Hunter Mabon

JOB ANALYSIS – Measurement problems and applications

HUNTER MABON

JOB ANALYSIS –

Measurement problems and applications

M&B fackboksförlaget ab 27

M&B fackboksförlaget AB
Cover by: Allan Mabon
Printed by: Göteborgs Offsettryckeri AB, Stockholm

Contents

Preface

This study deals with job analysis and a number of its applications. The work performed in this field has been extensive, as a subsequent review of the literature will show. The following quote reflects, however, a view of the current research situation with which the author concurs.

"This whole area of job analysis and description has probably not received enough attention in the past. It has just not had the glamour of other areas. But because of its central role in determining equity within an organization's classification and compensation structure – an area on which the majority of current research on compensation focuses, as will be outlined below – there is a need for additional involvement by industrial psychologists in this important nuts and bolts aspect of the field." (Hinrichs, 1970)

The deficiencies do not only concern the field of equitable payment. It has also been pointed out that

"... whole disciplines have mushroomed focusing on attitudes, values, performance style and motivation in relation to the work people, do, with only the slightest suggestion that understanding *what* the individuals do is at all salient. To these writers it would appear that any measurement and certainly interpretation is contingent upon a clear and concise description of the job being performed." (Prien and Ronan, 1971)

The analysis of the actual work performed and the requirements made of the job incumbents act as a corner-stone for any further examination of man at work. In attempting to analyze jobs we are, however, immediately faced with a number of problems. The concepts of job and analysis are both ambiguous. A study must be made of the extent to which distinctions can and should be made between concepts such as task, job, position and occupation; the idea of an analysis has also to be divided into descriptions, subjective evaluations, methods of investigation and types of applications. Should descriptions and analyses be trait-oriented, behaviour-oriented or task-oriented? What contributions can be made by methods of investigation such as observation, interview, checklist, critical incident or performance of the work by the analyst himself? To what extent can job analysis be used in connection with recruitment and selection, training, organisational studies and job evaluation?

In the fields of selection and placement the problems of measurement do not cease when a position has been filled. Validation of selection and placement procedures usually involves measuring the performance of the individual in the work situation at some later date, i. e. studying the agreement between individual prerequisites and job requirements. This

collection of criterion data has traditionelly been regarded as the Achille's heel of the selection procedure.

The initial impulse to carry out a study of the methodology of job analysis has come from the author's practical work in the field of personnel administration where, in his experience, much other work is vitiated by a lack of adequate job analysis methods.

The aims of the study may be summed up as follows:

a) To perform a methodological study of the techniques used in job analysis and to attempt to draw general conclusions as to how and under which conditions various measurement techniques can and should be applied.

b) To perform a more detailed study of the application of job analysis to classification and placement problems, with special attention being devoted to validation techniques.

Cognoscenti will immediately realise that this is not an easy subject to write about at present. Many of the traditional ideas concerning man at work are in a state of flux, as are theories concerning the organisation of work at all levels within companies. These new ideas can be seen as a reaction against the traditional static approach and, in addition, they presuppose a set of values different to those implicit in much of the research examined in the present study. The present author is acutely aware of many of these developments and has in fact dealt with them extensively in other publications (Mabon, 1972, 1973a, 1973b). The aims of the present study do not however extend to a study of motivation, values and new organisational techniques such as autonomous work groups. Instead it is the *technical* aspects of what has been and can be achieved with the aid of the traditional approach which is the focus of interest here. The central problems dealt with here will on the whole remain, irrespective of changes in values, motivation, organisation and power structure in working life.

The study is a purely theoretical one, but it is based on a wide range of empirical research, primarily in Sweden, the United States and Great Britain. A theoretical framework is developed and this is used to examine some of the basic measurement problems arising in job analysis and its applications. A study of the research within the area reveals, in the opinion of the author, "a plethora of data and a paucity of generalizable research insights and theory". (Hinrichs, 1970) Furthermore, once again quoting Hinrichs and bearing in mind that research is a major growth industry with a ready market for applicable results, that "too much of today's research continues to address problems of limited relevance to either the key issues of our times or the advancement of knowledge, to be *ad hoc* instead of integrated, and to use questionable or at least untested metrics" *(loc. cit.)*.

The framework of the study is as follows: Chapter 1 contains a theoretical framework for the analysis of basic measurement problems and techniques. Chapter 2 consists of a selective account of job analysis research, with emphasis being placed upon terminology and methodological advances. In Chapter 3 this research is examined against a background of the methodological framework presented in Chapter 1 and a number of models with

varying applications are derived. It is outwith the scope of the present study to make a detailed examination of all of the practical applications of job analysis. The concluding section of Chapter 3 presents nevertheless a discussion of the applicability of various job analysis models, these applications being analyzed on the basis of the preceding methodological discussions.

In Chapter 4 a study is made of the uses of job analysis in the selection and placement situation. In Chapter 5 the study is concluded with an investigation of the criterion problem and its relationship to job analysis techniques.

Work on this thesis has been carried out over a period of years in a number of different settings. The work was originally instigated and encouraged by Dr. Stig Borg and Mr. Arne Cox, when they and the author were employed by the Swedish Council for Personnel Administration. Professor Lars Persson of the Department of Business Administration, Stockholm University, has provided encouragement and useful comments on an earlier version of parts of the manuscript. The Institute of International Management in Berlin were kind enough to provide a refuge for a brief period in 1973. A seminar held there based on a draft of the present work elicited stimulating ideas, especially from Professor Sune Carlsson and Dr. Folke Ölander. Dr. Kurt Baneryd has also been kind enough to read through the final version of the manuscript and to make a number of technical suggestions which have been incorporated in the text. Professor Ruist has raised a number of statistical points in chapter 1 which I have attempted to clarify. I would like to take this opportunity to extend my thanks to all of the above.

My thanks and appreciation must, however, in the first instance go to Professor Gunnar Westerlund of the Stockholm School of Economics, my thesis supervisor. His painstaking advice and suggestions have transformed a rather rambling account of personnel administrative problems into as stringent a methodological study of job analysis as the author has been able to achieve.

Returning to the distant past, I have to thank my parents for their initial encouragement of my academic aspirations.

Finally, I would like to thank my wife, Sylvia, and my three sons for their forebearance over a number of years. As I now approach the second half of my allotted life-span, after thirty years of sporadic study, it is my fond hope that I will be able to devote more time to them than during the first half.

Stockholm, October 1974

Hunter Mabon

Chapter 1 A framework for the analysis of measurement and research

The framework presented in this chapter is in the first instance based on a study by Boalt (1965), relating to a series of measurement scales used in order to assess the value of research work. Although some of the scales, such as number of references and length of the treatise, may be presented tongue in cheek, most of the others appear to have genuine value as a framework for assessing not only academic theses but also a broad range of research within any area. Ten of these scales will be used to structure the present chapter, namely the following

1. *Type of measurement involved*
2. *Operationality of definitions*
3. *Theoretical approach to the measurements*
4. *Nature of the response and scaling level*
5. *Methods of data reduction*
6. *Reliability*
7. *Validity*
8. *Sampling techniques and representativity*
9. *The value of the research as a means of improving theory construction*
10. *The usefulness of the research to society*

Type of measurement involved

The standard work which has most influenced the author's view of measurement problems and approaches is that of Torgerson (1958). Initially we are faced with a number of basic problems such as the definition of measurement, the types of scales used and the types of measurement involved. Following Torgerson (1958), measurement will be defined as *the assignment of numbers to the properties of objects.* This assignment of numbers to properties gives rise to at least two different types of problems, namely the attributes of the number system utilized and the method by which the number is attributed to the property; in other words the problems of scaling and type of measurement respectively. The different types of systems of numbers or types of *scales* will become clear when some of the basic features of the system of real numbers are examined. Features of the real number series are that the numbers are ordered and that the differences between numbers are ordered. In other words, it is possible to state that 16 is greater than 6, and that 16 minus 10 is greater than 6 minus 4. In addition the real number series has a unique origin denoted by zero. These three properties set in fact a standard of measurement which it is difficult to attain in many behavioural fields.

It is, however, still possible to speak of measurement or scaling even if one or more of the properties of the real number system are not present. It is a matter of some debate and to some extent of definition whether or not the minimum requirement of ordering between numbers should be retained (cf. Stevens, 1951).

Starting initially from the definition of measurement as the assigning of numbers to properties, it has been shown how the number system applied provides varying amounts of information about the properties being studied. This definition of measurement gives, however, rise to the more fundamental problem of what is actually meant by the assignment of a number, i. e. the *meaning* of the measurement process. Torgerson (1958) distinguishes three different ways of making the assignment of numbers to properties meaningful. One method is to relate the property to other properties, density being an obvious example of this type. The density of an object attains meaning by virtue of the fact that is operationally defined in terms of fundamental measurements. This type of number assignment has been called *derived measurement.* In some respects it may be regarded as begging the fundamental question of meaning. A more forthright method of explaining the meaning of a number assignment is by arbitrary definition or *measurement by fiat.* This is normally achieved by presuming a relationship between obtainable observations and the property of the object in question. In view of the fact that the behavioural and social sciences have limited opportunities for obtaining derived measurements, this is perhaps the most common measurement technique applied. It is for instance possible to define job satisfaction in terms of absenteeism, or intelligence in terms of marks obtained in theoretical subjects at school.

The third type of measurement is *fundamental measurement* where the number assignment can be based on physical laws representing the property and not on direct or indirect relationships with other properties, as in the two previous cases. Basic properties in physics such as length and volume may be regarded as fundamental measurements. The possibilities of obtaining fundamental measurements in the behavioural sciences are extremely limited.

Operationality of definitions

The physicist Bridgman (1927) in his famous dictum "the concept is synonymous with the corresponding set of operations" has also attempted to resolve the problem of concept measurement by proposing *operational definitions,* i. e. definitions which state how the concept can be measured. Derived measurement, measurement by fiat and fundamental measurement may all in fact be defined operationally. In fundamental measurement and in derived measurement the concept involved is not normally the subject of controversy with respect to measurement techniques, whereas measurement by fiat may often involve a more tenuous and disputed relationship. Measurement problems are not of course solved instantaneously by using operational definitions. Often the reliability and validity of the corresponding measurements remain to be determined.

It has been maintained since the original work by Bridgman that his approach was in fact too restricted, there being more than a suggestion that he regarded the subject of measurement as being closed whenever a satisfactory operational definition had been determined for a specific variable. The term *"operationism"* has been coined (cf. Benjamin, 1955) in order to stress the fact that there are in fact many possible operational definitions for a specific concept, depending on the uses to which one wishes to put the measurement in question. The whole concept of operational definitions has also been examined in a classic article by Dodd (1943) with the telling title "Operational definitions operationally defined". In this article the author maintains that operational definitions can be used to solve a wide range of measurement problems.

Theoretical approach to the measurements

In general, Torgerson considers that measurement studies can be divided into three stages, the *theoretical,* the *experimental* and the *analytical,* the primary basis being the *theoretical one.* Theoretically, a distinction can also be made between the following three approaches to measurement:

- *the subject-centered approach:* Different responses on the part of subjects lead to the scaling of subjects. An obvious example of this type is *mental testing.*
- *the stimulus-centered or judgement approach:* Here the variation in the reaction of subjects is regarded as being caused by differences in the stimuli; this leads to a scaling of stimuli.
- *the response-centered approach:* Here the variation in reactions is due to variations in both subjects and stimuli. *Both* may be assigned scale values.

The *judgement* and *response* techniques will be of greatest interest in the study of job analysis. The distinction between the two approaches is relatively clear, although Torgerson points out that the distinctions tend to break down in æsthetic rating situations. We will also subsequently be examining the extent to which the distinction breaks down when it comes to job analysis.

The question of at which *level* judgement and response scaling measures are being made is relatively simple in the former case, the rule being that if transitivity applies, the *attribute* is at the *ordinal* level. In the response approach, however, this cannot be stated immediately and additional procedures are needed in order to specify the class of elements. One final distinction should perhaps be made; in the *judgement* methods the nature of the attribute is known or at least assumed, whereas in the *response* methods the meaning of the underlying continuum must be the subject of further analysis.

Let us now examine more closely the judgement method. The response approach can be examined more conveniently in the following section.

The judgement approach

Empirical results have shown that the different methods applied tend to give more or less the same results. This applies for instance to the case of *paired comparisons, ranking, single-stimulus rating,* and *sorting.* It is in the very nature of measurement methods such as ranking and sorting that *transitivity* is built into the scaling system, which means that measurement takes place at the ordinal level. The exception is the method of *paired comparisons.* It is of course this method which should be applied when transitivity is doubtful. It is the Law of Comparative Judgement which has aroused most interest in job analysis circles and it is this technique which will now be enlarged upon somewhat. No detailed presentation will, however, be given here, those interested being referred to Thurstone (1948) or Torgerson (1958). The method normally applied to obtain data is that of paired comparison and, at least when applying the case V assumptions, the stimuli will be measured at the interval level (cf. Björkman and Ekman, 1957, p. 141). In the case of single-ended continua a simple transformation will normally establish a zero point, thus providing measurement at the ratio scale level.

We have already mentioned the testing for transitivity with the aid of judgement methods. Transitivity may in fact be regarded as the basic test of the unidimensionality of the material being judged. If transitivity is not present in the data, we must assume that there is some degree of measurement error or that the data are *multidimensional.* The problem we are then faced with is that of establishing the minimum number of dimensions of the data and the scale values which each stimulus has on each of the dimensions involved. An initial distinction can be made between *distance models* which are concerned with the differences between each pair of stimuli, and *spatial* models which, in addition, involve establishing the dimensionality of the space and the values of the stimulus points on each axis. The distance model normally involves measurement at the interval scale, whereas the spatial models have a fixed origin and distances must be given on a ratio scale. This gives rise to the problem of how to determine an *"additive constant"* to transfer the measurements from the interval to the ratio level.

