

Published in *Psychological Bulletin*, 70 (5), 1968, 361-364.

On the Interpretation of Factor Analysis

J. Scott Armstrong

Sloan School, MIT, now at
The Wharton School, University of Pennsylvania

Peer Soelberg

University of Wisconsin, Milwaukee

Abstract

The importance of the researcher's interpretation of factor analysis is illustrated by means of an example. The results from this example appear to be meaningful and easily interpreted. The example omits any measure of reliability or validity. If a measure of reliability had been included, it would have indicated the worthlessness of the results. A survey of 46 recent papers from 6 journals supported the claim that the example is typical, two-thirds of the papers provide no measure of reliability. In fact, some papers did not even provide sufficient information to allow for replication. To improve the current situation some measure of factor reliability should accompany applied studies that utilize factor analysis. Three operational approaches are suggested for obtaining measures of factor reliability: use of split samples, Monte Carlo simulation, and a priori models.

With the advent of computers, factor analysis has become a relatively inexpensive technique. As a result, the number of published studies employing this method of analysis is rapidly increasing. Solomon (1960) reported that over 1,000 papers on factor analysis were published between 1900 and 1960 and that the rate of publication has increased steadily. These publications have been divided between papers on methods and papers on applied problems which utilize factor analysis as the primary statistical technique (Guilford, 1961). This paper is concerned with the latter category. Do the applied studies lead to advances in the description or understanding of real world situations?

To provide a framework for discussion, a simple example of applied work was selected. The data for this study were the analogue of asking 50 employees to rate their supervisors on 20 traits. These "responses" were then factor analyzed to determine whether underlying factors might be found that could summarize the results in a meaningful and economical fashion.

The factor analysis was carried out to conform to common practice. Pearson product moment correlations between all trait variables were used. Since the study was exploratory in nature, a principal components analysis was carried out using unity in the major diagonal. Following the rule of Kaiser (1960) and Guttman (1954), the solution was obtained by rotating all factors with eigenvalues greater than one.¹

Results

The results are presented in Table 1. Only loadings greater than 0.5 are shown.

¹ The calculations were done with the University of California, Los Angeles Biomedical 03M program (Dixon, 1965) at the Massachusetts Institute of Technology Computation Center.

Table 1. Nine Factors from Principle Components Solution

Variable	Loading	Variable	Loading
I. Facism		VI. Kindness	
Sensitive	-.72	Kind	.82
Democratic	-.72		
Responsible	.69		
II. Sincerity		VII. Humor	
Sincere	-.85	Humorous	-.88
		Aggressive	-.54
III. Social		VIII. Tactfulness	
Distance	.84	Tactful	.81
Formal	-.65		
Fair	-.54		
Humble			
IV. Reliability		IX. Social	
	.80	Leadership	.72
Trustworthy	.55	Warm	.71
Humble		Active	.64
		Enthusiastic	.55
		Ambitious	
V. Docility			
Patient	.69		
Shallow	.59		
Aggressive	-.53		

The results from this study seemed promising. It was possible to summarize 71% of the variance in the 20 traits with nine factors. We found also that it was rather easy to name each of the factors. In addition, it was easy to recall various studies in the literature which lent support to the solution. In short, it seemed rather simple to demonstrate internal validity for this study. A detailed analysis is not presented here, however, for reasons which will become obvious below.

Discussion

There is one key element which is missing from this “typical study.” The report provides no objective benchmark by which the results may be evaluated. If reliability had been evaluated, the results from the example would obviously have been discarded. The reason is that the “employee responses” were merely random normal deviates (selected from Rand, 1955). The arbitrary trait names were assigned prior to carrying out the analysis and only one solution was obtained.

Since it appears to be rather simple for a researcher to make sense out of the patterns provided by factor analysis, some benchmark or measure of factor reliability should be made a requirement for publication. Unfortunately, statistical tests for measuring factor reliability do not appear to be well developed. (For a recent discussion, see Cliff and Hamburger, 1967.) There are, however, some simple and operational methods for assessing reliability which will be summarized below.

