

Other titles by Paul Kline available from ROUTLEDGE

- Fact and Fantasy in Freudian Theory 1981
- Psychology and Freudian Theory: Introduction 1984
- Handbook of Test Construction: Introduction to
Psychometric Design 1986
- Psychology Exposed 1988
- Intelligence: The Psychometric View 1990
- The Handbook of Psychological Testing 1992
- Personality: The Psychometric View 1993

An easy guide to factor analysis

Paul Kline

1994



2.5 Computing correlations from raw scores	20
3.1 Correlation matrix	30
3.2 The first principal component	33
3.3 The residual matrix	34
3.4 The reflected residual matrix	34
6.1 Observed correlations and path coefficients	88
7.1 Direct Oblimin rotated factor structure matrix	106
7.2 The four factor solution of the shortened type A scale	110
7.3 Correlations between factor scores and psychometric variables	111
7.4 Factor pattern matrix of PPT scales and other variables	117
8.1 Structure matrix of a five factor Direct Oblimin rotation of a principal components solution	131
8.2 Structure matrix of a three factor Direct Oblimin rotation of a principal components solution for PPQ and EPQ scale scores	134
8.3 Number and percentage of items with their highest loading on keyed scale	136
8.4 Mean absolute factor loading of scale items	136
9.1 Key to variables	142
9.2 Direct Oblimin rotation of factors	143
9.3 Offender subgroup factors in analysis 1	148
9.4 Offender subgroup factors in analysis 2	149
9.5 Offender subgroup factors in analysis 3	150
10.1 Fit indices of first-order models	161

Chapter 1

A general description of factor analysis

AIMS OF THE BOOK

Factor analysis is a statistical technique widely used in psychology and the social sciences. Indeed in some branches of psychology, especially those in which tests or questionnaires have been administered, it is a necessity. With the advent of powerful computers and the dreaded statistical packages which go with them factor analysis and other multivariate methods are available to those who have never been trained to understand them. I have examined countless theses in which the factor analytic results were simply a source of fear and confusion to the students who were praying that they would not be asked to explain them.

Furthermore, I am regularly approached by colleagues not only in my own university but in other departments of psychology and education, especially, for advice on interpreting and carrying out factor analyses. None of these colleagues and students is stupid. There are several reasons why they have failed to understand factor analysis. First, in Great Britain, at least among social scientists there is a general fear of mathematics. Equations freeze the mind, rendering it blank. Second, in many departments of psychology and education factor analysis is not well taught, or not taught at all, as I have found in teaching postgraduates. As in my time as a post-graduate, students are referred to a number of 'excellent books, which tell you all you need to know'.

Regrettably, it is precisely these books which are the source of the problem. First let it be said that there are a number of excellent books on factor analysis and later in this *Easy Guide* I shall refer to them. However, except to mathematicians and those who have a reasonable knowledge and insight into factor analysis, they are

unreadable and close to incomprehensible. Indeed only one book, to this writer's knowledge, has ever attempted to simplify factor analysis to a level which students might be expected to understand - Dennis Child's *The Essentials of Factor Analysis* (latest edition 1990). Even this book many students find too hard and some important aspects of the technique are not included. Why these books are so difficult stems from the fact that they are usually written by good mathematicians. These find mathematical arguments and the equivalence of apparently completely different formulae so obvious as to require no explanation.

The aim of the *Easy Guide* is to provide an explication of the basic mathematics of factor analysis which anybody who can manage any form of tertiary education can follow, so that, at the end of the book, readers will understand factor analysis. All mathematical terms, even the most simple, will be explained. All mathematical processes will be illustrated by examples. Readers will not be expected to make mathematical inferences. Each step will be explained. If the basic mathematics of factor analysis is understood, readers will then be able to use the technique effectively for research and, perhaps more importantly, they will be able to evaluate its use in journal papers. For much factor analytic research, as has been shown by Cattell (1978), is technically inadequate, rendering the results valueless. In addition I believe that, after reading this book, the excellent books to which I have already referred will become the useful texts which they were intended to be.

Finally I should like to say that the desire to write this book derived from my own personal experiences in working with factor analysis over almost thirty years. I came into psychology with no mathematical background and would have been entirely defeated by factor analysis had I not had the good fortune to work with Professor Warburton of Manchester University, who had the patience to explain what I could not understand without that contempt for one's stupidity which all too often accompanies mathematical or statistical explanation. I hope that the *Easy Guide* will provide just such an explication.