With respect to spatial models, most research has been performed using Euclidean geometry. The original work on this problem was developed by Young and Householder (1938) in which a series of three theorems showed whether or not the stimuli can be represented by a real, Euclidean space, the minimum dimensionality of this space and the method by which the scale values on an arbitrary set of access in space can be determined.

Spatial models assume that the basic data for each pair of stimuli are available, whereas distance models are more concerned with the actual establishment of each pair of stimuli. In essence, each of the basic unidimensional models can be applied in these cases, the difference being that each basic judgement will refer to a pair of stimuli as opposed to single stimuli. A wide range of the unidimensional methods have been applied in practice and, as we shall see later, this includes the field of job analysis. The methods applied empirically include the method of equal-appearing intervals

(Ableson 1954), the Law of Categorical Judgement (Attneave, 1950), (Messick, 1954), the method of triadic combination (Richardson, 1938), the complete method of triads (Torgerson 1951, 1952; Messick, 1954) and the multidimensional method of rank order (Klingberg, 1941). The latter methods are all fairly similar and are based on Thurstone's Law of Comparative Judgement. With respect to the "additive constant" problem, Torgerson (1958) has stated "the basic assumption involved is that the appropriate value of the additive constant is that value which will allow the stimuli to be fitted by a real, Euclidean space of the smallest possible dimensionality" (page 269). Torgerson (1952) and Messick and Ableson (1956) have devised methods for obtaining values for this additive constant in specific situations. Although other methods for establishing differences between stimuli are conceivable, the above-mentioned ones appear to have been the most popular in practical applications.

Nature of the response and scaling level

As the nature of the response is to some extent dependent on the scaling level assumed, it is convenient to group the two areas in the heading within the same section.

It is usual, following Stevens (1951), to distinguish between four scaling levels, depending upon the extent to which the scale requirements on p. 11 are fulfilled. A primitive level of scaling where none of the conditions are satisfied is a *nominal scale.* For an *ordinal scale* the numbers are ordered but not the differences between them, nor is there any unique origin. The third scaling level is an *interval scale* where the numbers are ordered as well as differences between the numbers. The only remaining requirement which is not satisfied is that of a unique origin. The highest scaling level where all three of the requirements of the real number system are satisfied is called a *ratio scale.*

Although other attempts have been made to classify scales (cf. Torgerson, 1958, Coombs, 1952, and Carnap, 1950) the above distinction originally proposed by Stevens would appear to be satisfactory for the purpose in question here.

The four types of scales represent different methods of assigning numbers to properties. This assignment, which is more or less arbitrary in the case of a nominal scale, becomes more and more restricted as the scale increasingly satisfies the features of the real number series. This is most easily illustrated by the method first developed by Stevens (1951), namely that of determining permissible scale transformations at each scale level. In an ordinal scale the minimum requirement is that the numbers can be ranked, but this still provides freedom to make any monotonically increasing transformation, which may not necessarily be linear. At the interval scale level the only transformations which are permissible are linear transformations of the type $y = ax + b$. Finally, at the ratio scale level with its fixed origin, the only permissible transformation of the assigned number is a linear transformation of the type $y = ax$.

With respect to the nature of the response, an initial distinction may be made between on the one hand a response which records the relationship between a *stimulus* and a *subject*, an example being the subject agreeing with a statement and on the other hand the relationship between a *stimulus* and a *specified attribute*, for instance when the subject states that a stimulus has a certain degree of loudness or brightness. (This is, of course, the distinction made by Coombs between task A and task B). These are examined in greater detail on page 21.

A further important distinction is that between *comparative* and *categorical* responses. In a comparative response the subject might state that stimulus A is brighter than stimulus B, whereas in a categorical response the stimuli will be placed by the subject in specific classes or categories. (This is the distinction made by Coombs between *relative* and *irrelative* responses.)

Torgerson also distinguishes between *deterministic* and *probabilistic* models. In the former case it is assumed that all variation arising is due to variation in subjects and stimuli. The latter case assumes a certain amount of *error variance* both with respect to subjects and stimuli. It is obvious that the probabilistic models are more closely related to real-life scaling situations.

The response approach can be conveniently partitioned into *deterministic models for categorical data, probabilistic models for categorical data, deterministic models for comparative data* and *probabilistic models for comparative data.* As the last of these categories has been little developed, the mathematics involved rapidly becoming extremely complicated, only the first three categories will be examined here.

Deterministic models for categorical data

This is the area most closely associated with the name of Guttman (1941, 1944, 1950). In this situation there is assumed to be an underlying continuum on which each subject is placed, and which will determine the response alternative selected in a given situation. In the case of multiple category items we are faced with the situation of establishing the boundaries between the categories and their positions on the continuum, on the basis of subject responses. Being a deterministic model, Guttman scaling can also be examined as a special case of the later probabilistic models, i. e. the case where probabilities are either zero or one. This can be illustrated using the terminology first introduced by Lazersfeld (1950), namely that of the item *trace line* (see figure 1:1).

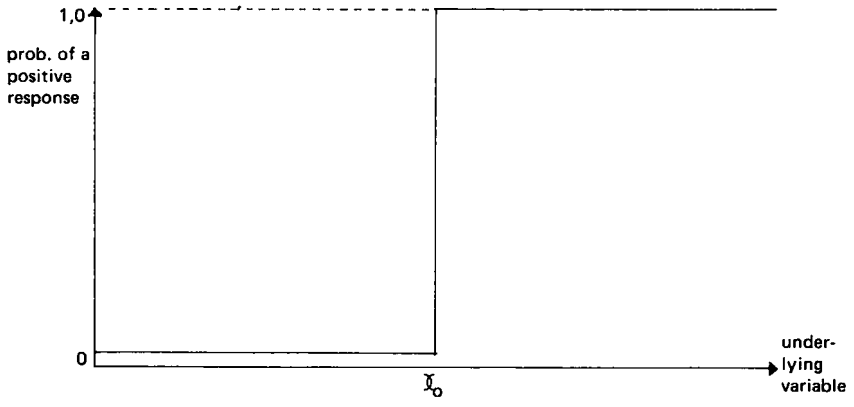


Fig. 1:1 The principles of Guttman scaling

The essence of this model is that the subjects' given responses are placed on an underlying continuum and that for each dichotomous item each subject will either answer yes or no, agree or disagree with 100 % certainty, depending on this position on the underlying continuum. The y-axis shows the probability of a positive response to the question and all subjects lying to the left of point x_0 will give a negative response, whereas all to the right will with 100 % certainty give a positive response. Space does not permit a detailed presentation of the Guttman methodology. Suffice to say that it is a relatively simple matter to test whether the data may be regarded as fulfilling specific scaling requirements.

Guttman's models have been extended to the multidimensional case although the mathematical complications are considerable. Instead of having a trace line referring to a single dimension we may have trace lines with only zero or 100 % probabilities on a number of different dimensions. An attempt to analyze the responses to a series of questions gives rise to the following terminology: in a *conjunctive* model the subject will reply positively to a question if he lies beyond the zero-level for each trace line on each of the dimensions in question. In a *disjunctive* model a positive response may be obtained if the subject is beyond the zero level on any trace line and negative if the subject is at the zero level on every trace line. Mathematically these two models are equivalent and it is in fact sufficient to examine the conjunctive one. Data can be analysed in terms of the conjunctive model, at least for two dimensions, but no general compact solution for this analysis problem is available (see Torgerson, 1958, p. 358). A further model to which some interest is devoted is the *compensatory* model in which a high score on one dimension may compensate for others at a lower level. Here the data have also shown themselves amenable to analysis up to the two-dimensional level. We shall have occasion to return to these models subsequently, as they are relevant in the selection situation when decisions are being made as to whether certain job requirements are essential, or whether they can be compensated for by others.

Probability models for categorical data

Here the tight restrictions of the deterministic model are relaxed to allow for probability in responses. The most important application of this type is that of the general *latent structure model* developed originally by Lazarsfeld. This model will be examined in some detail as it is beginning to be applied within the field of job analysis. The following account is based primarily on Lazarsfield and Henry (1968). The properties of the model may be investigated using figure 1:2 as a starting point.

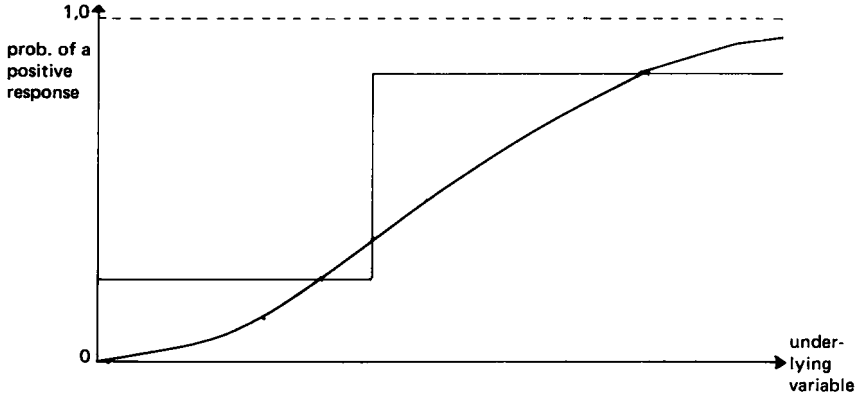


Fig. 1:2 Examples of different forms of trace lines

The structure is comparable to that of Guttman scaling. The decisive difference is, of course, that the trace lines no longer assume only the two extreme values (0 or 1) with respect to the probability of a positive response. The decisive feature of the latent structure model may be quoted *in extenso* from Torgerson (1958):

"In the latent structure model, it is postulated that no two items have any relationship to each other, above and beyond that which can be accounted for by their separate relations to the underlying variable. That is, if the underlying variable is held constant or partialed out, the items are all mutually independent of one another. For a group of subjects at any fixed value on the underlying variable, the items are uncorrelated." (Page 363) This is the decisive feature known as *local independence.* The starting-off point for latent structure analysis is to establish a series of general accounting equations as follows:

x = the underlying variable

$f_i(x)$ = the trace line of item i as a function of x

$g(x)$ = the density function of subjects along x

$g(x)\,dx$ = the proportion of subjects within $(x, x + dx)$

$f_i(x)\,g(x)\,dx$ = the proportion of the total number of subjects who are within the interval and who respond positively to item i

p_i = the total proportion who respond positively to item i

$$= \int_{-\infty}^{\infty} f_i(x)\, g(x)\, dx$$

p_{ij} = the total proportion who respond positively *both* to item i and item j

Noting the principle of local independence

$$p_{ij} = \int_{-\infty}^{\infty} f_i(x)\, f_j(x)\, g(x)\, dx$$

Similarly

$$p_{ijk\ldots m} = \int_{-\infty}^{\infty} f_i(x)\, f_j(x)\, f_k(x) \ldots f_m(x)\, g(x)\, dx$$

These *accounting equations* completely describe the relationships among the items and the problem arises as to how these equations can be solved for the "latent" parameters and the density function (distribution) of subjects. There are in fact an infinite number of solutions to this problem and in order to restrict the possible number of solutions restrictions must be introduced with respect to

a) the density function g(x)

b) the form of the trace line functions $f_i(x)$, for instance that all have the same given form.

A common restriction placed on the density function g(x) is that it be a *discrete* distribution with the subjects clustered at a given number of points. This is the *latent class model* and those subjects congregated at a single point form a latent class. As the items are independent within each class it then becomes possible to solve the accounting equations. It should, however, be noted that the assignment of subjects to each class is a *probabilistic* one which can be expressed in a series of *recruitment probabilities*; thus one specific subject may have different probabilities of being assigned to different classes and it is to be hoped that one of these probabilities is considerably greater than the others.

The other restriction, with respect to the form of the trace line functions $f_i(x)$ has also given rise to a number of different models. The *latent distance model* is for instance the probability analogy of the Guttman scale. Instead of postulating a trace line with only 0 or 1 probability it is here assumed that below a certain level the probability of a positive response is very low, whereas above this level the probability is very high. A further model presented by Lord (1954) utilises trace lines having the form of the *normal ogive*.

So far the manifest data, i. e. the p-values, have been regarded as dichotomous. The method can, however, be extended to situations where manifest data are non-dichotomous. This area has been developed primarily by Gibson (1959) and is called *latent profile analysis*. In this case latent classes can still be assumed but these are now determined by *mean values* rather than by *latent recruitment probabilities*.

Summing up, it is in fact theoretically possible to assume any number of combinations of manifest responses, trace line forms and subject distributions. Thus, the manifest responses may be dichotomous, a number of finite responses greater than two, or continuous. The trace lines may follow a "Guttman" distribution, or be linear, polynomial, following the latent distance pattern or that of the normal ogive. Similarly, the subject distribution may be discrete, discrete ordered, or continuous. Any one of these three combinations can be envisaged at least theoretically, although it is not certain that all combinations could be solved mathematically.

Multidimensional models are also feasible although this may primarily be regarded as work for future research.

The various possible models are summed up in the following table

Parameters	Alternative values
Manifest responses, i. e. p-values	discrete, 2 values; discrete > 2 values; continuous
Trace lines, i. e. $f_i(x)$	discrete, 2 values 0 and 1; discrete, 2 values near 0 and 1 respectively; linear, polynomial; normal ogive
Subject distribution, i. e. g(x)	discrete; discrete ordered; continuous

Various models may be defined in terms of these three parameters. Thus,

Guttman scaling:	discrete, 2 values; discrete, 2 values 0 and 1; discrete ordered.
Latent class model:	discrete, 2 values; *any* alternative; discrete.
Latent distance model:	discrete, 2 values; discrete, 2 values near 0 and 1 respectively; discrete.
Lord's model:	discrete, 2 values; normal ogive; discrete.
Latent profile model:	discrete > values *or* continuous; *any* alternative, discrete.

