CHAPTER 6
DESIGNING AND CONDUCTING FACTOR ANALYTIC STUDIES

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 6
DESIGNING AND CONDUCTING FACTOR ANALYTIC STUDIES

6.0. Introduction

In previous chapters the common factor model has been presented in the context of both a population and a sample, and a number of issues have been discussed which can influence obtained factor solutions. These issues include such things as the effects of random sampling, selective sampling, sample size, the presence of unique factors, standardization of attributes, selection of attributes, and number of attributes. The discussion of these issues has been carried out in essentially a theoretical context. The purpose of the current chapter is to translate this discussion into principles and guidelines for practice. This should serve to aid researchers in designing and conducting factor analytic studies and in determining a general approach to take to the actual factor analysis of their data. Subsequent chapters will be based on an assumption that studies have been carried out and that data have been obtained, and that a general approach to the analysis has been determined.

6.1. General Approach to Factor Analytic Research

Recall that the general objective of factor analytic research, as stated in Chapter 1, is to determine the nature of the underlying factors and to develop an understanding of their relationships to the surface attributes and to each other. A very important fundamental point about factor analytic research is that it is virtually impossible to achieve this objective via a single study. That is, it is not feasible to attempt to carry out a single study in which a domain and population are defined, a set of attributes adequately representing the factors in the domain are constructed, measures on these attributes are obtained from a sample, and the results of a factor analysis are valid, stable, and clearly interpretable. In all likelihood, the results of a first attempt to achieve this will reveal a variety of problems; e.g., attributes which do not represent factors as intended, the occurrence of unanticipated factors, the absence of anticipated factors, etc. It is very unlikely that such phenomena can be avoided in an initial study in a domain, simply because of the lack of knowledge about the nature and dynamics of the underlying factors.

The implication of all of this is that the achievement of the general objective of factor analytic research will almost always require a series of studies. In the initial studies a domain is defined with little prior idea of the nature of the factors in the domain. Attributes are selected and/or constructed with an effort to achieve wide representation of the domain. There may be some prior notion of the existence of particular common factors, and attributes should be constructed so as to attempt to represent those factors. But the guiding principle should be to obtain wide coverage of the domain, so as not to miss any important common factors. Once data are obtained, the factor analysis is conducted in an exploratory manner. Methodology for exploratory factor analysis is discussed in Chapters 7 through 13, and a more detailed discussion of this general approach is presented later in this chapter. For present purposes, the major point is that there is no testing of prior hypotheses; rather, the methodology is designed to explore the data to determine how many factors might be present, and to achieve some rough indication of, or conjecture as to, their nature and relationships to each other and to the surface attributes. Based on such results, the experimenter begins the task of refining the battery of attributes and designing subsequent studies to more closely investigate the factors in the domain under investigation. This process typically involves deleting attributes with undesirable properties (e.g., inadequate representation of factors), and adding new attributes so as to achieve a more complete understanding of the factors observed in the early studies. Distinctions between factors may be developed by addition of attributes which should "fan" between factors. Achievement of such distinctions should lead to better understanding of the nature of the factors. In these later studies, exploratory factor analysis methods would still be used because the objective is to continue to explore the data in search of information relevant to the impact of the modifications in the battery on the factor structure.

The succession of studies should be concern as aiding our understanding of the nature of the factors. In one sense, these studies involve changing samples of attributes included in the batteries used in the studies. This is a different question than the stability of results questions which would be approached by repeated studies with a constant battery of attributes having data collected for new samples of individuals.

The process of refinement of the battery, collection of additional data, and exploratory analysis may continue until the experimenter believe he or she has developed a clear and well-founded hypothesis about the number and nature of the factors in the domain under study, along with a battery of attributes which clearly reflect those factors. At this point, it would be most desirable to carry out a final study with the objective of testing the hypotheses about the factor structure in the domain. The study would employ the "final" version of the test battery, and the factor analysis methodology would be confirmatory rather than exploratory. This methodology is presented in Chapters 14 through 17, and a brief discussion of this approach is offered later in this chapter. For purposes of the present discussion, the major point is that confirmatory factor analysis provides for an explicit fitting and testing of the hypothesized factor structure, where the hypothesized structure is defined in terms of the number of factors and the hypothesized pattern of their relationships to each other and to the surface attributes. Results of the confirmatory analysis indicate the goodness of fit of the hypothesized structure to the data, and provide information to help evaluate what problems, if any, still exist. Clearly, the experimenter would hope that the hypotheses that were developed in the earlier studies would be strongly supported by the confirmatory analysis in the final study in the series. Such an outcome would support a claim that the factor structure in the domain in question is well understood and well represented in the battery of attributes. This general description of a series of factor analytic studies should clarify the notion that the final objective probably cannot be achieved via a single study.