GENERAL DESCRIPTION OF FACTOR ANALYSIS

I shall begin with a general, verbal description of factor analysis, describing the aims and logic of the method, the kinds of questions it can answer, and its different varieties. I shall do this because I

have found that the mathematical procedures become more comprehensible if the purpose and nature of factor analysis is already known.

What is factor analysis? Factor analysis consists of a number of statistical techniques the aim of which is to simplify complex sets of data. In the social sciences factor analysis is usually applied to correlations between variables and these must now be described.

Definition of a correlation coefficient A correlation is a numerical measure of the degree of agreement between two sets of scores. It runs from +1 to -1: +1 indicates full agreement, 0 no relationship and -1 complete disagreement.

Suppose we have two sets of scores in English and maths (Table 1.1).

Table 1.1 Scores in English and maths

Subjects	English	Maths
1	100	100
2	95	95
3	91	91
.	.	.
.	.	.
N	0	0

In this example (which would be virtually impossible in real life), in which the scores of N subjects (note that N describes the size of a sample) on English and maths are set out, it is evident that there is perfect agreement. In this case the correlation would be +1. If the opposite had occurred and the top person in English had been the bottom in maths and the second person in English had been second from bottom in maths and so on through the list the correlation would have been -1.

Notice that with correlations of ± 1 perfect predictions from one score to the other can be made. That is one of the important features of correlations in the social sciences: they enable prediction from one set of scores to another, although in real life correlations are almost never equal to ± 1 . The closer they get to 1 (regardless of sign) the higher the degree of agreement between the scores and

thus the better the possible prediction. If correlations are squared the percentage agreement of the sets of scores is indicated.

Thus a correlation of 0.8 indicates 64 per cent agreement and 0.2 shows only 4 per cent agreement (0.2 squared is 0.04, not 0.4). A correlation of 0 indicates that there is no relationship at all between the two sets of scores.

Definition of a correlation matrix A correlation matrix is a set of correlation coefficients between a number of variables.

Thus if we have five variables the correlation matrix will be of the kind shown in Table 1.2. The following points should be noted about this correlation matrix.

Table 1.2 A correlation matrix

Variable	Variable				
	1	2	3	4	5
1	1.0	0.31	0.40	0.24	0.35
2	0.31	1.0	0.53	0.12	0.04
3	0.40	0.53	1.0	0.01	0.42
4	0.24	0.12	0.01	1.0	0.25
5	0.35	0.04	0.42	0.25	1.0

1 Each variable is assumed to correlate perfectly with itself. In practice, which will be discussed in later chapters of this book, this is not always the case. What goes into the diagonals of the correlation matrix is important in understanding and interpreting factor analyses.

2 There is considerable redundancy in the matrix, each entry appearing twice, as it logically must.

3 Factor analysis is designed to simplify these correlation matrices. With as few as five variables there are 25 entries — which are hard enough to keep in one's head. In a large study with perhaps 100 variables there are 10,000 correlations. Without some simplifying procedure such a matrix would be incomprehensible.

4 In a large matrix of correlations, it is reasonable to ask what might account for the correlations. An example will clarify the point. Suppose that we have administered 100 different tests of ability and school attainment. In fact, the resulting correlation matrix would consist of positive and often high correlations in the region of 0.5 and 0.6. A factor analysis would reveal that

these could be accounted for by a small number of factors: intelligence, verbal ability and spatial ability. Thus instead of having to look at the scores on a 100 tests to understand these correlations, which no human being is able to do, we could understand them in terms of three scores — on intelligence, verbal ability and spatial ability.

The example in (4) asserts and illustrates that factor analysis can simplify a matrix of correlations, but it is not meaningful without further explanation and this is set out below.

Definition of a factor As Royce (1963) has demonstrated, whilst there have been many different definitions of a factor there is a common underlying trend to them all. Essentially a factor is a dimension or construct which is a condensed statement of the relationships between a set of variables. This has been more precisely refined by Royce (1963), who states that a factor is a *construct operationally defined by its factor loadings*.

This is an excellent definition although obviously factor loadings need to be defined.

Definition of factor loadings Factor loadings are the correlations of a variable with a factor.