Deterministic models for comparative response data

Here we are concerned with the *general unfolding model* developed by Coombs (1950). In its simplest form this may be expressed succintly by stating that if stimulus g is preferred to stimulus k by subject i, then this implies

$$s(g) - s(i) < s(k) - s(i) \text{ or } d_{ig} < d_{ik}$$

Both attributes and individuals can be placed on the same scale. On the basis of individual rankings of all or at least a number of the stimuli it is

possible to place both the individuals and the attributes along this scale. Work is also in progress here with respect to multidimensional extensions.

This fairly extensive coverage of various response techniques is felt by the author to be justified in order to establish at an early stage just what methods are available and what it is possible to achieve with the aid of these techniques. This review will prove essential when a subsequent study is made of the extent to which these methods can be, should be and have been applied in the field of job analysis.

A general structure for scaling techniques

Many of the previously mentioned rating techniques have been fitted into a general structure by Björkman and Ekman (1957) in which they use Coombs' theory of data. We may distinguish between four types of measurement; the terminology in this work is used in the first instance, while Coombs' terminology is mentioned in brackets:

	Categorical (Irrelative)	Comparative (Relative)
subject-centered (Task A)	I	II
stimulus-centered (Task B)	III	IV
response-centered	V	VI

To the original fourfold table (cf. Björkman and Ekman, 1957) have been added squares V and VI. As has been mentioned previously, the subject-centered approach lies outwith the scope of the present work, and squares I and II will not be enlarged upon

The stimulus-centered approach makes use of scaling techniques at two different levels, i. e. first and second order. Let us begin by assuming that four stimuli have scale positions R_a, R_b, R_c and R_d.

First order ratings may be categorical as in $R_a = c$, where c is a constant, or comparative as in $R_a = R_b$ or $R_a > R_b$. Similarly the ratings can be second order categorical as in $R_a - R_b = c$, or comparative as in $R_a - R_b = R_c - R_d$. Squares III and IV can thus be expanded further and examples given for each type of rating technique applied:

	Categorical	Comparative
First order	Rating $R_a = c$ (Single stimulus rating or sorting)	$R_a > R_b$ (paired comparisons or ranking)
Second order	$\frac{R_a}{R_b} = c$ (fractionation methods) $D_{ab} = c$ (Multidimensional distance measures)	$R_a - R_b = R_b - R_c$ (equal-appearing intervals) $D_{ab} > D_{cd}$ (Multidimensional tetrads) $D_{ab} > D_{bc}$ (Multidimensional triads)

In a similar manner a fourfold table can be derived for squares V and VI, although it will be of a different nature:

	Categorical	Comparative
Deterministic	Guttman scaling	Coombs unfolding technique
Probabilistic	Latent class analysis	?

It would, at least theoretically, be possible to further expand upon this table by introducing first and second order measurements for at least some of these squares. For reasons of parsimony in data processing this does not appear to be justified at present. Conversely, it might be asked why the distinction made between deterministic and probabilistic measurement is made at the response-centered level, but not at the stimulus-centered one. Theoretically there would not appear to be any difficulty in introducing this further complication, but this is not in fact usually done. One possible explanation for this is that the very nature of probability seems to involve variation on the part of the *rater*; stimulus-centered ratings can thus conveniently be regarded as in the first instance *deterministic.*

These various distinctions between different scaling techniques will be studied in greater detail in chapter 3.

Methods of data reduction

By this is meant the methods whereby a large array or matrix of data is grouped together into units such as factors or profiles, in a way which increases the insight into the data obtained or simplifies the description. The three methods which are relevant in the present study are factor analysis, multidimensional scaling and latent profile analysis.

Factor analysis has dominated (or in the opinion of some researchers plagued) the art of variable reduction for half a century. The basic techniques involved will not be referred to here; a modern account is provided in Harman (1967). A necessary prerequisite for the application of factor analysis is that the multivariate distributions involved are normally distributed, that the variables are linearly related and that measurement takes place at least at the interval level. The basic problem in the behavioural and social sciences where factor analysis has its widest field of application is that *these three requirements are rarely satisfied.* One method of solving this problem is to ignore it; there is a myriad of factor analytic studies where there is no discussion as to whether or not the data satisfy the above rather strict requirements and often there is strong *prima facie* evidence that this is not in fact the case.

Further problems connected with the application of factor analysis are well-known and widely documented (cf. Harman, 1967). Question-marks concern "the problem of communality" and whether or not iterative procedures should be applied to estimate the values in the diagonal; the arbitrariness of the method of determining the number of significant factors; the concept of simple structure and the arbitrariness of the orthogonal or oblique transformations required to achieve this, and finally the interpretation of the factors obtained.

Factor analysis has, of course, traditionally been a sitting duck for criticism of this type, but, if we are to believe a leading expert such as Harman (1968), work in the last decade has gradually contributed to reducing the arbitrariness of some of the traditional manipulations. One factor contributing to the increased respectability of the method is the advent of high speed computers. This has removed the drudgery of computation and made possible many new types of exploratory studies which, by virtue of the quantities of data involved, were previously impracticable.

Factor analysis was developed by Spearman primarily in order to illustrate his one-factor theory of intelligence and to show that all other tests would be linear functions of this factor. Later developments by Thurstone do not, to be sure, have this restriction, but they still continue to assume that the factors are linear functions of test scores.

Latent profile analysis, the mathematics of which has been presented on pages 16–18, shows many similarities to factor analysis, without the above list of restrictive encumbrances. Green (1952) has in fact shown that factor analysis is a *special case* of latent structure analysis. Lazarsfeld (1959) points out that this observation leads the way to non-linear factor analysis, non-linearity being the main distinguishing feature of latent structure and latent profile analysis. It is this reduction of the requirements with respect to the data which is one of the primary strengths of LPA when compared to factor analysis.

A final technique which has certain advantages, and which has been touched upon in previous sections, is that of *multidimensional scaling.* As Schultz and Siegel (1964) point out, it is often very difficult *a priori* to

determine which factors, dimensions or variables are liable to be involved in a job analysis. Multidimensional scaling presupposes much less than factor analysis. In the latter case the variables to be rated are determined in advance, whereas in the former case the dimensions are produced by the raters themselves. Once again a more detailed examination of these matters will be reserved for chapters 2 and 3.

Reliability

It has been common to distinguish between the reliability and the validity of a measurement technique. Reliability refers to the extent to which a measurement technique produces the same results when the measurement procedure is repeated, or the extent to which different people applying the same measurement procedure at the same time arrive at the same results. A distinction can be made between *inter-rater reliability* where the reliability measurement refers to the agreement between two different raters using the same measurement, and *intra-rater reliability* which refers to two different measurements or ratings made by the same rater at different times. The measurement of reliability can normally be defined as the correlation between the two different measures of the same property. The reader is referred to a standard source such as Guilford (1954) for a more detailed account of the mathematics of measuring reliability.

Validity has been defined briefly as the extent to which a measurement measures what it *intends* to measure. In the theory of mental testing a careful distinction has been made between reliability and validity but recently a number of researchers have begun to feel that the distinction in some cases is an artificial one. This point will be elaborated upon in the following section.

Validity

According to Magnusson (1967), the validity of a method is the accuracy with which meaningful and relevant measurements can be made when applying it, in the sense that the measurement method actually measures the properties it was intended to measure. The question arising is of course, what is the property it was intended to measure? In the theory of mental testing, including industrial psychology, a number of techniques have been devised in order to validate a measurement procedure. *Predictive validity* means that the measurement is validated using a subsequent criterion measurement, while *concurrent validity* means that two measures of the same trait or property are available simultaneously and can be compared. *Content validity*, a somewhat nebulous concept which cannot be expressed in a simple coefficient, means that an attempt is made to estimate the extent to which a specific measurement adequately represents some area of interest.

A fourth type of validity, much used in clinical psychology, is that of *construct validity*. This concept is applied in situations where it is more or less impossible to obtain any satisfactory external criteria for the correctness

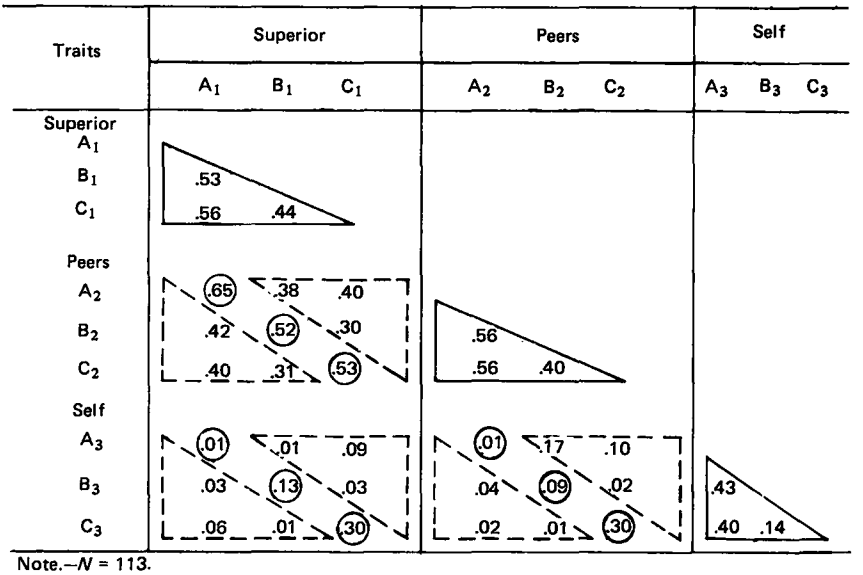
of a measurement. (It is, for instance, impossible to obtain a criterion measure for a measurement of latent aggression in an individual where the results have been obtained from a projective test.) The only method of validating measurements of this type is to attempt to study the extent to which *predictions* based on the measurements can be confirmed. Validity of this type cannot be expressed in a single coefficient either. It is in fact an application of the classical hypothetical-deductive system of investigation. Methods of investigating construct validity (cf. Cronbach and Meehl, 1955) include the following:

- the study of the extent to which individuals or groups who, according to the theory and measurement, should differ in some respect do in fact differ.
- the extent to which individual and external changes give rise to individual measurements in accordance with the theory.
- the correlation between measures which according to the theory should measure the same variable.

Considerable caution must, however, be shown in the interpretation of such coefficients as there is a tendency for them to be indistinguishable from simple reliability measures.

A further type of validation which has aroused some interest is *synthetic validity*, a term coined by Lawshe (1952). The basic idea underlying this type of validity is an attempt to analyse the system or order underlying the actual relationships between predictors and criteria. As we shall subsequently see this method has been applied in a number of job analysis and selection studies.

An attempt which succeeds in clarifying what is sometimes only a tenuous distinction between reliability and validity has been presented by Campbell and Fiske (1959). This classic article contains an analysis of construct validity by means of the *multitrait/multimethod schedule.* One application of this technique involves changing from a multitrait-multimethod approach to that of a *multitrait/multirater approach.* Basically, a distinction is made between *convergent validity*, i. e. the agreement between measurements of the same type using different methods or different rater categories, and *discriminant validity,* which is said to be in evidence when the measurement of a variable correlates more closely with measurements of the same type carried out by some other rater, than with a measurement of some other type carried out by the same rater. In addition, discriminant validity requires that the validity coefficient for a given variable must be greater than the correlations between the measurements of this variable and the measurements of all other variables by any other rater. This can be clarified somewhat with the aid of the following figure from an empirical study.



Note.—N = 113.

Fig. 1:3 The multitrait/multirater design (from Lawler, 1967)

Convergent validity arises when there is good agreement between measurements of the same type by different raters, i. e. the figures shown within the rings. The traits rated in figure 1:3 were quality of job performance, ability to perform the job, and effort exerted on the job. The figures within the rings show that the superior and average peer ratings have good convergent validity. The remaining figures for correlations between each pair of self, peer and superior ratings do not, however, show any signs of convergent validity and it is in fact a matter of some doubt as to how these figures should be interpreted.

With respect to *discriminant validity* the following more specific requirements should be met.

"*First,* the validity diagonal correlation value should be higher than the values lying in its column and row in the heterotrait-heterorater triangles. That is, a trait measure should correlate more highly with another measure of the same trait than with any other variable having neither trait nor rater in common.

Second, a trait measure should correlate higher with an independent effort to measure the same trait than with measures designed to get at different traits which employ the same rater. For a given variable, this involves comparing its values in the validity diagonal with its values in the heterotrait-monorater triangles.

Third, it is desirable for the same pattern of trait interrelationships to be shown in all of the heterotrait-triangles of both the monorater and the heterorater blocks." (Lawler, 1967).

The data in the figure for superior and peer ratings meet the first two requirements for discriminant validity and almost meet the third, although this is an extremely stringent requirement seldom met in the behavioural

sciences. Divergent validity requirements are not met by the self ratings together with superiors and peers. With respect to the formal requirements for coefficient values to be significant, a number of measures have been proposed by Evans (1969).

The terminology originally introduced by Campbell and Fiske has been presented at some length as it will have considerable bearing on the examination of empirical work in the field of job analysis and selection, not least with respect to the question of criterion data. This matter will be returned to in chapter 5.

