Psychological Reports, 1963, 12, 247-250.

EMPIRICAL NOTE ON THE "NUMBER OF FACTORS" PROBLEM IN FACTOR ANALYSIS

KENNETH I. HOWARD AND ROBERT A. GORDON
University of Chicago

This paper illustrates a practical method for determining the number of factors to be extracted from a matrix of sample correlations. Although a number of criteria have been proposed (cf. Cattell, 1952; Harman, 1960), many are cumbersome, while those that are not often lead to contradictory decisions. Tryon (1960) has shown that in the use of tests suggested by Coombs (1941) and Tucker (Thurstone, 1938), the extraction of an additional factor after the criterion indicates that factoring should end often leads to a decision to extract still further factors. That is, in some cases the criterion for stopping is reached after extracting four factors, but if a fifth factor is extracted the test may indicate further extraction. It is not likely that statistical tests, such as Lawley's (1940), which cause the number of factors to be a function of sample size (and significance level) will satisfy all factor analysts, e.g., if the same battery of tests is administered to samples of different sizes, Lawley's test may indicate different numbers of factors in each case, complicating the problem of factorial invariance.

The general practice has been to rely upon rules of thumb. Among these is the recommendation that principal axis factoring end at the point when eigenvalues fall below 1.00 (Harman, 1960, p. 363). Another common practice is to cease factoring when there is a sharp break in the amount of variance reduction between two adjacent factors. Recently, Wrigley (1960) has offered a practical method for limiting factor extraction to common factors. This paper presents an illustration and extension of Wrigley's method.

Wrigley (1960) has proposed overfactoring (drawing more than the assumed number of factors) and then using Varimax (Kaiser, 1958) or Quartimax (Neuhaus & Wrigley, 1954) to rotate successive numbers of factors. In this procedure one searches among the rotated factors of a solution for one having the highest loading of only one variable. If such a specific factor is found, the last factor from the principal axis solution is omitted and Varimax (or Quartimax) is reapplied. The procedure is repeated until that solution is found in which each factor contains the highest loadings of at least two (experimentally independent) variables. This method serves to limit factoring to the extraction of common factors. However, there may still be certain ambiguities associated with some of these factors. The following example illustrates this and suggests a way of handling such ambiguities.

As part of a larger study[1] behavioral data were obtained on the activities

[1]Youth Studies Program, University of Chicago, NIMH Grant M-3301. All computations were done on UNIVAC I.

TABLE 1
SALIENT LOADINGS FOR EACH FACTOR IN EACH ROTATED SOLUTION

Variable	I^6	I^6	I^7	II^5	II^6	II^7	III^5	III^6	III^7
1. Individ. Fight	79	78	80						
2. Group Fight	77	76	77						
3. Concealed Weapon	68	66	68						
4. Assault	67	63	68						
5. Individ. Sports				71	71	72			
6. Team Sports				68	66	67			
7. Social Acts				60	57	57			
8. Gambling				49	50	50			
9. Sex. Intercourse							78	79	80
10. Statutory Rape	40	39	36				68	68	70
11. Petting							67	68	67
12. Signifying							52	54	55
13. Work Experience							35	33	*30*
14. Narcotics									
15. Marihuana									
16. Homosexuality									
17. Illegitimate Child									
18. Commonlaw Marriage									
19. Suicide									
20. Auto Theft									
21. Drive w/o license									
22. Public Nuisance	44	41	49	35	42	41			
23. Theft	44	38	35	*32*	38	36			
24. Gang Bang									
25. Bribery									
26. Domestic Chores	*31*	35	*33*	36	*34*	*34*			
27. Alcohol	36	*34*	49				39	38	*33*
28. Robbery	51	45	48						
29. Forgery	*24*	*27*	35						
30. Hanging				40	43	44	43	45	43
31. Joy Riding				46	51	51			
32. Truancy				43	48	47			
33. Runaway									
34. Arson									
35. Homicide									
36. Pimping									
37. Rape									

of 598 street-corner gang boys. These data were combined into 37 variables, the variables were intercorrelated, and the resulting matrix (with communality estimates in the diagonal) was factored using the principal axis method. Eleven factors were extracted. Varimax rotations were performed using the first 2, 3, 4, 5, and so on, up to 11 of the principal axis factors.