Table 1.3 Illustration of a factor analysis

Variables	Factor 1	Factor 2	Factor 3
Intelligence	0.82	0.63	0.44
Non-verbal IQ	0.78	0.35	0.51
Vocabulary	0.68	0.64	0.21
Rhyming	0.28	0.59	0.18
Algebra	0.45	0.20	0.38
Geometry	0.50	0.17	0.69
Physics	0.41	0.13	0.37
Latin	0.58	0.70	0.20
French	0.32	0.68	0.17
History	0.25	0.43	0.12
Engineering	0.49	0.09	0.60

An artificial illustration will clarify this definition. Suppose that we have computed the factor analysis of a set of ability and attainment tests. At this point how this is done (which will be fully explained in later chapters of this book) does not matter. The factor analysis

might well look like the analysis set out in Table 1.3. This example shows three factors in a typical factor analysis of abilities. There would be other factors but these illustrate the nature of factors and factor loadings extremely well.

The interpretation and meaning of factors

- 1 The results of a factor analysis simply set out a number of factors, as shown above. The meaning of these factors has to be deduced from the factor loadings – the coefficients set out in Table 1.3. The factor loadings are what are computed in the factor analysis. It must be stressed, however, that all interpretations of factors, based on loadings, should be validated against external criteria. Thus if we think we have an intelligence factor we might investigate whether high scorers showed other signs of intelligence, e.g. held good jobs or were highly successful at exams and so on.
- 2 The factor loadings are correlations of the variables with the factors. It is usual to regard factor loadings as high if they are greater than 0.6 (the positive or negative sign is irrelevant) and moderately high if they are above 0.3. Other loadings can be ignored. A more precise account of the significance and importance of factor loadings will be given later. From these loadings the following three deductions can be made.
- 3 Factor 1 correlates with all the school subjects but most highly with the intelligence tests. (Actually the technical term is loads rather than correlates and I shall use this from now on in the text.) This suggests that this is a general ability, which is important in all these tests. Since intelligence is usually defined as a general reasoning ability, it appears that this first factor should be identified from its loadings as general intelligence. Note that some school subjects load more highly on this factor than others – suggesting that intelligence is not so important for some subjects. History, in this example, appears not to require as much intelligence as Latin or geometry. Although this is an artificial example there is some evidence for such a difference (see Kline 1991).
- 4 Factor 2, by the same process of deduction as used above, must be verbal ability, as it loads highly on all those tests and subjects which involve language but rather low on other tests.
- 5 Factor 3 loads highly on intelligence and on geometry and engi-

neering. Both of the latter subjects require an ability to orient things in space and this suggests that this factor is spatial ability.

These five points, taken from the artificial example given in Table 1.3, illustrate clearly the nature of factors as constructs or dimensions defined by their loadings. They also illustrate how a complex set of data, in this instance a large number of scores on different tests, can be reduced and simplified. Thus if we wished to predict how well a subject would perform in these subjects in the future we could do so from scores on these three factors. We would not need to know all the test scores.

This, therefore, is the rationale of specifying factors as constructs or dimensions defined by their factor loadings. Given this definition I shall now discuss the utility of factor analysis in the social sciences, its aims and its logic.

THE UTILITY OF FACTOR ANALYSIS

Given, then, that factor analysis can be used to simplify correlation matrices, an important question still remains: what can be done with it and how can it be useful in the social sciences?

To answer this, one further point needs to be made about factor analysis. In the application of this method a distinction needs to be drawn between exploratory and confirmatory factor analysis.

Exploratory factor analysis: an illustration

In exploratory analysis the aim is to explore the field, to discover the main constructs or dimensions. It was for this purpose that factor analysis was originally developed by Spearman (1904), in the area of human abilities. He attempted to answer the question of why it was that human abilities are always positively correlated, what is referred to in the factor analysis of abilities as the positive manifold, i.e. all the correlations in the correlation matrix are positive.

From our description and interpretation of the factor analysis given in Table 1.3 it is clear how factor analysis can be used to answer this question. Rephrased in factor analytic terms the question becomes: what constructs or dimensions could account for the correlations between abilities? It was from this early work that the notion of general intelligence was developed – as factor analysis

revealed that a general factor, loading on all variables, was highly important in accounting for the correlations. More precise definitions of what is meant by 'account for' and 'highly important', when referring to factors, will be given in Chapters 3 and 4.

This first example illustrates clearly one of the main uses of factor analysis in the social sciences. When there is a highly complex field, as there almost always is in real-life human affairs, as distinct from laboratory studies of human beings where one variable may be manipulated at a time, factor analysis simplifies the field by indicating what the important variables are. Let us take the educational example again. Suppose that we had before us a huge matrix of correlations between abilities and skills and we tried, without factor analysis, to answer the question of why these are positively correlated. A huge number of possibilities come to mind:

- 1 some people are good at school work and clever and this accounts for the results (this, however, is not an explanation since it simply throws the original question back a further stage: what makes these people clever and others not clever?);
- 2 some might argue that social class is the determiner;
- 3 others may argue that it is the ability to concentrate;
- 4 another hypothesis suggests that it is quality of education.