A brief description of an illustration of a series of factor analytic studies may help to further clarify the points made in the preceding paragraph. Suppose a researcher wishes to investigate the factorial structure in the domain of mental abilities. Such a project would begin with the assembly of a large battery of tests of a wide range of abilities. The experimenter probably would have some prior notions about basic internal attributes which might exist, such as mathematical, verbal, analytical, and spatial abilities, etc. The initial battery of attributes would include attributes that were thought to represent these factors, along with a variety of other attributes intended to represent the range of abilities in the domain. As stated above, the objective is to use a large and diverse initial battery so as not to overlook the existence of important common factors. Of course, the degree to which the prior notions are formed and supported will influence the size and diversity of the initial battery. However, even when prior hypotheses about the factors are present in early studies, they should be regarded as crude and subject to modification as research progresses. Regardless, the exploratory factor analysis in the initial study provides an initial indication of the number of common factors, along with estimates of the parameters of the model (factor weights, factor intercorrelations, and unique variances of the attributes). Based on this information, the researcher begins the process of refining the battery of attributes. Attributes which do not measure factors in a useful manner can be deleted. For instance, if there are a large number of attributes which measure the factor "numerical facility," some of those attributes can be deleted without sacrificing representation of this factor in the battery. More specifically, one may wish to delete those measures of the "numerical facility" factor which have the lowest communalities, as long as those attributes are not critical measures of other factors also. It may also be desirable to add attributes to the battery. For instance, if there is some indication of a factor which seems to represent "reading comprehension," additional attributes can be constructed and added to the battery which are intended to measure this factor directly. This would be an effort to more clearly determine this factor and to enhance its representation in the battery. Distinctions between related factors such as word fluency and verbal ability could be studied by the construction of several new tests which range between emphasizing word fluency and verbal ability. After such refinements are determined and carried out, a new sample of data could be collected and another exploratory factor analysis conducted. The results in this second study should be clearer, but may indicate the need for further modifications of the battery. Obviously, this process can continue until the researcher believes that a clear hypothesis about the factors in the domain of mental ability has been developed, and that a battery of attributes has been defined which provides a clear representation of these factors. At that point, a final study could be conducted, with the data being subjected to confirmatory factor analysis. Positive results from this analysis would support the hypotheses about the factor structure in the domain of mental abilities, as well as the construct validity of the tests in the battery. Negative results would indicate problems of some type with the hypotheses and/or the data, necessitating further investigation. Specific techniques for evaluating these results will be discussed in the chapters on confirmatory factor analysis.

The critical point in this discussion is clearly that the achievement of the objective of factor analytic research requires a series of studies, proceeding from initial studies where hypotheses are only loosely formed and analyses are exploratory, to final studies where confirmatory analyses are conducted to test well-developed hypotheses. While the actual number of studies involved in a series will vary greatly from one domain to another, progress in understanding the factor structure in a domain is achieved in a step by step fashion through such a series of studies by one or more investigators.

Given this general view of factor analytic research, let us now turn our attention to the practical implications of a number of issues discussed in Chapter 5.

6.2. Selecting Observations from a Population

In Chapter 5 we discussed in a theoretical context some effects and issues involved in the process of sampling observations from a population. We will now consider these issues in a more practical context, and will follow the distinction employed in Chapter 5 between random and selective sampling.