Is the example used in this paper a fair basis for criticizing current research? To answer this question, a survey was carried out to assess current practice. Papers from six academic journals were examined to see whether the authors provided some benchmark for comparison—either a measure of reliability or a measure of validity.

This literature survey was selective: Factor-analytic reports excluded from the survey were those studies that tested the reliability of some previously reported, factor-derived scaling of items (like the MMPI scale), or replicated a previously reported set of factors. Most of the decisions concerning studies analyzed were reasonably

clear-cut. No evaluation was made of how successful the author was at establishing the reliability of his factors or the validity of his results, but merely of whether or not he had made an attempt to provide such information.

Table 2. Summary of Literature Survey

Journal	Years	N of studies	N of studies without a reliability measure	N of studies without measure of validity	N of studies with neither reliability nor validity measures
<i>Educational and Psychological Measurement</i>	1964, 1965, 1966	11	4	6	4
<i>Journal of Applied Psychology</i>	1963, 1965	15	11	10	8
<i>Journal of Consulting Psychology</i>	1964, 1965, 1966	7	4	2	2
<i>Journal of Marketing Research</i>	1964, 1965, 1966	3	2	2	1
<i>Journal of Personality</i>	1964, 1966	1	1	1	1
<i>Personnel Psychology</i>	1962, 1963, 1965	9	8	6	5
Totals		46	30	27	21

Table 2 indicates that about two-thirds of the studies did not report a measure of reliability and 46% reported on neither reliability nor validity. One wonders how many of the delinquent studies would have been published had the reliability of the factors been measured and reported.

Measurement of Reliability

Three rather simple approaches to measuring factor reliability are suggested: split samples, a priori analysis, and Monte Carlo simulation.

Split Samples

The original data might be split into two (or more) random subsamples. Separate factor analyses can then be run for each subset, and a comparison made of them. Although more sophisticated methods are available (Lawley & Maxwell, 1963), a simple approach is to correlate the factor loadings as suggested, for example, by Burt (1948). It is recommended that the results from each subsample be published separately, so that “eye-ball” comparisons may be made. In the studies that were surveyed in Table 2, the sample sizes were generally large enough relative to the number of variables that it would have been easy for the authors to have split their samples.

A Priori Analysis

Before analyzing the data, the researcher should work out, in as much detail as possible, the structure of the solution that he expects to find. He might, for example, postulate the number of factors he expects to appear, which variables should load together, relationships which should exist among factors, or what variables he expects will dominate which factors. His predictions could be based on behavioral models, previous findings reported in the literature, or merely on “well-educated” hunches. Few of the studies that we examined made any attempt to develop an a priori model.

The results may, of course, agree with an a priori model as the result of luck or chance. The key point, however, is that the a priori specification of a model provides a much more objective benchmark than is provided by a posteriori appeals. Horn (1967) discussed problems associated with this approach. The key problem relates to the subjectivity required in deciding upon factor-analytic procedures. Reference to this problem is found below under “Other Problems of Interpretation.”

Monte Carlo Simulation

In some cases sample sizes are so small that it is not practical to split the sample. In addition, one may have very weak prior information about the underlying behavioral processes. For such cases researchers should try to simulate their results by factor analyzing suitable samples of random data. By “suitable samples” is meant sets of random data chosen to conform to the actual data in terms of sample size, number of variables, and assumed underlying distributions. The type of factor analysis should also conform to that used on the actual data.

The analysis of random data could be replicated many times in order to obtain distributions of the various factor statistics. By comparing the results based on actual data with the distributions from Monte Carlo simulations one could judge whether the former appeared to be “significantly” different from the latter. For example, comparisons could be made in terms of the number of factors having eigenvalues greater than 1.0, the average loading on most important variables in each factor, or the percentage of variance explained by a given number of factors.