These are four obvious hypotheses and the moderate correlations between social class, educational level and test scores make two of them possible. The notion of concentration is also appealing.

Exploratory factor analysis would answer the original question and thus implicitly refute or confirm these four hypotheses. In fact, as has been mentioned, a general ability factor can be shown to be important in accounting for these correlations. As in our example of Table 1.3, there would be a general factor loading on intelligence tests and school attainment measures. Social class loads only low on such a factor. This means that social class is implicated to some extent, but not much, in performance at school. We all know of dim aristocrats and brilliant people from poor backgrounds. Of course, the actual findings are not important. If social class were a powerful determinant of school performance it would load highly on the factor on which the attainment tests loaded.

If a measure of the ability to concentrate were to be included in the matrix of correlations, this again would load only moderately on school performance. It plays a part, of course, but it is not critical. This ability is probably itself in any case related to

intelligence. Actually a moment's thought would show that concentration could not be a major determinant of performance. If this were the case there would be little variation in school performance since concentration is needed for all tasks. Again actual results are not what matters in this illustration of the utility of factor analysis. If concentration were important it would load highly on the attainment test factor. Identical arguments apply to the final hypothesis concerning quality of education. If this were influential it would load on the attainment test factor. If it were not, it would not do so.

From this example of attainment testing it is clear that exploratory factor analysis is a powerful tool in elucidating its important determiners and associated variables. In general in exploratory analysis the rule is to put in as many variables as possible and see what loads on the relevant factor. If our interest is in achievement at school we simply look for all the variables which load on the same factor as the achievement measures. If our interest were more specific, say science performance, then we would look for a factor specific to that and see what variables loaded on it.

Further examples of exploratory factor analysis

These are all practical questions. Some psychometrists have used factor analysis to examine more theoretical issues. Thus in the field of human abilities the question has concerned the number of human abilities and their structure. This has been convincingly answered by factor analysis which shows that there are two intelligence factors and three other smaller factors – the ability to visualize, a speed factor and a factor of fluency, the ability to retrieve material from memory, as has been fully discussed by Cattell (1971). Not only is this of considerable theoretical interest, but the selection of the most important factors is essential for further study of the area. Thus when the important factors are known, environmental and genetic influences can be investigated and other determinants of their variation, thus building up a psychology of abilities. Exploratory factor analysis is an essential first step in the investigation of complex areas of human psychology.

It must not be thought that factor analysis is only suitable for the study of attainment and ability. It is widely used in the study of personality. Thus an important question concerns what are the most important personality variables. Exploratory factor analysis

has proved excellent for this purpose. As many measures as possible are administered to subjects and the results are factored. In fact there is a general consensus that there are five pervasive factors, as is fully discussed in my *Personality: The Psychometric View* (Kline 1993).

Another important use of factor analysis is in the construction of psychological tests. The logic of this procedure is simple. Any test, ideally, should measure only one variable. To ensure this a large number of possible items are administered to subjects and the correlations between the items are subjected to factor analysis. Items which load the general factor are selected for the test.

A final example will illustrate the power of exploratory factor analysis as a statistical method. In a small-scale study of occupational choice we factored occupational choice along with a number of personality and interest tests. We then looked at what variables loaded on the factors concerned with such choice. Thus it was found, for example, that introversion and intelligence went along with science as a preferred career (Parker and Kline 1993).

In summary, it can be seen that exploratory factor analysis is ideal where data are complex and it is uncertain what the most important variables in the field are. This is hardly surprising given our definition of factors as constructs or dimensions which account for the relationships between variables and which are defined by their factor loadings.

Confirmatory analysis

Originally factor analysis was simply an exploratory statistical method. Recently however it has become possible to test hypotheses using factor analysis, a method developed by Joreskog (1973) and called confirmatory analysis. In this method, based upon previous studies or on relevant theory, factor loadings for the variables are hypothesized. Confirmatory factor analysis then proceeds to fit these loadings in the target matrix, as it is called, as closely as possible. How good the fit is can also be measured. Since the scientific method, as it is generally conceived, e.g. by Popper (1959), involves testing hypotheses confirmatory analysis has become acceptable to psychologists who were previously resistant to exploratory methods.