6.2.1. Practical Implications Under Random Sampling

Let us begin by briefly reviewing how factor analytic results are affected under random sampling. As discussed in Chapters 4 and 5, two primary effects can be seen. First, random sampling affects the covariances among the factors; that is, the covariances among the factors in a sample will be affected by the random characteristics of that sample. This is represented in Eq. (4.13). Second, under random sampling the assumption that unique factors are uncorrelated with each other and with the common factors will generally be violated. As a result of the invalidity of this assumption, sampling error affects the common factor weights obtained in a sample. Additional sampling error effects arise as a result of the standardization of the common factors and of the attributes in the sample. These phenomena were discussed in Chapter 4 and demonstrated in Chapter 5.

An important point is that the magnitude of these effects of sampling error is dependent in part on several other characteristics of the data, as discussed in Chapter 5. First, the lower the unique factor weights, the less influence sampling error will have on the obtained common factor weights. Second, standardization of attributes in a sample introduces additional sampling error affecting the common factor weights; i.e., the analysis of a correlation matrix rather than a covariance matrix adds additional error in the recovery of the population common factor weights. Third, the use of a large sample reduces the effects of sampling error described in the previous paragraph by improving the stability of the obtained solutions and the correspondence between obtained results and population parameters. Finally, it is very important to recognize that there is an interactive effect between sample size and the other two influence just mentioned-- the unique factor weights, and standardization of attributes. When sample size is large, the amount of sampling error arising from these influences is reduced.

These phenomena have direct implications for applied factor analytic research. In particular, they suggest some basic guidelines for practice. First, since the impact of sampling error increase as unique factor weights become larger, it is desirable to select attributes which will have small unique variances. When a succession of factor analytic studies is conducted, this is achieved by eliminating those attributes with low communalities and retaining those with high communalities. Alternatively, one might add attributes to a battery in order to more strongly represent certain common factors, thus increasing communalities of some of the original attributes. The general point is that one should seek to develop a battery containing attributes with high communalities. Not only will this help to provide clear representation of the common factors, but it will also reduce the impact of sampling error on the obtained results.

A second point is that it is desirable to apply factor analysis methods to covariance matrices rather than correlation matrices. This will eliminate the process of standardization of the attributes within the sample, and will thus avoid this type of sampling error. However, at the same time, it must be understood that standardization is often desirable in practice when measured attributes are characterized by radically different scales of measurement. Standardization of the attributes would then eliminate the influence of these highly disparate scales on the results of the factor analysis, thus rendering the solutions easier to interpret. In practice, then, the investigator must determine whether the scales of measurement of the attributes are sufficiently different to warrant standardization, thus implying a willingness to accept the additional source of sampling error introduced by this process.

Thirdly, it is most desirable to obtain a large sample. This will have two important effects. It will improve the stability of the results and the correspondence between the sample results and population parameters. It will also reduce the impact of sampling error arising from the presence of unique factors and the standardization of attributes. That is, a researcher can be less concerned about sampling error arising from those sources if the sample size is very large. This is an important point, especially with respect to the issue of unique factors. In early studies in a domain, it is likely that substantial unique factor weights may be present. Since results obtained from early studies provide the basis for subsequent studies, it is important to do whatever possible to reduce the impact of sampling error. A large sample will thus be quite important in the early stages, both to enhance the stability of the obtained results and also to reduce the influence of sampling error arising from unique factors and possible standardization of attributes.

6.2.2. Practical Implications Under Selective Sampling

It is most important for researchers to understand that selective sample is the rule rather than the exception when samples are obtained in factor analytic research. As discussed in detail in Chapter 5, actions of the experimenter, of institutions, or of the subjects themselves routinely affect the nature of the obtained sample in such a way that it becomes more homogeneous with respect to certain variables. This occurs when subjects are included in the sample on the basis of their score on one or more variables, called selection variables. In some situations the experimenter may conduct selective sampling deliberately in order to obtain one or more samples which are homogeneous in certain ways; e.g., this would be the case when a researcher wished to compare children of different ages. In other situations, selective sampling may be intentional or unrecognized; e.g., when college students are employed as subjects in psychological research. Regardless of this issue, some degree of selection operates routinely in practice and researchers must be aware of its potential impact on obtained results, as well as of any procedures they can employ to reduce or control that impact.