If the actual results yielded fewer factors, had consistently higher loadings, and had a higher percentage of variance explained than the simulation data, the investigator would gain support for the claim that his factor analysis provided a good way of summarizing the data. On the other hand, if actual results did not differ substantially from simulated results, one would have reason to question the reported factor analysis. Several of the papers included in this survey reported results which seemed no more impressive than the outcome of our random number experiment.

Each of the previous approaches has its own advantages and disadvantages. In some situations, however, it may be useful to try more than one approach for measuring reliability.

Measurement of Validity

In addition to the measures of reliability mentioned above, it would be desirable if a researcher provided some measure of the validity of his factor analysis. Authors often claim that their study has been useful in an exploratory sense. But how is such a statement to be interpreted? If the study merely provides the author with an intuitive understanding, then it is not always clear what the next step should be. If there are other payoffs, say for prediction or control, then perhaps the author should attempt to demonstrate them.

In order to judge the validity of a study it may be helpful to specify at least one dependent variable which the factor analysis was designed to help explain or predict. But in many of the reports that we read it appeared as if factor analysis had been used when the investigator saw no other way of “massaging” his data.

Other Problems of Interpretation

The example in this paper implied that the researcher followed some well-accepted conventions in carrying out the factor analysis. In fact, there is a lack of well-accepted conventions. The researcher must make many decisions in order to reduce his problem to a manageable size. Such decisions include: (a) how should the variables be measured? (b) how many variables should be included in the analysis? (c) how large should the sample size be? (d) what type of correlation coefficient should be used? (e) what estimates should be used for communalities? (f) how should the factors be extracted? (g) how many factors should be extracted? (h) what type of rotation should be performed?

The subjectivity imposed by the analyst in making the above decisions compounds the problem of interpretation. The factor solutions are affected significantly by these decisions.

The survey of the literature also pointed up that some papers did not provide enough information about the decisions made by the analyst. As a result, it is not always possible for other researchers to replicate the study.

Conclusion

Intercorrelations of distributions of random normal deviates were factored by a standard method and a seemingly good structure obtained. Although trait names were assigned at random, factors appeared to be readily interpretable. This random data example is typical of many published studies in that no measure of reliability or validity was presented. To guard against the publication of meaningless results, it is suggested that all published studies include a measure of factor reliability. Where possible, some attempt should also be made to demonstrate the validity of the factors.

References

- Burr, C. (1948), "The factorial study of temperamental traits," *British Journal of Psychology, Statistical Section*, 1, 178-203.
- Cliff, N. & Hamburger, C. D. (1967), "The study of sampling errors in factor analysis by means of artificial experiments," *Psychological Bulletin*, 68, 430-445.
- Dixon, W. J. (1965), *BMD: Biomedical Computer Programs*. University of California, Los Angeles, School of Medicine, Department of Preventive Medicine and Public Health.
- Guilford, J. P. (1961), "Psychological measurement a hundred and twenty-five years later," *Psychometrika*, 26, 101-127.
- Guttman, L. (1954), "Some necessary conditions for common-factor analysis," *Psychometrika*, 19, 149-161.
- Horn, J. L. (1967) "On subjectivity in factor analysis," *Educational and Psychological Measurement*, 27, 811-820.
- Kaiser, H. F. (1960), "The application of electronic computers to factor analysis," *Educational and Psychological Measurement*, 20, 141-151.
- Lawlay, D. N. & Maxwell, A. E. (1963), *Factor Analysis as a Statistical Method*. London: Butterworth.
- Rand Corporation (1955), *A Million Random Digits with 100,000 Normal Deviates*. New York: Free Press.
- Solomon, H. (1960), "A survey of mathematical models in factor analysis," in H. Solomon (Ed.), *Mathematical Thinking in the Measurement of Behavior*. New York: Free Press.