair loadings. All three parceling techniques produced better fits in CFAs after eliminating the Psychoticism factor and its related parcels.

Based solely on statistical criteria, therefore, there seems no basis for preferring one parceling technique over the others examined in this study. In some sense, this finding parallels that of Hase and Goldberg (1967) which found four strategies, including essentially those used in creating the three parcel sets in this study, for creating scales using CPI items were equivalent on 13 diverse criteria. Nonetheless, a number of considerations lead this author to advocate forming parcels on a more rational basis.

SIMPLICITY

Content parcels, relative at least to correlated parcels, were quick and easy to form. It took me only a few hours to examine the EPQ-R items and to identify parcels of items within scales that had similar meanings or contents. On the other hand, examining correlation matrices, some with nearly 500 entries, and identifying triads with the highest mean intercorrelations by calculating average Z scores was time consuming, laborious, and boring. Unfortunately, this author knew of no existing programs for such a task and blanched at the prospects of writing one. It is true that randomly aggregating items from particular scales into triads was even easier and faster, but that seemed too much like acting as though I had no ideas whatsoever about shared meanings and potential relationships between items.

THE ARTICULATION OF MEANING

An apparent advantage of content parcels over correlated and random parcels was the assistance they gave in determining what the proposed scales for the EPQ-R were, and were not, measuring. Content parcels seemed best at identifying what still seem to be problems, at least in this South African refugee sample, with the Psychoticism scale. The content parcels, however, also seemed to corroborate that the E, N, and L scales were indeed measuring what they were intended to measure.

The content parcels seemed to be useful in understanding whatever was being weakly measured by the Psychoticism factor. In advance of any statistical analyses, content parcels defined in this author's view as disobedient, distrusting, intimidating, and callous best characterized this tenuous factor. The meaning of such attributes would need to be couched in an appreciation that all of these South African refugees had fled Apartheid into exile, a majority reported having been tortured in detention, and many had loved ones killed in raids by the South African army. Also, if one were inclined to cast these attributes into a psychopathological light, they would seem more definitional of psychopathy than psychoticism.

The content parcels seemed to confirm the intended interpretations of the more satisfactory factors. Content parcels a priori labeled outgoing, lively and sociable seem most defining of a factor intended to measure extraversion. Content parcels predefined as ennui, moodiness, irritability, and sensitivity load highest on Neuroticism, but that worry, nervousness, and guilt/shame load lower might put a more depressive than anxious cast on the neuroticism being measured. Finally, the L factor identified in this study seems to relate more strongly to denying minor flaws and less to the affirmation of desirable attributes. L scores therefore seem most related to what Paulhus (1984) would identify as conscious dissembling or impression management rather than to the more unconscious process of self-deception or ego enhancement.

ARE CONTENT PARCELS TOO SUBJECTIVE?

Recall that Cattell and Burdsal (1975) rejected parceling items on content as too subjective, essentially in their process of formation by any particular researcher. Undoubtedly, some readers might disagree with my content parcels and/or the labels that I gave to those parcels and could provide their own alternatives. Had space permitted, I might have provided more explicit rationales for the content parcels presented in Table 1; these would have represented some combination of common language analysis and psychological theory. Another researcher's rationale for grouping items differently might even be persuasive to me. I would argue, however, that the degree of subjectivity about where ideas come from is as irrelevant to the outcome as the chemist Kekul reporting that his idea for the benzene ring came from his dream of snakes dancing in a ring with each others' tails in their mouths. What seems more important to me is that, once formed by whatever basis, parcel sets are public and explicit; we can empirically test whether any set does a better job. Just as the MMPI-2 has a number of different scalings — the clinical, content, and supplemental scalings — it might also be that different parcels have different, but equivalently good (or bad) stories to tell about our data.

GENERAL APPLICATION

While the example that I have presented only deals with CFA, there seem to be a number of other multivariate analysis situations where content parceling might recommend itself as a potential solution to the *n:k* ratio or categorical variable problems. Even in very exploratory analyses, rarely would researchers be totally uninformed (nor should they be) about what variables might be meaningfully combined to make an exploratory factor, regression, or canonical analysis more manageable. Psychological variables are rarely created de novo; they typically have histories of being researched or at least of researchers/theorists contemplating their meaning and how they relate to other variables. As Harvey Skinner (1984) argued, "The single most powerful data analytic technique is a 'rational mind.' "

REFERENCES

Aleamoni, L. M. (1976). The relation of sample size to the number of variables in using factor analysis techniques. *Educational and Psychological Measurement*, *36*, 879-883.

Anderson, J. C. & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 15-173.

Balderjahn, I. (1985). The robustness of LISREL unweighted least squares estimation against small sample size in confirmatory factor analysis models. In W. Gaul & M. Schader (Eds.), *Classification as a tool of research* (pp. 3-10). Amsterdam: Elsevier Science Publishers.

Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample size effects on chi-square and other statistics used in evaluating causal models. *Journal of Marketing Research*, *19*, 425-430.

Bentler, P. (1993). *EQS structural equation program manual* [Program manual]. Los Angeles: BMDP Statistical Software.

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin*, *105*, 467-477.

Bock, R. D., & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika*, *46*, 443-459.

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation* (Part I, pp. 149-174). Amsterdam: North-Holland.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of personality*, *54*(1), 106-148.

Cattell, R. B. (1956). A shortened "basic English" version (Form C) of the 16 PF Questionnaire. *Journal of Social Psychology*, *44*, 257-278. (a)

Cattell, R. B. (1956). Validation and intensification of the Sixteen Personality Factor Questionnaire. *Journal of Clinical Psychology*, *12*, 205-214. (b)

Cattell, R. B., & Burdsal, C. A., Jr. (1975). The radial parcel double factoring design: A solution to the item-vs-parcel controversy. *Multivariate Behavioral Research*, *9*, 165-179.

Comrey, A. L. (1970). *A manual for the Comrey Personality Scales*. San Diego: Education and Industrial Testing Service.

Comrey, A. L. (1973). *A first course in factor analysis*. New York: Academic Press.

Comrey, A. L. & Duffy, K. E. (1968). Cattell and Eysenck factor scores related to Comrey personality factors. *Multivariate Behavioral Research*, *4*, 379-392.

Cook, L. L., Dorans, N. J., & Eignor, D. R. (1988). An assessment of the dimensionality of three SAT-Verbal Test editions. *Journal of Educational Statistics*, *12*, 19-43.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in diverse populations. *Journal of Applied Psychology*, *70*, 662-680.

Eysenck, H. J., & Eysenck, S. B. G. (1969). *Personality structure and temperament.* San Diego: R. R. Knapp.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire.* London: Hodder & Stoughton.

Eysenck, H. J., Eysenck, S., and Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences, 6*(1), 21-29.

Gallini, J. K., & Mandeville, G. K. (1984). An investigation of the effect of sample size and specification error on the fit of structural equation models. *Journal of Experimental Education, 53*, 9-19.

Gerbing, D. W., & Anderson, J. C. (1985). The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. *Mulitivariate Behavioral Research, 20*, 255-271.

Geweke, J. F., & Singleton, K. J. (1980). Interpreting the likelihood ratio statistic in factor models when the sample size is small. *Journal of the American Statistical Association, 75*, 133-137.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Hase, H. D., & Goldberg, L. R. (1967). Comparative validity of different strategies of constructing personality inventory scales. *Psychological Bulletin, 67*, 231-248.

Howarth, E. & Browne, J. A. An item analysis of the 16 PF Personality Questionnaire. *Personality: An International Journal, 2*, 117-139.

Huba, G. J., & Bentler, P. M. (1983). Test of a drug use causal model using asymptomtotically distribution free methods. *Journal of Drug Education, 13*, 3-14.

Jöreskog, K., & Sörbom, D. (1984). *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods* [Program manual]. Mooresville, IN: Scientific Software, Inc.

Kunce, J. T., Cook, D. W., & Miller, D. E. (1975). Random variables and correlational overkill. *Educational and Psychological Measurement, 35*, 529-534.

Loehlin, J. C. (1987). *Latent variable models: An introduction to factor, path, and structural analysis.* Hillsdale, NJ: Erlbaum.

Mpumlwana, V. B. (1991). PTSD symptoms in South African exiles (Doctoral dissertation, University of Manitoba, 1991). *Dissertation Abstracts International, 52A*, 2308-2309.

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research, 22*, 267-305.

Muthèn, B. (1978). Contributions of factor analysis to dichotomous variables. *Psychometrika, 43*, 551-5

Muthèn, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115-132.

Muthèn, B. (1987). *LISCOMP: Software for advanced analysis of linear structural equation models with a comprehensive measurement model* [Program manual]. Mooresville, IN: Scientific Software.

Muthèn, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171-189.

Muthèn, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables; A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, *45*, 19-30.

Nunnaly, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598-609.

Skinner, H. A. (1984). Correlational methods in clinical research. In A. S. Bellack & M. Hersen, *Research method in clinical psychology*, pp. 139-156. N.Y.: Pergamon.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics* (2nd ed.). New York: HarperCollins.

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393-408.

Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, *58*, 134-146.

Watkins, D. (1989). The role of confirmatory factor analysis in cross-cultural research. *International Journal of Psychology*, *24*, 685-701.

Author Notes

Appreciation is expressed to Vuyo Mpumlwana for the use of her dissertation data in these analyses. Correspondence concerning this article should be sent to Dr. John R. Schallow, Department of Psychology, University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2.

Footnotes

1. We should note that an empirical literature exists that calls into question the utility of any of these suggested guidelines (Arindell & van der Ende, 1985; Barrett & Kline, 1981).