Reverting to the previous point, concerning the relativity of the concepts of reliability and validity, it will be seen that convergent validity in the multirater approach is in fact the same as inter-rater reliability. In general, the two concepts have primarily been developed in connection with the theory of mental testing, i. e. the subject-centered approach, whereas the distinction tends to break down in the stimulus-centered and response-centered approaches. This will be expanded upon with the aid of the following table:

	Variation in	
Approach	Subject	Stimulus
subject-centered	Yes	No
stimulus-centered	No	Yes
response-centered	Yes	Yes

In the subject-centered approach the only variation is assumed to be among subjects. The concept of reliability thus revolves around the stability of the *stimulus,* i. e. the extent to which the measurement stimulus will provide the same results for each individual subject reacting to it. Validity in this situation thus refers to the extent to which reliable stimuli and the variations in subjects on the basis of these stimuli can be used for practical purposes, for instance in assigning people to jobs or courses of study.

What significance do these two basic concepts have in the stimulus-centered and response-centered cases?

In the former case the situation is reversed, in that it is the stimulus and not the subject which is liable to variation, whereas in the latter both subject and stimulus can be subject to variation. In the stimulus-centered case it is instructive to define reliability and validity exactly analogously to the subject-centered approach. Reliability in this case would thus refer to the stability of the *subject,* i. e. the extent to which one and the same subject would tend to rate a series of stimuli in the same way on repeating the rating. Validity would then refer to the extent to which the ratings of stimuli by reliably rating subjects can be used for practical purposes, such as constructing an attitude scale. The concepts of reliability and validity become even more complex in the case of the response-centered approach. Once again

continuing the increasingly nebulous analogy of the subject-centered approach it would be possible to talk of *subject-reliability* and *stimulus-reliability*. It is something approaching this distinction which Campbell and Fiske are trying to get at when they specify the requirements for *discriminant validity*.

A definition of validity in the response-centered case does not appear to be particularly meaningful at the present stage, other than in the most general terms of the measurement finding a meaningful application. This and other points arising from the concepts of reliability and validity in job analysis will however be returned to repeatedly in subsequent chapters.

Sampling techniques and representativity

This and the two final measurement scales require only the briefest of comments. Along with many others, Boalt (1965) has stressed that any study of measurement and research must carefully examine the sampling techniques used in order to obtain the basic data and the checks which have been made, to ensure that the area in question is representative of a wider area with respect to which conclusions may be drawn.

The value of the research as a means of improving theory construction

The object of research is the advancement of knowledge. Advances can normally only be made if there is some substantive structure behind individual research projects. It is reasonable to assess research work with respect to the extent to which this substantive structure (theory or model) is illuminated by each piece of research.

The usefulness of the research to society

Suffice to say at this stage that it is this criterion which is perhaps most often neglected by researchers whose primary goal is unfortunately not always that of the advancement of knowledge.

This chapter contains a broad general presentation of certain measurement areas, scales or criteria. They will be used for two separate purposes in later chapters; firstly to analyse the research work performed by others within the field of job analysis, and secondly the first seven measurement scales will be used as a loose framework for a more general discussion of the problems arising in connection with job analysis and the extent to which they can be solved in various specific situations, including the selection procedure.

Chapter 2 Job analysis – terminology and a review of the literature

Terminology

A study of the literature in the field suggests that terminology is by no means uniform. For instance, careful distinctions are seldom made between analyses, descriptions, specifications and evaluations, or between descriptions which are work-oriented and those which are worker-oriented. The conceptual confusion would appear to have arisen partly because different researchers and practitioners have had greatly varying approaches and purposes when dealing with the problems of job analysis.

It is suitable to begin by clarifying the concept of "work". Following Shartle (1952), a widely accepted distinction is that between the three levels of

position – job – occupation.

A *position* consists of a collection of duties which are performed by a single individual, for example the position of being marketing manager of a given company. There are as many positions as there are (full-time) employees in a company.

A *job* consists of a group of closely related positions, where the duties are more or less the same for different incumbents, i. e. the job of typists or of manual lathe operators.

An *occupation* consists of a group of closely related jobs, each of which has a number of basic features in common, for instance the occupation of metal workers as opposed to that of building workers.

The distinction between a position and a job is more concrete than that between a job and an occupation, but the classification has nevertheless proved to be a worthwhile one and will be applied in this study.

It will be necessary to make a further terminological distinction at the micro-level, namely between *task, behaviour* and *trait*:

A *task* is a duty to be performed by someone in a given position. This task can normally be expressed in concrete terms without reference being made to the person performing the task. A task may thus consist of "distributing letters within a department", "designing a specific factory" or "selling shoes within a specific department". The task may therefore be a rather extensive one, encompassing almost the whole of a position, or a very limited one.

A *behaviour* is the way in which a task or position will be performed by a job incumbent. The distinction between this and a task is that the behaviour obviously refers to *how* the task is performed and not *what* is performed. A position described in behavioural terms would consist of expressions such as

"is pleasant to the customers", "shows understanding of the problems of others", "is prepared to help workmates who are lagging behind".

A *trait* is normally a psychologically measurable variable which is used to describe the job incumbent. "Verbal ability", "the capacity to work with others", "finger dexterity" are all examples of traits.

The subsequent study of the research on job analysis will show that a clear distinction is not always maintained between these three concepts. This has meant that many job analysis results are difficult to interpret and to apply.

The confusion with respect to terminology is illustrated in the two standard works by Ghiselli and Brown (1948) and Thorndike (1949). These will not be elaborated upon in detail here but it should be noted that Thorndike appears to be primarily interested in an initial *job description*, which may then be used to analyse, in the case of selection and recruitment, *trait requirements*. Ghiselli & Brown consider on the other hand that an initial *job analysis* can be used to give rise to a *job description* and *specification*, in other words a diametrically opposed terminology.

The concepts of *analysis* and *description* have also been the subject of general discussion. A number of writers clearly consider that the term *"job analysis"* may be used to denote any attempt whatsoever to describe a position, job or occupation. Other writers, especially psychologists concerned with recruitment and vocational guidance, maintain on the other hand that the actual analysis stage must be preceded by a *description* aiming at providing a sufficiently exhaustive and objective information base concerning the position – job – occupation, which may then be used for a number of different specialized analyses.

The following suggestion (to be applied throughout this study) is made as to *terminology*: any job analysis should commence with as objective a job description as possible. This *job description* will normally be expressed in terms of *tasks* involved. It will of course be objected that there is no such thing as an objective description of anything, least of all the *tasks* involved in any given job or position. This is of course quite true, but it is the contention of the author that it is more objective to describe a job in terms of the *tasks* involved than to express it in terms of the *behaviour* manifested by a successful or unsuccessful job encumbent, or in terms of the psychological *traits* which are required in order to succeed in the job in question.

This basic task job description can then be used for a number of different possible *analyses*, where the degree of objectivity becomes considerably less. One possible application would, for instance, be to express a job in *behavioural* terms, which could then be used for *training* or *personnel appraisal* purposes, another would be to express the job in terms of *trait* requirements which could then be used in connection with psychological selection techniques.

This distinction raises the whole issue of *work-oriented* or *worker-oriented* studies. There will be occasion to return to this question subsequently; for the present it will be apparent that what has here been called a task job description will correspond to a work-oriented study, a trait requirement analysis will be worker-oriented, while a description in behavioral terms will

fall somewhere in between.

This attempt to clarify the concepts involved should not conceal the fact that the very nature of concepts such as *job* and *worker* are *relative* ones. Jobs do not always have exactly specified duties and changes in job requirements will tend to be an organic process which is partly dependent on the surrounding jobs and their job incumbents. Similarly, the worker or employee will tend to show a varying capacity for performing the duties in a satisfactory way, both in the short term and in the long run. Furthermore, none but the staunchest supporters of Taylor will maintain that there is only the "one best way to do a job". In this connection mention can be made of the systems theory launched by Buckley (1967). Applying his *morphogenetic model* to job analysis we may regard a job as consisting of a complex, compound system in continual contact with the surroundings. Conflicts between different jobs and job incumbents may be expected to contribute to this process of change and adaptation. The morphogenetic model also gives rise to two other interesting aspects, the principles of *equifinality* and *multifinality*. The principle of equifinality states that the final state of a system, for instance that of a job and a job incumbent, may arise in a number of different ways, i. e. from the same initial state but following different development paths or from different initial states. The principle of multifinality states that a single initial state of a system may have many different final states. Applying this dynamic systems approach to job analysis and the relativity of job and worker, it will be seen that there may be many different combinations of jobs, positions and job incumbents which may lead to satisfactory (or unsatisfactory) outcomes for the system as a whole.

In order to further clarify the use of the above *terminology* a very brief account will be given of some standard applications of job analysis.

Applications

Personnel appraisal. The job description and subsequent analysis can be used to produce a job specification stating performance requirements for the incumbent in behavioural terms. A comparison between these performance requirements and actual performance can then be used to provide a basis for reviewing requirements, salary increases, promotion, and relocation of employees within the company.

Organisational studies. It is unusual for employees to work entirely independently and it is in fact more normal for them to have contact with other employees and outsiders. Task job descriptions which include the interactions between different employees may provide a basis for decisions concerning control measures, planning, communications, division of labour etc.

Job evaluation. American studies suggest that job analyses are primarily used for this purpose. (Jones and de Cotiis, 1969) The factors included in the current job evaluation systems applied in Sweden consist of an interesting

combination of task job description and analyses in behavioural and trait terms. As the object of job evaluation is to establish the *difficulty* of the job for salary and other purposes, this mixture provides difficulties of interpretation as will be shown in chapter 3.

Recruitment and selection. A position or job description provides a basis for an analysis of trait requirements. This trait analysis is normally expressed in psychologically measurable variables which are then included in a selection test battery.

Later *criterion* data which may be expressed in *task, trait* or *behavioural* terms, see chapter 5, provide an idea as to how well the analysis has succeeded in determining the most important job requirements. It is this application to which most attention will be devoted in the present study.

Training and educational purposes. Job descriptions may be used as a basis for qualification analyses expressed in *trait* or *behavioural* terms. When there is a shortage of labour which is directly suitable for the duties in question, these requirements may then be related to available applicants in order to determine the training measures required for the applicants to become suitable. Detailed job descriptions may also simplify the production of training manuals and other on-the-job instructional materials, as well as perhaps improving tools, equipment and routines.

Vocational guidance. This involves not so much position or job descriptions as *occupational descriptions* and analyses which in *general* trait terms state the requirements involved for different occupations. The trait profile of the person seeking guidance which may be determined by psychological testing or otherwise is related to different vocational requirements. The guidance is then based on a matching of individual traits and occupational requirements.

Biotechnology. In recent years increasing attention has been devoted to analysing duties from the point of view of increasing job satisfaction, safety and in general in improving the working environment. This area requires specialists from many fields such as technology, medicine and psychology.

Methods for data collection

The problems of carrying out task job descriptions and of applying job analyses to the determination of trait or behaviour requirements has been the subject of very extensive study, the majority of which has not however contained any advanced methodological discussion. An even more striking defect is the absence of financial considerations, which are of course a vital factor in practice.

The basic methods of collecting data in the behavioral sciences are limited. Information can be obtained by introspection, observation or by asking other people directly or indirectly about the area in question. The classic *introspective* method in job analysis was for the job analyst himself to perform the job and to draw conclusions as to the tasks, behaviours and traits

required. Although Gilbreth and others no doubt derived considerable regressive pleasure from driving tram cars in the early days of industrial psychology, the method is scarcely a serious practical alternative, bearing in mind the cost, the intuitively low reliability and the limited fields of application. An even more serious criticism is that the job analyst who is a beginner at the job in question will not see the job and the problems involved in the same way as an experienced operator.

The method of *direct observation* including filming and sampling techniques has primarily been used in the field of work study. This technique tends to provide rather superficial information and is normally restricted to manual work. A further difficulty is that of determining the period of study necessary in order to establish the complete pattern of tasks in a work cycle.

Finally, people can be *interviewed* directly or indirectly with the aid of *questionnaires* and *checklists*. The deficiencies of the interview as a method of collecting data are well known and need not to be expanded upon here. Additional problems in connection with job analysis involve establishing who is to be interviewed and what questions are to be asked. A number of studies have shown that superiors, subordinates, peers and perhaps the outgoing job incumbent have widely varying cognitive interpretations as to the tasks, behaviours and traits involved in the job in question. The agreement is often low even *within* rater categories.

With respect to questions asked, it is often difficult to distinguish between task and behaviour descriptions on the one hand and trait assessments on the other. We shall have occasion to return to these matters in chapter 3.

Questionnaires and checklists are perhaps the methods which are applied most often in industry by consultants, organisations and personnel departments in Sweden at present. In the experience of the author there are literally hundreds of questionnaires and checklists used by companies and consultants in Sweden. An extensive selection is available at the Swedish Council for Personnel Administration, where the author has performed some of the research presented in the present chapter. Many of these questionnaires involve only minor changes from American prototypes and it is worth commenting briefly on certain recurring features.

The above-mentioned linguistic and terminological confusion is readily apparent, many questionnaires consisting of combinations of questions about tasks, behaviours and traits. Questionnaires often concentrate upon determining the total time distribution for different tasks, with the incorrect underlying assumption that this will in some way provide a ranking of the difficulty or importance of these tasks.

Summing up, many of the checklists and questionnaires at present in use are rather unsatisfactory. They involve a mixture of "objective" task description and "subjective" trait or behaviour assessment, the terminology is vague ("is in charge of", "is in contact with", "general supervision") and interrater reliability studies have provided unsatisfactory results. The application of modern organisation terminology such as decision making, search behaviour and control mechanisms is conspicuous by its absence.

Structuring job analysis data

The question of data collection techniques is only part of the complex problem of structuring data. The questionnaire methods mentioned above often involve a considerable degree of structuring, whereby the responses obtained from the questionnaire will tend to be compiled question by question in accordance with the basic form of the questionnaire.