Table 1 shows the salient loadings (.35 and above) of each factor for the rotational solutions using five, six, and seven factors, which suffice for this illustration. As can be seen, Factor VII in the last solution was a "specific factor" in that it contained only one salient loading. Every subsequent rotated factor was also a specific. On the basis of this evidence, according to Wrigley's criterion, six factors would have been regarded as "common" and retained for interpretation. However, comparison of the loadings of Factor VI in the six-

TABLE 1 (cont'd)
SALIENT LOADINGS FOR EACH FACTOR IN EACH ROTATED SOLUTION

IV^5	IV^6	IV^7	V^5	V^6	V^7	VI^6	VI^7	VII^7	$(h^2)^5$	$(h^2)^6$	$(h^2)^7$
									72	72	73
									73	74	74
									57	57	57
									61	62	62
									56	58	58
									47	48	49
									48	49	49
									35	35	35
									71	71	73
									66	66	68
									55	55	55
									42	45	45
									18	19	21
56	39	*34*				51	57		42	44	46
56	37	*33*				56	53		47	53	54
53	36	*34*				48	48		37	39	39
49	52	52							29	32	32
48	54	55							24	30	31
36	37	37							14	16	16
			69	58	59	39	*31*		58	58	59
			65	62	62				50	53	53
			57	45	42				65	66	66
			52	*29*	38	52	*22*	59	62	65	81
			31	35	*33*				29	34	34
			29	37	39				12	16	19
									27	28	28
			48	45	*33*	*29*	43		59	60	75
36	*23*	*24*				45	*28*		51	52	53
			30	42	42				16	27	28
									41	43	44
			40	*30*	*25*				39	39	42
			38	*26*	*26*				41	41	41
			43	*24*	*25*	42	*30*		28	32	32
									03	05	05
									03	05	06
									11	12	12
									10	13	13

Note.—Arabic superscripts to Roman numerals indicate number of factors rotated. Roman numerals identify factors. Except for loadings in italics, which are provided for comparison, only loadings of .35 or above are listed. Decimal points are omitted.

and seven-factor solutions called into question its status as a common factor. Of the eight salient loadings on this factor in the six-factor solution five fell below .35 with the inclusion of the seventh factor, a specific, which was now defined by what was previously the second highest loading variable (Theft) on Factor VI. The transient appearance of Theft (on its way to becoming a specific) on this factor, coupled with its general instability, suggested that in the six-factor solution Factor VI represented a kind of residual common variance factor, which splits apart with the addition of another reference vector. Interpretation of such a factor, composed of temporary residual or accidental elements, appeared unwarranted. Furthermore, in the context of this research

and in comparison with the meaningfulness of the first five factors, the salient variables of Factor VI made little psychological or sociological sense. [For interpretation of these factors, see Short, Tennyson, and Howard (1963).] These considerations led us to postulate only five common factors.

We feel that this example illustrates a practical solution to the number of factors problem, when a high speed computer is available. One overfactors, then rotations, using Varimax, are carried out on an increasing number of factors until one recognizable as a specific appears. At this point an evaluation is made of the stability of the loadings of the remaining common factors. In the case where the stability of a common factor is in doubt one may decide to exclude or include that factor on the basis of its interpretability. An alternative method would be to examine rotational solutions using more specific factors in order to assess the relative tendencies toward decomposition of any common factors suspected as unstable. In this way interpretation is restricted to stable common factors.

Summary.—In the application of factor analytic procedures to a matrix of sample correlations it is necessary to adopt some criterion for ceasing to factor. Although there are a variety of such criteria available, many lead to contradictory and unstable decisions. Wrigley (1960) has proposed a procedure which insures the retention of the maximum number of common factors. This paper illustrates a practical modification of Wrigley's method in which extraction and interpretation can be restricted to the maximum number of *stable* common factors.

REFERENCES

CATTELL, R. B. *Factor analysis.* New York: Harper, 1952.

COOMBS, C. H. A criterion for the significance of common factor variance. *Psychometrika,* 1941, 6, 267-272.

HARMAN, H. H. *Modern factor analysis.* Chicago: Univer. of Chicago Press, 1960.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika,* 1958, 23, 187-200.

LAWLEY, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edin.,* 1940, 60, 64-82.

NEUHAUS, J. O., & WRIGLEY, C. The quartimax method: an analytical approach to orthogonal simple structure. *Brit. J. stat. Psychol.,* 1954, 7, 81-91.

SHORT, J. F., TENNYSON, R. A., & HOWARD, K. I. Behavior factors in gang delinquency. *Proc. Amer. Stat. Assn,* 1962, 16, 61-67.

THURSTONE, L. L. *Primary mental abilities.* Chicago: Univer. of Chicago Press, 1938.

TRYON, R. C. Salient dimensionality vs. the fallacy of 'minimal rank.' Paper presented at the first annual meeting of the Society of Multivariate Experimental Psychology, 1960.

WRIGLEY, C. A procedure for objective factor analysis. Paper presented at the first annual meeting of the Society of Multivariate Experimental Psychology, 1960.

Accepted January 5, 1963.

ERRATUM

HOWARD, K. I., & GORDON, R. A. Empirical note on the "number of factors" problem in factor analysis. *Psychological Reports*, 1963, 12, 247-250.

Add to Table 1, p. 249, for the variable, 4. Assault, Factor VI^6, .36; Factor VI^7, .23.

Erratum published in:
Psychological Reports, 1963, 12, 299E.