Confirmatory analysis will be scrutinized carefully in Chapter 6,

although a full mathematical coverage will be beyond the scope of this book. However, it is sufficient to say at this point that in the social sciences it is often so difficult to specify with any precision what the factor loadings should be that confirmatory analysis is not highly useful. However, if the target matrix is specified in a more general fashion, e.g. each variable being specified as a high, low or zero loading, then it is difficult to reject the hypothesized target matrix. All these problems and difficulties together with the advantages of confirmatory factor analysis will be discussed, as has been stated, in Chapter 6.

It can be seen from all these points and arguments that factor analysis is a powerful technique of investigation with wide application in the social sciences. Yet although it is much used (or more strictly misused, which is part of the reason for writing the *Easy Guide*) many researchers, for whom it would be valuable, do not employ it because of a number of objections to the method. These will be briefly mentioned now, although they are all easily dealt with, as will be shown in the relevant chapters of this book.

OBJECTIONS TO FACTOR ANALYSIS

- 1 The main objection to factor analysis is that there is an infinity of mathematically equivalent solutions. While this is true, it is also the case that psychometrists have developed powerful methods of choosing the right solution. All these methods are fully explained in Chapter 5 of this book.
- 2 Factor analysts frequently disagree as to what are the most important factors in the field. This is especially so, for example, in personality. This is really a specific instance of the first objection as it appears in practice. Often, disparity of results is due to poor factor analytic methods which readers of the *Easy Guide* will learn to spot.
- 3 It is difficult to replicate factor analyses. Again this stems from the first basic objection, although with sound methodology it can be overcome.
- 4 It is sometimes said that with factor analysis you only get what you put in so that it is difficult to see how the method can be useful. This objection is based upon faulty logic. It is the case that if, for example, in a study of abilities no measures of musical ability were included, then no factor of musical ability could emerge. From this it follows that in exploratory analyses it is

essential to sample variables as widely as possible. However, it does not follow from this that in factor analysis you only get out what you put in.

In test construction some authors include items which are essentially paraphrases of each other. These certainly will load a factor and in this case the objection does indeed hold. Cattell (1978) calls these useless factors bloated specifics and how to eliminate and identify these factors will be discussed in Chapters 8 and 10. However, it does not follow that this is always the case. For example, if I were to factor simply a set of school examination results, including only school subjects, three factors or so would account for the correlations: the ones which we discussed above - intelligence, verbal ability and spatial ability. Although no specific measures of these factors had been included in the study they would emerge because, in fact, they do determine the relationships between measures of school achievement.

In brief, this objection is sometimes valid. However, generally this is not so and, ironically, one of the most attractive aspects of factor analysis as a statistical method is that it can reveal constructs which were previously unknown.

CONCLUDING REMARKS

So far in this chapter I have explained the purpose and function of factor analysis and have defined factors. However, this verbal explanation has to be accepted on trust. In the following chapters I shall explicate the surprisingly simple algebra of factor analysis in such a way that its ability to account for correlations, the status of factors as constructs, the definition of factors in terms of loadings and the need to validate factors against external criteria will become obvious. Thus the algebra in the *Easy Guide* will aid understanding. With the basics of factor analysis understood, readers who wish to do so will be able to tackle the more comprehensive texts to which reference will be made in the relevant chapters. These books describe some of the more mathematically sophisticated procedures of factor analysis.

SUMMARY

- 1 Factor analysis is defined generally as a method for simplifying complex sets of data.

- 2 Factor analysis is usually, in the social sciences, applied to correlation matrices.
- 3 A factor is defined as a construct or dimension which can account for the relationships (correlations) between variables.
- 4 Factor loadings are defined as the correlations between variables and factors.
- 5 The definition in (4) enables a more precise definition of factors as constructs defined by their factor loadings.
- 6 The meaning and interpretation of factors is derived from their loadings, although these meanings must be externally validated.
- 7 Examples of the use of factor analysis, both practical and theoretical, in the fields of ability and personality are given.
- 8 A distinction is drawn between exploratory analysis, the most common use of the technique, and confirmatory analysis in which hypotheses can be tested.
- 9 Some objections to factor analysis are briefly discussed.
- 10 The most important of these - the infinity of equivalent solutions - is answered.
- 11 The claim that from factor analysis no more than is put in can be taken out is also discussed and is found to hold only in a small number of instances.