The traditional approach. The traditional method of performing a task job description using the terminology applied in this study is to divide the work into a number of independent aspects such as the following.

A. Job, name or title.
B. Present methods and procedures of work.
C. Physical conditions of the work environment.
D. Relation of each job to other jobs.
E. Conditions of employment.

These headings (and further sub-headings not elaborated upon here) provide opportunities for describing the job in detail. However, a number of question-marks remain concerning the use of this information in practice. Methods and procedures employed in carrying out tasks are of central importance, but no specifications are provided as to how this information should be obtained.

The systems approach. Primarily thanks to the pioneer work of Barnard (1938) and Simon (1947), interest in modern organisation theory has to an increasing extent been devoted to decision making, information systems, control systems and search behaviour.

Using the systems approach, a position or job may be regarded as the basic unit to be studied. This position will involve information and material being treated, processed and passed on in accordance with certain guiding impulses and information. Only modest attempts have been made to describe jobs in these terms in connection with job analysis, but it seems likely that systems thinking will play an increasing part in the future, not only with respect to administrative rationalisation and performance measurements, but also with respect to job and training requirement analysis.

This may be illustrated with the aid of the following terminology: When rating trait or behaviour requirements for a given position, a considerably greater amount of information is obtained, if it is known that the duties are *programme controlled,* i. e. the job incumbent has only to place certain figures in the programme in order to obtain information as to which course of action is to be selected, than it is to be told that the job incumbent "is in charge of" or "is responsible for" this duty. A systems theory description also includes contacts with other job incumbents, i.e. whence and whither guiding impulses and communications come and go, to whom information or material is to be supplied, the manner in which new impulses are generated, etc. This approach also provides information as to which tasks and positions are most demanding (in general one might for instance say that *goal controlled* duties

place greater intellectual demands on the job incumbent than *directly controlled* duties). A pioneering attempt to develop these ideas has been made by Gagné (1962), who discusses the human factor in systems, especially man-machine systems, and arrives at the conclusion that human functions consist of *sensing, identifying* and *interpreting* various types of signals. Although this terminology has been little used in connection with industrial psychology, it is readily apparent that it can be used to provide a more exact description of the tasks involved in different jobs. In a further contribution to the same book, Miller (1962) states that each *duty* or *task activity* consists of "an indicator on which the activity-relevant indication appears, the indication or cue which calls for a response, the control object to be activated, the activation or manipulation to be made, the indication of response adequacy or feedback". Descriptions in system terms would appear to improve stringency in industrial psychology, if only at the micro-descriptive level. Preliminary attempts to apply this approach to job analysis have been made by some of the author's graduate students, cf. Almroth and Jonsson (1971), Benson and Hellberg (1970).

The critical incident method. This method was proposed in the early 1950:s by Flanagan (1949) and has also been applied to some extent in Sweden by, *inter alios,* Andersson and Nilsson (1966). The method applied is to collect information about tasks or behaviours in the work situation and the manner of reacting towards them which is to be regarded as "critical" in the positive or negative sense. The respondents are asked to provide a list of such critical situations and to suggest positive and/or negative reactions on the part of the job incumbent. These critical incidents and the positive and negative reactions towards them are then made the subject of a systematic classification. (This approach has influenced Herzberg, 1959, in his work on two-factor motivation theory.)

The method has a number of obvious advantages in connection with training analyses, trait requirements analyses and perhaps even as a basis for performance appraisal of individuals. The basic problem with the critical incident method is the classification and codification of the incidents. There are also practical difficulties involved in using these classes of incidents as a basis for establishing job trait requirements in a selection situation.

Differences between rater categories. A general difficulty mentioned on page 33 and arising in all data collection and processing is that different categories of respondents will tend to provide different responses. We may expect considerable differences in impressions obtained, irrespective of the data collection in question, when subordinates, peers, superiors and present job incumbents provide information regarding present duties as well as their own interpretation of the required traits and behaviors (cf. Mårdberg, 1971d and Triandis, 1959). It is well-known that present job incumbents tend to exaggerate the extent of their tasks and the demands involved. Superiors may underestimate the work performed by subordinates or in a recruitment situation exaggerate the difficulty of subordinates tasks in order to obtain

overqualified subordinates, thereby raising their own status. Peers may tend to underestimate the tasks, if there is an element of competition in the situation in question, or overestimate them if their own tasks are entirely comparable. Although it is unusual in practice for subordinates to provide job description or job analysis data, we may expect a similar process to occur here; *anecdotally,* few people are impressed by the abilities of and the demands placed on their immediate superiors. In other cases the duties of the superior may be overestimated through lack of knowledge of the total hierarchical structure of the decision-making process within the company. The words "overestimate" and "underestimate" have been deliberately abused in the present context, as they assume that it is possible for some person within or outside the system to make an objective assessment of the duties and difficulties involved. This is, of course, not always necessarily the case. Limited research has been conducted in Sweden as to the reliability of job descriptions. The types of reliability measure of greatest interest would be inter-rater agreement *within* the category of, for instance, superiors, and inter-rater agreement *between* categories, such as present job incumbents and superiors. Research reported from the United States suggests that in some circumstances the reliability is satisfactorily high, whereas in others it is so low that it would tend to nullify any attempts made to use the description as a basis for an analysis of job requirements. The multitrait-multirater approach presented in chapter 1 is of some relevance here and we shall have occasion to return to these matters in chapters 3 and 5.

A selective study of the literature on job analysis

Elsewhere (cf. Mabon 1968a, 1968b) the author has provided a fairly extensive review of this literature. In this study the review will be restricted to work which is regarded as having an important bearing on methodological and terminological advances and which will primarily be used for the analysis in subsequent chapters.

Studies made by the US Department of Defence and the US Labor Office and Employment Service

a) A large number of the studies which will subsequently be quoted have been performed under the auspices of the American Department of Defence. Studies of this type, which will be reported on in detail are American Institute for Research (1951), Gragg (1962), Kershner (1955), Morsh, Madden and Christal (1961), Morsh (1962), Norris (1956), Office of Naval Research (1959), Rupe (1955, 1956) and Rupe and Westen (1955).

b) The work by the *United States Labor Office* in producing a Dictionary of Occupational Titles which has been used as source material for a considerable range of studies is also worthy of note. This *Dictionary of Occupational Titles* (DOT) (1949) now comprises some 30.000 jobs and contains brief descriptions of 13.000 of them. Studdiford (1953) suggested a new occupational classification structure whereby the job requirements for any

job could be expressed in terms of *aptitudes, interests, temperaments, physical capacities, working conditions* and *training time.* Estimates of worker trait requirements for a representative sample of 4.000 jobs defined in DOT were then developed by the *United States Employment Service* (1956). Prior to this study, Trattner, Fine and Kubis (1955) had compared worker requirements based on the job descriptions and by direct observation. Fine (1957) compared the USES occupational classification with the Minnesota Occupational Rating Scale and drew the conclusion that continued research was necessary in order to determine the best method of classifying jobs. Fine and Heinz (1957) also presented a brief review and summary of the previously mentioned report on estimates of worker trait requirements for 4.000 jobs. A major study by McCormick, Finn and Scheips (1957) will be dealt with in detail subsequently. A study by Newman and Fine (1957) showed that the USES ratings of worker trait requirement for the 4.000 jobs were satisfactory when it came to physical capacities and work condition information. Boling and Fine (1959) studied the cues used by raters in the rating of temperament requirements of jobs. Mosel, Fine and Boling (1960) also studied whether or not traits such as *verbal ability* and *motor reaction* time were in fact unidimensional. The *interest* requirements appeared to be quantifiable, about half of the *temperament* requirements and only a few of the *aptitude* requirements. This study appears to question the findings of much of the previous work in this field.

A study using the 4.000 jobs and the aptitude trait was that performed by Orr (1960). This was however a methodological study aimed primarily at comparing two multivariate analysis techniques.

The work conducted on and around the DOT and the estimates of worker trait requirements suggest that the DOT is in many ways an excellent method of indirectly obtaining useful information about job content and trait requirements. The results obtained by McCormick, Finn and Scheips (1957) show for instance that the range of trait patterns is considerably smaller than might have been imagined.

General introductions

Little need be said of the general introductions to the subject; these tend to be fairly repetitive and an analysis of them would serve primarily to point out the wide range of usage with respect to the terms involved. The comparison of approaches in the standard works of Thorndike (1949) and Ghiselli and Brown (1968) presented on page 30 may be taken to exemplify this confusion.

Job analysis of a specific job or occupation

Many of the studies involving analysis of a given job or occupation are intended as practical aids in specific situations where scientific requirements are not always high. This type of study will not be examined in this report.

However, one of the better examples of this type will be mentioned: Christensen (1950), in which a sampling technique of activities is applied to analyze the job of navigator. The method used was a type of *time sampling* which provided information as to the order in which various tasks were performed, how often they were performed, and the proportion of total time devoted to each task. A study of a similar type has been performed in Sweden by Wirdenius (1958).

Comparisons between different methods of data collection

A discussion has been presented on page 32 as to the relative advantages of different methods of data collection. A number of studies have been performed which throw more detailed light upon this question.

Das (1960) compared worker analysis ratings based on job descriptions on the one hand and motion-time study on the other. This is in fact a rather unusual comparison, although the possibilities should be apparent of using motion-time study in order to determine some of the more elementary aspects of how work can be performed. Three observers studied a number of jobs within the textile industry, using job descriptions and MTM methods. The rater agreement *between* the methods was 79 %, whereas the rating agreement *within* each method was 77 % for MTM and 78 % for the job description method. The author concludes that the majority of the tasks performed in the jobs could be identified by either method.

Newman and Fine (1957) studied the validity of job descriptions for *physical requirements* and *work condition information.* This study was performed within the framework of the job descriptions provided by the United States Employment Service. A comparison was made between *indirect* and *direct* methods concerning information within the two areas in question. The direct methods were performed by making ratings on the basis of studying the jobs in question directly, while the indirect ratings were performed on the basis of the written descriptions provided. The physical requirements consisted of 6 factors and the working conditions of 7 factors. Somewhat contrary to expectation it was shown that the indirect methods were just as good as the direct ones when it came to assessing these types of factors. The conclusion was drawn that the USES descriptions of 4.000 jobs were as satisfactory a basis for performing job analyses as direct observations.

A further blow to the direct observation technique was provided by O'Neil and Kubany (1959). This technique was used to study 85 supervisory positions. Four subjective and four objective criteria were selected but no significant discrimination was obtained after crossvalidation. In addition the information obtained was not of sufficient value to justify the direct observation method when compared with the cheaper methods of interviewing and questionnaires. It was concluded that the method of direct observation was *not* the most suitable one for measuring supervisory behaviour.

A comparison of a different kind was performed by Porter (1961), in which 64 supervisors and 74 middle managers used *paired comparisons* to

rank 13 personality traits with respect to their importance for success in supervisory work. A high correlation was obtained between the two sets of rankings. Both groups, especially the supervisors, stressed the importance of cooperation. This result tends to contradict the studies performed in Sweden by Mårdberg (1971d) in which it was shown that different categories of employees had widely different views as to the duties and rules governing specific jobs.

The most comprehensive series of comparative studies was by Rupe (1955, 1956) and Rupe and Westen (1956). This was a series of studies in the United States Air Force in order to determine the most satisfactory method of performing job analyses for 4 different types of occupational categories within the Air Force.

The 5 methods applied were the *questionnaire, group interview, individual interview, observation interview* and a so-called *technical conference.* The methods were evaluated in terms of the job elements identified by expert judges. The questionnaire and group interview method provided less information about job elements than the three other methods. Almost the same quantity of information was obtained from the three other methods, but in terms of time devoted to the method in question the individual inerview was best when it came to senior appointments.

The same comparison of methods was made in a later study, this time with respect to four technically complex jobs. The previous findings were confirmed, namely that the questionnaire and the group interview gave less information. The questionnaire was, however, the cheapest method. The quantity of information as well as the cost in man-hours have the following ranking: technical conference, observation interview and individual interview. The differences were very small.

A summing up of the series of studies was presented in a final report. The results were as follows:

- no significant differences could be determined among the methods with respect to reporting on tools, materials and equipment.
- the group interview is the least efficient when it comes to obtaining information about job elements (i. e. sub-tasks), but it is the second cheapest with respect to man-hours involved.
- the questionnaire is the least reliable method. It is ranked second last when it comes to obtaining information about job elements, but is at the same time the cheapest method.
- technical conferences and the observation interview are almost equal with respect to job elements and are better than the questionnaire and group interview. They are however the two most expensive methods for data collection.
- the individual interview is most reliable, most efficient and average with respect to costs.
- in a military study of this type, differences with respect to the efficiency of the analysts are greater than the differences among the efficiencies of the methods.

There are a number of features of this study which are worth stressing. Firstly, Rupe is one of the few researchers to take into account *financial considerations*. Most other studies confine themselves to expressing differences in interrater reliability between and within methods, while Rupe's studies make it possible to establish more or less the reliability which can be attained, if a given sum of money is made available to the job analyst.

The final finding of Rupe should serve as a reminder that in the last instance all ratings are performed by *individuals* and not by rating scales, as well as in specific situations for specific workers.

A further study aimed at investigating the quality of the information provided by the USES was performed by Trattner, Fine and Kubis (1955). In this study worker requirement ratings made by reading job descriptions were compared with direct job information. Worker requirements were subdivided into four main areas by the USES, namely *temperament, interests, aptitude* and *physical capacities*. The reliability and validity of various ratings were compared when direct and indirect methods were applied. The criterion measures were obtained by testing 60 workers on each job. The results showed that reliability was higher for the indirect method; that the rating of *essential* factors was more consistent for the direct method and that validity was higher for the direct method. Validity was however high for both methods when it came to psychological and perceptual factors, contrary to the case for physical factors. The study is interesting in itself but would of course be of considerably greater value to a decision maker if information as to the *costs* of the two alternatives were available.