As discussed in Chapter 5, the general effect of selective sampling is to increase the homogeneity of the sample and thus reduce variances and intercorrelations of attributes. This effect influences results of factor analysis applied to such samples. In Chapter 5 we developed a theoretical framework for linear selective sampling which postulated the dynamics of this phenomenon and provided a basis for studying the effects of selective sampling on obtained factor analysis solutions. The results of this approach showed that common factor weights and intercorrelations in a selected subpopulation will be affected by the selection process when the factors are standardized in that subpopulation. The usual effect will be for those weights and intercorrelations to be reduced due to the reduced variability in the subpopulation. Of greater interest were results involving the relations of solutions obtained from different selected subpopulations. These results showed that the common factor intercorrelations and weights will be different in the two groups. Furthermore, when covariance matrices are analyzed, the common factor weights in the two groups will be proportional by columns. However, when correlation matrices are analyzed, the rows of the weight matrices also will undergo a rescaling which results in an incomparability of the weights from the two groups. As a result of this effect, it is not advisable to analyze correlation matrices in such a setting. The entire issue of factor analysis in multiple groups will be considered in detail in Chapter 20.

The effects just described represent the case in which the selection variables are related only to the common factors, and not to the specific or error factors. When there are nonzero relations of the selection variables to the specific or error factors, considerable complexities arise with regard to the effects of selection on obtained factor solutions. This can occur fairly easily. For instance, whenever the selection variables themselves, or parallel tests to those variables, are included in the battery being factor analyzed, this type of relation will be present. In such a case, complex effects on the factor weights will occur, and additional common factors will arise and will be characterized by some negative weights. Given the complexity of such effects and the difficulty in recognizing and interpreting them in practice, we consider it highly advisable to avoid this situation whenever possible. That is, we recommend against including the selection variables themselves, or parallel tests to those variables, in the battery to be factor analyzed.

A final comment about the effects of selective sampling is that they will not occur in isolation in practice. Rather, they will occur in conjunction with other effects discussed in this and earlier chapters: the effects of random sampling and the effect of model error. These effects cannot be separated in practice. A major task for the researcher is to strive to design and conduct research in such a way that undesirable effects of these phenomena can be minimized, or at least recognized and understood.

6.3. <u>Selecting Attributes from a Domain</u>

In Chapter 4 we discussed and illustrated potential effects of attribute selection on the obtained factor solution. The major points made then were that (a) the deletion of attributes from a battery can cause common factors to vanish altogether or to become specific factors, and (b) the addition of attributes to a battery can cause specific factors to become common factors, or can give rise to new common factors altogether. It is important that applied researchers understand these phenomena because they are quite relevant to the process of constructing and modifying attribute batteries in a sequence of factor analytic studies.

As stated earlier, when an initial battery of attributes is constructed, attributes are selected so as to represent any hypothesized common factors, and also to provide wide coverage of the domain under study so as to enhance the opportunity to discover whatever important factors exist in the domain. In subsequent studies the initial battery is modified by deleting or adding attributes. This is done so as to enhance the representation of the common factors which are being identified. To achieve this goal, the researcher must understand and make use of the phenomena discussed above, regarding the effect of attribute selection on obtained factor solutions. To enhance the representation of a common factor, new attributes which are intended to measure that factor can be added to the battery, and attributes from the original battery which were designed to measure that factor but which failed to do so can be deleted. When an important specific factor is suspected to exist (e.g., when an attribute has a very high reliability but a very low communality), new attributes can be added to the battery in an effort to bring out that factor as a common factor. In such ways, the researcher can carefully modify the battery so as to improve the representation of the common factors.

Another important issue regarding selection of attributes involves the number of attributes to include in a battery. In Chapter 4 this issue was discussed in the context of the concept of overdetermination of common factors. That is, it is necessary to have a sufficient number of attributes so as to achieve an adequate degree of mathematical overdetermination of the common factor weights. Table 4.10 provided information about the minimum number of attributes necessary to provide a given level of overdetermination for a given number of common factors. This information is very useful in the process of designing factor analytic studies. In early studies in a domain, the researcher may have only a crude estimate of the number of common factors present. In such a case, the number of attributes included in the battery should be sufficient to provide a high level of overdetermination of the common factors. As represented in Table 4.10, a coefficient of overdetermination of 3.0 or 3.5 may be desirable in such research. In later studies, when the number of common factors is known more accurately, a smaller coefficient of overdetermination would be sufficient; a value of 2.0 or 2.5 probably would be acceptable in most cases. In practical terms, the effect of including too few attributes in the battery to adequately overdetermine the factors would be to obtain unstable parameter estimates and poorly defined factors. Results would very likely be difficult to interpret, and these problems would have detrimental effects on the design and conduct of subsequent studies in the domain.