Multivariate analysis of job descriptions and traits

The basic aim of studies of this type has been to determine a limited number of dimensions, along which a job may be described or to determine a number of basic traits which are important for different kinds of jobs and positions. It is the application of studies of this type which gives rise to difficulties in distinguishing between the work-oriented task descriptions and worker-oriented trait descriptions. These were briefly dealt with in the previous section on terminology. Many studies appear to be work-oriented, i. e. they aim at establishing a limited number of basic tasks which are important in different kinds of work. As has been made clear previously, this type of study is of less value than a worker-oriented trait study in the selection process, since a further study must be made of the basic work-oriented task dimensions in order to translate these into psychologically meaningful traits when determining selection instruments.

Almost without exception the method of analysis applied has been *factor analysis*. Latterly, however, other techniques such as *latent profile analysis* have come to be discussed and applied. This section will be confined to an account of research conducted within the area. The suitability of various multivariate and multidimensional techniques will be treated in chapter 3.

Work-oriented analyses (task analyses)

An example of a work-oriented factor analysis is provided by Chalupsky (1962). A sample of 192 office jobs were analyzed independently by means of a *knowledge check list* and a *task check list.* The intercorrelations between the items on both lists were then the subject of factor analysis, four factors being obtained which were common to both lists.

A similar study was performed by Coombs and Salter (1949). This was an attempt to study job families, i. e. at the occupational level, with the aid of factor analysis. A formula based on the number of elements which two variables had in common was used in order to compute the correlations between jobs. A sample of 20 jobs gave a first order general factor and 4 common factors.

A study was performed by Saunders (1956) in which 24 activity areas on a check list were completed by engineers from 23 companies and then factor analyzed. 5 factors were obtained and on the basis of these factors engineers could be divided into 7 categories.

A *cluster analysis* of office operations has been performed by Thomas (1952). A task description check list containing 139 items was completed for 112 office jobs and then cluster analyzed. Eight components were obtained.

A final example of a factor analysis which is work-oriented is provided by Brumback and Vincent (1970). This consisted of a study of work-performed data for a sample of administrative, professional and scientific positions in the United States Public Health Service. A total of 196 task descriptions were obtained from some 8.000 work statements extracted from 4.000 narrative job descriptions, and the 196 job descriptions made up a so-called position inventory. Each duty was then rated on a seven point scale with respect to the significance of the duty within the position as a whole. The main use of the study was for the development of a new officer rating instrument. The authors regard this as an exploratory study and it is not exactly clear how their results are to be used and interpreted.

Most work-oriented analyses have been made with respect to office and supervisory positions. The main use of these analyses appears to have been to classify different types of jobs and job families. The problem of selecting people to fill these various types of jobs requires of course a translation of the job description into a trait analysis.

Worker-oriented analyses (trait and behaviour analyses)

A considerable number of worker-oriented studies have also been performed. A factor study of worker characteristics performed by Jaspen (1949) on a sample of 275 jobs showed that 20 of the 45 items on the worker characteristics form appeared on more than 10 % of the jobs. Factor analysis reduced the number of factors to 6: strength, intelligence, inspection, physically unpleasant working conditions, manual skill and mechanical knowledge. (How physically unpleasant working conditions can be regarded as a worker characteristic is not enlarged upon.)

A further worker-oriented study is reported by Prien (1963). A behavioural checklist for supervisors was developed and data were obtained from 24 supervisors and 6 managers. A factor analysis was performed on the check list and 7 factors were obtained. A second factor analysis provided two factors, one on *initiating structure* and the other on *consideration.*

A study which has become a minor classic was performed by McCormick, Finn and Scheips (1957). 4.000 jobs from the Dictionary of Occupational Titles were described by the United States Employment Service primarily with respect to the following 6 areas: length of training time, aptitude, physical capacities, temperament, interests and working conditions. Five of these are thus worker-oriented and the final one, working conditions, work-oriented. These 6 headings contained a total of 44 specific variables. These variables were then factor analyzed and 7 factors were obtained. In addition it was possible to obtain the 4 specific variables which were best able to identify each of the factors. The factors were

- mental and educational development *versus* adaptation to routine
- adjustment to precision tasks
- body mobility
- artistic ability and æsthetic taste
- manual artistic ability
- personal contact ability *versus* adaptation to routine
- heavy physical work *versus* aptitude for office work.

It was then possible to compute a score on each factor for each of the 4.000 jobs. The factor values were then divided into two levels except for the first one (mental and educational development *versus* adaptation to routine) which was scored on three levels because of its great importance. It was thus possible to have 192 different work patterns (2^6 x 3).

When studying the 4.000 jobs it appeared that 12 patterns encompassed 60 % of the jobs and 33 patterns 88 %. Seventy-seven patterns contained no jobs whatsoever. The conclusion would appear to be that with only a few exceptions *most jobs fit into standard and readily identifiable trait patterns.* The study is a very interesting and carefully conducted one.

An attempt to determine dimensions of job success for mechanics was performed by McQuitty, Wrigley and Gaier (1954) and a similar study was performed by Norris (1956).

A worker-oriented check list of 177 items dealing with working duties was used by Palmer and McCormick (1961) in order to study 250 selected jobs within a large steel works which had task job descriptions for 5.000 jobs. The 250 jobs selected were described in terms of 14 primarily bipolar trait dimensions. Eight doctoral students then performed independent analyses of from 2 to 5 jobs which were then compared with the results obtained by the project leader. This gave a reliability measure, which was shown to be satisfactory. It was also shown that tasks could be identified and measured and that the number of different activity areas could be grouped with greater simplicity and economy in terms of a small number of comparatively independent dimensions. A higher factor analysis provided 4 general factors

of which the most important one was decision making, a result well in agreement with modern organization theory. As in a number of studies of this type there is unfortunately considerable movement back and forward between work and worker orientation..

An attempt to study the performance dimensions of supervisory positions was carried out by Peres (1962). 414 supervisors were asked to write essays about the best supervisor they knew and this was used to produce a check list of 303 items about supervisory behaviour. 372 supervisors then rated the best and the worst supervisor they knew on a five point scale for each of these items. A subsequent factor analysis produced 6 factors plus a general halo factor.

A further study to determine the requirements of Air Force job families is reported by Thorndike, Hagen, Orr and Rosner (1957). An activity (task) questionnaire was developed giving scores for 14 types of trait requirements considered to be comparatively independent. Each person studied reported how often he performed each of a number of specific tasks. Trait profiles were then obtained for 25 Air Force positions but the method was not entirely successful, primarily due to difficulties in selecting suitable items. For this reason the researchers are not optimistic about the future use of the method.

A study investigating whether or not ratings of this type are satisfactory measurements at all was performed by Mosel, Fine and Boling (1960). An attempt was made to study whether trait requirements such as verbal ability and motor reaction time were in fact unidimensional. If this does not prove to be the case, it is meaningless to rank jobs on the basis of these requirements and impossible to determine whether the ability of analysts is specific for each trait or is a more general ability. Seven analysts rated 33 traits for 50 jobs on a *yes-no*-scale. It was shown that all of the interest traits, half of the 13 personality traits, but only 3 of the 10 ability traits were quantifiable. In addition the ability of the analysts was very *specific* for each requirement. This finding should be compared with the previous results obtained by Rupe with respect to analyst ability.

The study of *leadership dimensions* is an extensive one which has been studied using many different approaches. One example is by Wofford (1970) who performed a factor analysis of managerial behaviour variables. Variables such as "encourages his men to set unit-wide goals and targets", describe how the job incumbent performs the work but also imply the underlying ability to act in this way. In an earlier study, Wofford (1967) presented the theory that human behaviour can be categorized in terms of 6 dimensions: security and maintenance, order and structure, personal interaction, achievement, personal enhancement, and group achievement.

The researcher who has perhaps devoted most time to worker-oriented studies is McCormick (1964). McCormick, Cunningham and Gordon (1967) based a study of job dimensions on the factor analysis of worker-oriented variables. This seminal study contains as an introduction a lucid discussion of the essential differences between work- and worker-oriented analyses. The

important point is stressed that *work-oriented analyses tend to be restricted to a specific occupational or technological area and can only be used for the transfer of personnel to related jobs or for considering the content of minor retraining programs within the area in question.* Worker-oriented variables, as has been mentioned previously, are more *fundamental* and can be used to study likenesses and differences in trait requirements over the whole spectrum of jobs. A Worker Activity Profile consisting of 162 more or less objective items is used in describing the worker-oriented elements of jobs. Some of the items are however outwith the worker-oriented area, the items being grouped into the following categories:

job activities – discrimination activities, mental activities, body and limb activities, supervisor activities, rhythm of work activities;

general characteristics – general characteristics of the job;

situation variables – physical environment, psychological and social aspects.

Two representative samples of about 400 jobs were selected. A total of 32 variables were obtained, grouped into the dimensions overall dimensions, mediation dimensions, physical output dimensions, communication dimensions, situational dimensions and environmental dimensions. Of the 32 variables, 22 were considered to be congruent when a comparison was made between the 2 samples. This could be compared with a previous study by Gordon and McCormick (1963), where 22 of 28 factors were found to be congruent. Four of the 10 incongruent factors were from the overall factor analyses.

The authors draw the conclusion that there is substantial structure in the domain of human work when this is described in terms of human behaviour and contextual and environment attributes of the work situation. They consider the study to be preliminary, however, and they intend to continue work by improving the worker activity profile and by making *direct* studies of jobs instead of using job descriptions.

Multidimensional scaling

Factor analysis is not the only method applied to reduce the number of worker-oriented or work-oriented dimensions in job analyses. In recent years a number of attempts have been made to apply the techniques of multidimensional scaling to processing of job analysis data.

Important early theoretical studies within this field were provided by Toops (1945) and Thorndike (1953). One of the first empirical studies was performed by Orr (1960), who presents what he regards as a new method for clustering jobs. Once again the 4.000 jobs selected by the United States Employment Service from the Dictionary of Occupational Titles have provided the basis for the study. Orr selects only one of the 8 components used in these studies namely the *aptitudes* (i. e. traits) necessary for successful job performance and he applies a distance (D) measure for classifying jobs into clusters with respect to these aptitudes. The job descriptions were taken

from the USES aptitude descriptions and 9 aptitudes were rated on a five point scale. Thorndike's (1953) method of determining the optimal number of clusters was employed and 6 were tentatively identified. This was basically an exploratory study, aimed at investigating the consistency and applicability of the techniques in question. Comparisons between different samples suggested that the clusters obtained were reasonably consistent. Three different kinds of clusters appeared, differentiated both with respect to type and level, namely *intellectual-supervisory, mechanical-manual* and *clerical.*

A number of other studies have used *multidimensional scaling* with the aid of *similarity ratings* to study the structure of tasks, occupations and behaviour descriptions for specific jobs.

Gonyea and Lunneborg (1963) made a study of perceived *occupational* similarity. A previous similarity by Gonyea (1961) using the method of Andrews and Ray (1957) was performed on 30 occupational titles. Intercorrelations were then estimated and 12 oblique factors were obtained which in second order analysis provided 5 orthogonal factors regarded as representing more general dimensions of job perceptions. This method involved certain difficulties however and the latter study applied a method of *triad comparisons* in which groups of 3 occupations were presented together and the subjects were asked to delete the one which was most unlike the other two. Twentytwo occupations were selected for study, the selection being partly based on a broad coverage of the areas in the Strong Vocational Interest Blank. The matrix of similarity ratings was transformed into an estimated intercorrelation matrix and this was factor analyzed using the principal axes method. The factor analysis yielded 7 factors, two of which were of marginal significance. Although there were some discrepancies, it was felt that the 5 higher order factors in the previous study by Gonyea (1961) and the 5 main factors obtained in this study had considerable likenesses.

A similar higher level study, this time comparing different jobs in a foundry, was performed by Mårdberg and Baneryd (1971). The purpose of this study was primarily to experiment with different *computational* techniques. One interesting feature was that the 15 variables to be used in the similarity ratings were obtained by means of *latent profile analysis* from a considerably larger number.

At the level of individual jobs a similarity rating of 18 tasks in a naval job was performed by Schultz and Siegel (1964). An 11-point scale was used to judge the similarity between each pair of tasks and the method of equal appearing intervals was used to scale the obtained inter-stimulus distance judgements. After the adjustment necessary to transform the data from interval scale to a ratio scale the matrix of scalar products was factored using the method of principal components. An orthogonal rotation followed and 4 dimensions were obtained. The job being studied was a naval-electronic one and the dimensions were thought to be meaningful for the job in question. The 4 dimensions were also considered amenable to unidimensional scaling.

A later study of this type was performed by Brown (1967) who carried out *similarity ratings* of behaviour descriptions for a specific job. Eighteen

behaviour statements were obtained with respect to the position of management analyst in the United States Federal Government. A methodological feature of interest in this study was the comparison of results obtained from the method of successive intervals (cf. Messick 1956) and the method of non-serial matching originally proposed by Andrews and Ray (1957), and already used in the study by Gonyea and Lunneborg (1963). In the first method 120 employees in the position of management analyst rated the behaviours for similarity on a 7 point scale; in the latter case they were asked, for each type of behaviour, to select the other behaviour which was most similar to it. The data processing followed previously reported lines. In the first method 5 factors were identified and in the second, the A-technique analysis, 7 factors were obtained. A comparison of factors suggested however that the 2 techniques provided much the same results. The results obtained were related to previous research on *consideration* and *initiating structure.* The main advantage of the A-technique is the ease of administration and reduction in the amount of work involved by the rater.

Mårdberg's work on LPA

The work of Mårdberg in Sweden is worthy of special consideration. He has initially studied the application of *latent profile analysis* to job analysis data and this has successively led to a more general study of the methods which should be used to process data of this type.

In a theoretical study which also made use of secondary data, Mårdberg (1966) made a comparative study of a wide range of multidimensional and multivariate techniques for the reduction of data dimensionality. The methods investigated include multiple discriminant analysis, component and factor analysis, profile (distance) analysis, Q-factor analysis, and latent profile analysis. Mårdberg arrived at the conclusion that factor analysis is often a restricted and in some cases incorrect method of attempting to reduce the dimensionality of the data. We will return to the general problems involved in the following chapter.

The method of *latent profile analysis* (LPA) was applied to foundry data by Mårdberg and Baneryd (1971). Part of the study consisted of an analysis of this type of 38 jobs rated on 30 variables. The variables covered a wide range of psychological, technological and physiological factors in the work situation. A 13 profile solution was selected by the researchers as being the most suitable one and these 13 profiles were judged to be meaningful. The authors stressed, however, that the application of LPA had primary been used as a *mathematical computation* and that generalization to other jobs was not permissible.

Mårdberg then used this study as a starting-off point for a more general study of different methods of utilizing data of this type. One study, (Mårdberg, 1971a), took up a general discussion of the possibilities of clustering jobs by latent profile analysis. The method was considered to be an alternative one for clustering jobs comparable to that reported by Orr (1960).

Thirtyseven jobs in a foundry were rated on 8 psychological requirement variables. A 7 profile solution was judged to be the most interesting one. In a comparison made with a 3 profile alternative it was shown how an increasing number of profiles tended to give a more differentiated representation of the basic 3 profiles. Classification scores were then used to determine the jobs nearest to each latent profile and these jobs can then be used subjectively to constitute the nucleus of each latent profile. The profiles obtained were felt to be psychologically meaningful. A further study (Mårdberg, 1971b) discussed latent profile analysis as a non-linear alternative to factor analysis. The fact that data in LPA does not have to be linear, nor follow a normal distribution is stressed. A distinction is made between LPA which aims at classifying in this case jobs with respect to variable profiles and Q-LPA where the variables can be classed with reference to job profiles. Of the original 7 profiles obtained in the previous study (Mårdberg, 1971a), the 4 most important job clusters are used in the subsequent analysis aiming at reducing the 8 variables to non-linear domains, i. e. performing a Q-LPA. When this Q-LPA is performed a solution corresponding to 3 assumed profiles was considered to be the best one. These variable groups appeared to divide jobs into 3 non-linear requirement domains, namely *general ability, perceptual ability* and *motor ability.*

Mårdberg then continued his mathematical manipulations by attempting to classify the 38 jobs into groups by performing LPA on the profile matrix obtained from the Q-LPA. Four assumed profiles were judged to give the best discriminating ability. These latent profiles could then be used to classify the jobs on the basis of the principle of the nearest latent profile (Gibson, 1962). The 4 job classes obtained were considered meaningful. Mårdberg stressed that the methods applied do not permit significance testing, but he did not consider this a serious disadvantage, since multivariate analysis in the behavioural sciences tends to be of an explorative nature.

A subsequent study made a more detailed analysis of the Q-technique of latent profile analysis (Mårdberg, 1971c). Fiftyfour variables were selected for ratings with respect to 38 foundry jobs. Twentythree theoretical profiles were selected as the best alternative and these profiles were then used to classify the *variables* into homogeneous clusters. Some of these were heterogeneous, and only variables with high classification scores were included when interpreting them. The analysis resulted in 8 variable domains: sensory attentiveness, production responsibility, heaviness of the work, social freedom-constraint, variability of the work, independence of the work, responsibility level and, finally, variations-monotony. A factor analysis was then performed to give rise to second order factors. Three were obtained, namely *heaviness, responsibility* and *social constraint.* These 3 linear factors could be compared with the finding in the previous study, namely the variable groups of *general ability, perceptual ability* and *motor ability.* Jobs with *responsibility* required *general ability, heavy work* placed demands on *motor ability* and jobs with *social constraints* placed primarily demands on *perceptual abilities* such as *attentiveness.* Thus the two methods arrived along

different paths at psychologically meaningful and consistent results.

The purpose of this chapter up to now has been to present a general account of methodological advances in structuring job analysis data. A critical examination concerning the value of the various methods and an attempt to present general models will be presented in chapter 3.

Critical reviews and bibliographies

The purpose of this section is to briefly mention and comment upon some of the most important general studies, reviews and bibliographies covering the field of job analysis. Ghiselli (1956) in the course of an address to the American Psychological Association in 1955 deals with "the placement of workers; concepts and problems". After a broad review Ghiselli concludes that there is no entirely satisfactory classification system and this suggests that the validity of a selection procedure must always be extremely specific. Jones *et al.* (1953) present a survey of the literature on job analysis of technical positions in the course of which it appears that technical work is by no means easy to define. Kershner (1955) presents a valuable critical study of job analysis, with sections covering many of the areas dealt with in this study. A critical study of the job analysis methods used by the American Department of Defence is presented by a Militärpsykologisk arbeidsgruppe (1956). Studies in Denmark showed that job analyses were *unable* to reveal traits of which officers were not already aware. A further result was that those supervisors who were judged to be incompetent were found to be unsatisfactory for reasons which would have made them incompetent for other jobs as well. In other words the reasons for the incompetence were *not* specific to the jobs in question. Morsh (1962) presents a bibliography comprising 1.500 titles for the period 1911–1961, including a complete list of reports from the U.S. Department of Defence.

The most extensive review of job analysis research findings in recent years has been presented by Prien and Ronan (1971). This report, which became available at a fairly late date in the writing of the present study, covers a number of the studies reported here and has the additional advantage of reporting on a number of studies not easily obtained in Sweden. The study is primarily restricted to the possibilities of using job analysis within the *clerical* and *management sector*, but a wide range of other problems are also covered. Prien and Ronan make the point that the analysis of jobs can be compared with the traditional methods of work measurement applied by industrial engineers. Little of the work in this field has, however, permitted generalization. Attempts at producing general job taxonomies have on the whole been unsuccessful, although the work of Fleishman (1967) aiming at providing a behavioural taxonomy for describing jobs has produced 10 basic psychomotor factors and 9 physical proficiency factors. Prien and Ronan feel, however, and in this the present author concurs, that the application of these 19 factors will not necessarily make it easier to understand what constitutes the total job performance of a typist, machinist or car driver.

Finally, Prien and Ronan draw the only conclusion which seems to be possible when studying the research in the field, to wit, that many of the studies appear to be a goal in themselves and that "relatively few studies have been performed which extend or apply the results of broad task analysis studies" (*op. cit.* page 389). Prien and Ronan also draw a conclusion which has in fact been the *leitmotiv* of the present study, namely that "whole disciplines have mushroomed focusing on attitudes, values, performance style and motivation in relation to the work people do with only the slightest suggestion that understanding *what* the individuals do is at all salient. To these writers it would appear that any measurement and certainly interpretation is contingent upon a clear and concise description of the job being performed" (*op. cit.* page 391).

Chapter 3 Job analysis methodology – a critical analysis

Initially, we will here examine job analysis research and theory in terms of the various measurement scales or criteria presented in chapter 1. The chapter will then be continued with a more detailed analysis of some of the basic models involved and a discussion of their suitability in various practical applications.

Measurement scales

Type of measurement involved

Torgerson (1958) distinguished between measurement defined in terms of derived measurement, measurement by fiat or arbitrary definitions, and fundamental measurement where the number assigned can be determined in terms of basic physical properties, such as length and volume. The method of measurement primarily applied in job analysis has been to ask individuals one at a time to state what duties are being performed and which traits are required to perform them; alternatively the respondents have been provided with some form of framework which can be used to structure their responses. A third alternative is for the researcher himself to make observations and to develop job descriptions or analyses, or to have some conceptual framework within which these can be structured. The basis of the measurements is thus subjective *assessment* made by individuals. This must be regarded as *measurement by fiat.*

Operationality of definitions

The section on terminology in the previous chapter has shown that there is little agreement as to what is meant by job descriptions, job analyses, job specifications. In some cases this has been due to the fact that different researchers have interpreted and defined the concepts differently. More frequently there has been an unwillingness to give exact operational definitions both of the major concepts and the various sub-units involved. In many studies where basic concepts such as tasks, traits and even jobs themselves are shadowy, it is hard to see how any sort of conclusions can be drawn from the ratings performed by different respondents.

The tasks performed in different jobs have normally been established by obtaining a large list of tasks from knowledgable respondents and by applying some selection and reduction technique in order to obtain a manageable number of tasks. Ratings in worker-oriented studies where worker trait

requirements have been involved have not always been operationally defined; at any rate this consideration has not been dealt with in detail in the research reports. The many indirect studies which have used the USES data about job descriptions and ratings of trait requirements may be said to provide more careful definitions; it is hardly surprising that many studies have gratefully used the secondary USES data. The study by Mårdberg and Baneryd (1971) is an exception in that the psychological, social and medical requirements of foundry workers are carefully defined and the scalability of these requirements is discussed.

Theoretical approach to the measurements

It will not always be possible to separate the subject matter in this section and the two following ones. In chapter 1 a distinction was made between the subject-, stimulus- and response-centered approaches and it was stated that the first-named approach had little relevance for the present subject. The stimulus-centered or judgement approach was divided into a fourfold table (page 22) with respect to the nature of the stimulus rating involved, while the response approach was studied in a different form of fourfold table (page 22).

Initially one is struck by the fact that, although it is normally assumed that job analyses are in fact *judgement methods* as opposed to *response methods,* it is obvious that there is much confusion as to the nature of the assessment being made. In situations where jobs are being rated by a group of homogeneous raters, we may conclude that this is intended to be a judgement study, where each job in turn can be rated with respect to each task or behaviour description. If however the raters are heterogeneous an extra variable is being introduced and we are in fact dealing with a *response situation* where the cognitive structure of individuals or groups should be investigated. In any type of suspected heterogeneous situation it is highly doubtful whether a factor analysis of the data matrix involved will provide any form of meaningful results. The whole problem of the manner in which the response is made will be a recurring feature of this chapter. Is it possible to state whether any one (group) of the judgement techniques presented in chapter 1 is inherently better than any other? If there are some doubts as to the dimensionality of the data, it would be unwise to apply a technique such as a first-order categorical rating with built-in transitivity. As many of the variables such as tasks, behaviours and traits are by no means certain to be unidimensional, the most interesting techniques would be those of paired comparisons (in order to test transitivity) and the multidimensional ratings of distance, similarity, or the methods of tetrads or triads. A variety of these techniques has been applied by Orr (1960), Gonyea (1961), Gonyea & Lunneborg (1963), Brown (1967) and Schultz and Siegel (1964).

Although these methods provide some check as to the dimensionality of the data, the fact still remains that response methods are often more suitable for job analysis studies, in that the researcher is often interested *both* in differences among jobs *and* differences among raters.

Nature of the response and scaling level

Few of the studies presented have examined the problem of scaling in detail. Most check-lists and other attempts to quantify the degree to which a trait is required or the extent to which some duty is performed are at the ordinal scaling level, although researchers by their choice of statistical techniques often tacitly assume that measurement is taking place at the interval level. In one of the few studies to descuss this matter, Mårdberg and Baneryd (1971), it is pointed out that some direct rating methods such as *noise scale* can be regarded as measurement at least at the interval level, whereas other scales such as *vibrations* can only be divided up roughly along a three-point scale such as *non-existent, present but not troublesome, present and troublesome.* In some ways this type of scale is no more than a nominal one, but Mårdberg and Baneryd nevertheless consider that a clear ranking can be made between the groups. Their subsequent data processing assumes that the values are summable.

Methods of multidimensional scaling such as similarity ratings produce, with certain assumptions, measurement at the interval level and thus permit the application of techniques such as factor analysis.

The only study dealing with the general scalability of job analysis variables is that by Mosel, Fine and Boling (1960) (previously reported). This study investigated the scalability of *interests, personality* and *attitude requirements.* The indices of reproducibility of Guttman (1950) and Jackson (1949) were used to study scalability; only *interests* appeared to be scalable with respect to both of these criteria. The *personality* and *attitude* requirements appeared to constitute some form of quasi-scale, the rank-orders being however perfectly efficient when correlated with outside variables (Guttman, 1950).

What then of the response approach and the fourfold table presented on page 22? In the opinion of the present author it is this type of technique which is liable to throw most light upon the nature of job analysis ratings. With the exception of the work of Mårdberg, which will be examined in detail subsequently, none of the studies investigated has applied these techniques, not even the Coombs unfolding technique, which *a priori* might have been expected to have produced a few adherents.

Methods of data reduction

The three basic methods used for data reduction in job analysis studies have been factor analysis, multidimensional scaling and latent profile analysis. None of the studies making use of factor analysis which have been reported on in this study has investigated to what extent their data satisfy the requirements for factor analysis, nor has space been found in these studies to discuss the application of the factor analysis technique adopted.

The most common technique is the principal component method using Kaiser's criterion to determine the number of factors and applying the varimax solution when performing the orthogonal transformation. The fact that IBM has a programme where the correlation matrix is fed in and

complete varimax results are obtained has undoubtedly contributed to the popularity of this technique.

In the traditional methods of multidimensional scaling, a matrix of scalar products is obtained from the rating data. This can then be factor analysed. A different technique is that suggested by Andrews and Ray (1957), the so-called A-technique, which directly transforms the proportion of common elements between stimuli into estimates of correlation coefficients for subsequent factor analysis. These similarity measures can be performed either by means of paired comparisons or by the method of triads.