A final important issue in this area involves the interaction of sample size with the degree of overdetermination of the common factors. It has been shown (MacCallum, Widaman, and Lee, 1986) that a large sample is less necessary when common factors are highly overdetermined. An implication of this finding for practice is that when large samples are not easily obtained, an experimenter can enhance the quality of results by employing a battery of attributes which strongly overdetermines the common factors. That is, when the common factors are represented strongly by a number of attributes in the battery, results from relatively small samples may be quite stable and interpretable. Conversely, when the study is characterized by a low coefficient of overdetermination, the need for a large sample in order to obtain stable and interpretable results is greatly increased.

Given that the issue of sample size has been discussed in different contexts in this chapter, it is useful to briefly state a number of issues in combination. Combining the point made in the previous paragraph with points made in Section 6.2.1, it can be seen that a large sample is most important when unique factor weights are not small, attributes are standardized, and the coefficient of overdetermination is low. On the other hand, it is possible to obtain stable and interpretable results with relatively small samples when unique factor weights are small, attributes are not standardized in the sample, and the coefficient of overdetermination is high. Under these conditions, the impact of sampling error arising from unique factors and standardization will be low, and the high coefficient of overdetermination will allow for stable representation of the common factors even in smaller samples.

6.4. Final Issues in Conducting Factor Analytic Studies

After a researcher has obtained data from a sample of observations on a battery of attributes, there are a few final issues which should be considered before factor analysis techniques are applied to the data. Three such issues will be discussed in the subsequent sections.

6.4.1. Standardization of Attributes and Factors

The issue of standardization of attributes and of factors has been discussed a number of times in various contexts. These developments have some important practical implications. Let us first consider the issue of standardization of attributes; i.e., the factor analysis of a correlation matrix rather than a covariance matrix. Though this is by far the most common approach to factor analysis in the applied literature, it is not necessarily the most desirable approach. It is an attractive approach in practice because it simplifies a number of aspects of the analysis. As data, correlations are simpler to interpret than covariances. In addition, factor weights are simpler to interpret and compare when attributes have been standardized. As noted earlier in this chapter, standardization eliminates influences of widely different raw scales of measurement of the attributes. However, as discussed and illustrated in Chapter 5, this standardization does introduce a source of at least a slight amount of additional error. Furthermore, potentially meaningful information about differences in variances of the attributes is lost. As a result, we encourage the analysis of covariance rather than correlation matrices when possible.

Most factor analysis is conducted under the imposed condition that the factors are standardized. Researchers must keep in mind that this is a separate and independent issue from standardization of attributes. Despite the fact that this technique introduces an additional sampling effect on the solution, it is generally acceptable. The sampling effects caused by this standardization are simple multiplicative effects on the weights for each common factor. Furthermore, standardization of the factors in the sample serves the dual purpose of simplifying interpretation of solutions (e.g., since they can be interpreted in terms of factor intercorrelations rather than intercovariances) as well as resolving the identification problem for the factors. This problem, discussed briefly in Chapter 4, involves the fact that it is necessary to establish a scale for the factors in order to estimate parameters involving those factors. If this problem is not resolved via standardization, then some other step must be taken to resolve it. While alternative approaches will be discussed in some subsequent chapters, standardization of the factors is clearly the simplest and most common way to solve the identification problem.

A final point about standardization of attributes and factors involves the matter of selective sampling. It was shown in Chapter 5 that when solutions are obtained from different subpopulations, the standardization of attributes and factors within the subpopulations will give rise to systematic differences in common factor weights. This phenomenon argues further for the analysis of covariance matrices, since that would eliminate one of the sources of such differences. The issue of standardization of factors is more complex since, as noted above, if factors are not standardized some other approach must be taken to identify their scales. This issue will be considered for the case of subpopulations in Chapter 20.

6.4.2. Exploratory vs. Confirmatory Factor Analysis

Prior to conducting a factor analysis, the researcher must determine whether to use exploratory or confirmatory methods of analysis. The primary issue upon which this decision is based is the degree to which clear prior hypotheses are present regarding the factor structure underlying the battery of attributes. When such hypotheses are absent, or cannot be stated very explicitly, exploratory methods of factor analysis should be employed. This will generally be the case in early stages of research in a domain. Hypotheses will be at best loosely defined, and the general objective of the research will be to explore the factorial structure of the domain. The factor analysis methods employed in such studies involve estimation of all the parameters of the model (common factor weights and intercorrelations, and unique variances), and provide information to aid the researcher in determining the number of factors, interpreting the nature of those factors, and refining the battery of attributes for the purpose of further study of the domain. Exploratory factor analysis methods will be the subject of Chapters 7 through 13.

In later stages of research in a domain, the investigator is likely to have developed very specific hypotheses about the factorial structure of the battery of attributes. These hypotheses will concern the number of common factors, their pattern of intercorrelation, and the pattern of common factor weights. The presence of such explicit hypotheses can be taken into account via the use of confirmatory methods of factor analysis. Such methods allow the investigator to fit the common factor model to observed data under various types of constraints. For instance, the number of common factors would be defined to be a given number, and some of the parameters would be assigned fixed numerical values, rather than be estimated; e.g., certain common factor weights or intercorrelations could be fixed to zero. The remaining parameters of the model would then be estimated, and the goodness of fit of the solution to the data would be evaluated. The degree to which the solution fit the data would provide evidence for or against the prior hypotheses. A solution which fit well would lend support to the hypotheses, and would provide evidence for the construct validity of the attributes and the hypothesized factorial structure of the domain as represented by the battery of attributes. A solution which fit poorly would indicate problems with the hypotheses and/or the data, and would call for further diagnosis and study. Methods of confirmatory factor analysis are presented in Chapters 14 through 17. A major issue to be resolved prior to conducting any factor analysis is to determine whether the study calls for exploratory or confirmatory analysis.

6.4.3. Lack of Fit of the Model to the Data

Regardless of what type of factor analytic methods are employed, researchers should not lose sight of the fact that all such methods involve fitting a model to data. As emphasized in numerous contexts previously in this book, the common factor model is a mathematical model developed from a theoretical view about the underlying structure of multiattribute data. Furthermore, it must be kept in mind that that model will virtually never represent the real world exactly. Thus, in factor analytic research, there will virtually always be some lack of fit of the model to the observed data. It is important that researchers understand the sources of this lack of fit, as well as how to attempt to reduce it.

As discussed in Chapter 4, there are two identifiable sources of lack of fit of model to the data. These are referred to as model error and sampling error. Model error can best be understood in the context of the population, where the common factor model simply will generally not precisely account for the variances and covariances of the surface attributes. The presence of model error is the reason for the distinction between surface attributes and modeled attributes used throughout this book. Sampling error is an additional source of error manifested by the fact that the random characteristics of a sample contribute additional error to the estimation of the model parameters. As discussed in detail in Chapter 4, this error arises from the fact that the simplifying assumption that unique factors will be uncorrelated with each other and with common factors will not hold exactly in a sample.

In empirical applications of factor analysis, the effects of model error and sampling error are combined and are not separable. Their presence is manifested by lack of perfect fit of model to the data, along with instability and error in the parameter estimates. Despite the fact that it is not possible to separate the effects of these two sources of error in practice, it is possible to take steps to attempt to reduce their impact. In particular, the impact of model error should be reduced as a battery of attributes is refined via successive studies in a domain. The gradual development of a battery of attributes in which a given set of common factors are strongly and clearly represented will have the effect of improving the correspondence between the theoretical model and the observed battery simply because this process eliminates problems causing a lack of such correspondence. The effect of sampling error, on the other hand, can be reduced by methods discussed earlier in this chapter; i.e., by employing a large sample, attempting to eliminate attributes with high unique variances, and analyzing covariance rather than correlation matrices.

A major concern of researchers in practice should be to make efforts to reduce the impact of both model error and sampling error. Regardless of the degree of success of these efforts, some lack of fit will almost always occur in fitting the model to real data, and the measurement and evaluation of this lack of fit will be an important part of any empirical factor analysis study.