Brown (1967) has shown the similarities between traditional multidimensional scaling and the A-technique; Gonyea and Lunneborg (1967) have shown that the paired comparison and triad A-techniques also produce similar results, while Schultz and Siegel (1964) show that the basic method of multidimensional scaling thus appears to be an extremely viable method for studying job analysis, especially in situations of an exploratory nature where the determination in advance of the variables, as in the case with factor analysis, may be felt to be unnecessarily restrictive.

The empirical applications of latent profile analysis within this field have been restricted to the work of Mårdberg and his associates, as far as the author has been able to determine. The basic advantages of LPA compared with factor analysis, over and above the less stringent data requirements, can be illustrated with the aid of the following general model, derived from Mårdberg:

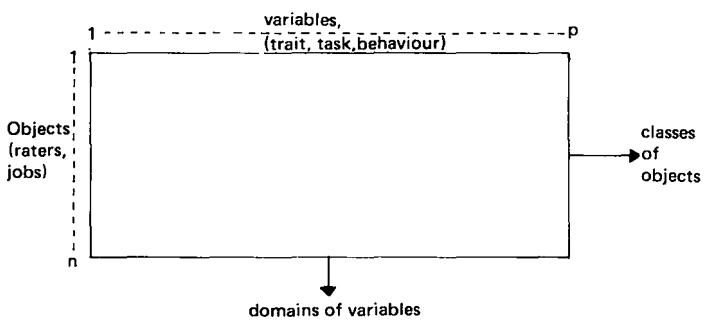


Fig. 3:1 The derivation of classes of objects and domains of variables

The model shows the basic situation concerning any form of dimensional reduction. On the one hand we have the n objects and on the other we have information concerning them with respect to p variables. Following Mårdberg (1969, 1971c) there are two basic types of dimensional reduction: on the one hand we can attempt to class the *objects* into *groups* on the basis of their variable scores, on the other hand we can attempt to group the *variables* into *domains* which are homogeneous with respect to object scores. In addition we must distinguish between situations where $n > p$ and $p > n$. We thus are presented with four different types of dimensional reduction as in the table below.

No of variables and objects	Reduction of	
	Variables (p)	Objects (n)
$p < n$	A (FA)	B (LPA)
$p > n$	C (Q-LPA)	D (Q-FA)

Cell A is the traditional situation for factor analysis, i. e. the reduction of the dimensionality of the variables on the basis of object scores. Data in cell C can also, at least in theory, be processed by the Q-method of factor analysis. This is extremely complicated if more than two or three factors are involved, as the technique necessitates the graphic plotting of factor scores. In practice this method is scarcely used. Cell B data can also in theory be processed by means of the standard methods of factor analysis, but as this case also involves the problem of graphic plotting of factor scores, its application is limited. Finally, in cell D, the reduction of the number of classes of objects where the number of variables is greater than the number of objects, the Q-method of factor analysis can be applied straightforwardly.

One of the merits of latent profile analysis is that it can be used with much less restrictions than factor analysis in all four situations. The primary application of LPA is to classify objects, such as jobs or raters, in homogeneous classes. This makes LPA the most suitable technique to apply in cell B. Similarly the reduction of variable domains can be achieved in a single step by applying Q-LPA to cell C. In cells A and D the situation is more complicated, as it is necessary to perform a two-step procedure to arrive at the required results. In the first case it is necessary to perform a Q-LPA on the object profiles obtained after LPA in cell B, and this will lead to a series of variable profiles. Conversely, an LPA performed on the classes of variables obtained from cell C will produce the required grouping of objects in cell D. As it is perhaps difficult to follow this brief theoretical account, the studies performed by Mårdberg (1971a, b, c), (summary in chapter 2), will be outlined within this new framework in order to demonstrate the scope of LPA for this type of data.

It is most convenient to begin from cell C. Here the data consist of 54 variables, covering a wide range of social, psychological, medical and technical job requirements. The extent to which these are necessary is rated for 38 foundry jobs. In other words, $p = 54$, $n = 38$, and the object is to obtain a smaller number of variable profiles for the jobs in question. This is performed by means of Q-LPA and the result obtained is 8 variable profiles for the 38 jobs. This 38 x 8 matrix can then be subjected to a further LPA-treatment which gives 4 object profiles, i. e. the result required in cell D. A start can also be made from cell B, using only 8 variables (obtained from the previous variable reduction) and, for some unstated reason this time only 37 jobs. Thus, $p = 8$, $n = 37$ and the object is to obtain homogeneous classes of

objects, i. e. jobs. An LPA is performed and 7 profiles are obtained, 4 of which contain at least 3 jobs. This 4 x 8 matrix can then be subjected to a Q-LPA in order to produce variable domains. Three of these are obtained, constituting a solution for cell A.

Latent profile analysis (LPA) is an extension of LCA to the situation where the manifest variables are either discrete but at least trichotomous or are continuous variables. The respondents are still grouped in discrete latent classes, where average values can be expressed with respect to each manifest value, thus producing a latent profile. Many data which are available in the behavioural sciences cannot even be said to approach a normal distribution. If the assumption of normality is nevertheless made, this means that only the first two moments of the data can be used to determine interrelationships*, as is the case with factor analysis. In LPA this assumption can be dropped and it is possible to use higher moments to determine the nature of the relationships between the variables. The data obtained from an LPA can thus be presented in the form of a profile. The data may for instance take the following form:

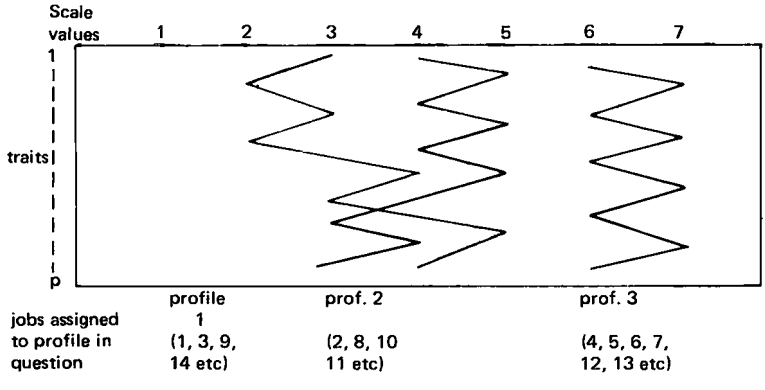


Fig. 3:2 Profiles obtained from an LPA

The alternative model, that of performing a Q-LPA would give rise to data of the following type:

Domain 1 (homogeneous with respect to job scores)
traits 1, 3, 4, 7, 10 etc.
Domain 2
traits 5, 6, 8, 12, 13 etc.
Domain 3
traits 2, 9, 11, 14 etc.

* $E(X-\mu)^k = \begin{cases} 0, \text{ if k is odd} \\ \sigma^k (1)(3)(5) \ldots . (k-1), \text{ if k is even} \end{cases}$
for normal distributions

The problem of determining the number of domains is similar to that of determining the number of job profiles. A further problem common to both treatments is the assignment of jobs or traits to profiles or domains. In LPA each of the traits is given the same weight, when determining the probability assignment of each job to a given profile. It is possible that the profile to which a given job is assigned will be determined by the proximity of the job to this profile when it comes to traits which are of *minor* interest, at the same time as there may be large deviations with respect to more important traits. Concentration upon traits of central importance might have resulted in the job being assigned to a different profile.

The primary value of an LPA is the increase in information provided when compared with other alternatives such as factor analysis. A profile of values with respect to different traits contributes considerably more information for a series of jobs than the factor loadings on different traits which constitute the end product of factor analysis. A further important feature of LPA is that the profile data is of the type used in specific job analysis applications such as in the selection situation. If we have access to ready-made profiles for a series of jobs, it will be easier to determine the suitability of applicants, not only for specific jobs, but also for a whole series of jobs in a classification situation.

The principle in a classification situation of that type would then be that the applicants are assigned to the profile which they most closely ressemble. Some simple distance measurement from each profile trait score would suffice to assign respondents to a specific profile. There would be difficulties in a situation such as this, of course. The profiles themselves are approximations or averages. If the individual applicant deviates radically from profile scores, there would be considerable risk involved in assigning him or her to one of the specific jobs in the profile in question.

In the case of a Q-LPA, the determination of domains for a given series of jobs would also be of considerable value in order to reduce the number of traits included. In future analyses the domains instead of the individual traits could then be used in order to produce latent job profiles.

Much of the work done by Mårdberg has broken new ground and he has been careful to document the many practical difficulties arising in applying LPA problems of this type. Although data requirements are less restricted, compared with factor analysis, there are still many stages where arbitrary decisions must be made. The main difficulty documented in practical work is choosing *the number of latent profiles.* There is no objective method of determining the number of possible latent profiles which is most satisfactory. This must be determined from situation to situation by means of a stepwise procedure, whereby an increasing number of latent profiles is computed, the different numbers of profile solutions compared and the one most easy to interpret selected. Other difficulties involve variable reduction into domains. As Mårdberg (1971c) points out, psychological theories will tend to be conserved if the same type of methods continue to be applied to them. Factor theory, as we have seen previously, was developed on the basis of trait

theory and the idea of independent dimensions. *Classes* of variables may be more interesting concepts in dynamic psychology theory.

A further restriction in the use of LPA is that the methods at present can to some extent only be regarded as *mathematical manipulations* and it is impossible to derive significance tests for the results, contrary to the situation in factor analysis. This restriction is not necessarily a serious one, as many studies of this type are explorative ones, scarcely suitable for hypothesis testing.

Furthermore, LPA cannot be used to process symmetric matrices from multidimensional scaling. As the literature has shown, these are normally analysed by factor analysis, although a non-linear alternative has been developed by Sheppard (1962).

A final quote *in extenso* from Mårdberg (1971 b, p. 30) summarizes the remaining complications as well as the technical advantages of using LPA as opposed to factor analysis.

"The main advantages of using LPA instead of factor analysis are the possibilities of conducting analyses on raw data without transformation and on scales of the type used by behavioural scientists, ordinal scales (which are assumed to be summable, however), rating variables, attitude scores and qualitative data. Assumptions as to normal distributions and linear relationships are not necessary. In those cases where the two methods give the same results, the linear information is sufficient to explain the relationships in the material. In cases where differences arise in analyses made by factor analysis and by LPA, this is a result of significant nonlinear information in the material. In these cases LPA is to be preferred. As a rule LPA may therefore be applied since the risk will then be smaller of losing information in the analysis. In addition one avoids the whole problem of indeterminability in factor analysis arising in connection with the problem of rotation. The use of triple correlations means that this is avoided in LPA."

Summing up, the vast majority of the variable reduction studies reported here belongs to what Mårdberg would call *class A*, in other words the situation which is most suited to factor analysis. The ratings performed are basically of two different types; in the first case a number of jobs are selected and these are rated by presumably competent judges with respect to a wide range of duties or traits required. Palmer and McCormick (1961), McCormick, Cunningham and Gordon (1967) and Chalupsky (1962) may be taken as examples of this type. The second situation consists of a number of judges using some sort of check-list or inventory to rate their own work as in the case of Brumback and Vincent (1970) or that of their superiors, as in the case of Wofford (1970). In some of the studies the factors obtained are presumably specific to a single job or job cluster, whereas other studies, such as those of McCormick and his colleagues, aim at arriving at more general work dimensions. These types of studies are of course valuable, but the detailed examination of Mårdberg's work shows that latent profile analysis greatly widens the scope and makes it possible to process data in a number of different and fruitful ways.

In situations where the objects being rated are jobs, LPA can of course be used not only to reduce the variable domain but also to group the objects together to form homogeneous classes of jobs with more or less the same requirements. When different rater categories make up the objects and rate one or more jobs it is also instructive to use LPA in order to study the extent to which the different classes of raters form homogeneous groups.

Reliability

With the exception of Mårdberg (1972), which deals with job evaluation, the author knows of no studies concerning job analysis where intraindividual comparisons have been made. A number of studies previously mentioned have shown that interindividual ratings within one category of raters tend to differ considerably. Some have, however, been quite satisfactory as for instance Madden, Hazel and Christal (1964), which deals with job incumbent responses to job activity questionnaires, and Palmer and McCormick (1961) with a median interrater correlation of 0.75 when applying a check-list to written job descriptions. Satisfactory reliability in a similar rating is also reported by McCormick, Finn and Scheips (1957). Still more perturbing is the low agreement with respect to duties and requirements manifested *between* rater categories such as job incumbents and supervisors, as shown by Lennerlöf (1966). This problem was briefly touched upon in the section in chapter 2 dealing with who should perform job analyses. The degree to which the values held by the different rater categories tend to produce different assessments must be regarded as a very serious weakness.

Morsh (1964) also showed that reliability varies depending on the time interval between ratings, thus suggesting that jobs, not unexpectedly, change over time. A number of studies covering the reliability of different methods of applying job analyses has previously been reported. The summing-up report by Rupe (1956) for instance showed, somewhat surprisingly, that the questionnaire was the least reliable method, while the interview was the most reliable one. It should once again be stressed that differences in the efficiency of the analyses were greater than differences in the efficiency of the methods. Many of the studies emanating from the USES data have also shown that in many respects indirect ratings based on DOT job descriptions or the trait requirement ratings on the basis of these descriptions are as reliable as direct observations. Carroll and Taylor (1969) studied the validity of estimates by clerical personnel of job time proportions and established that the average correlation between estimated and actual time or allocations for individual workers was 0.88, with only 2 of 16 correlations under 0.80. The researchers regard this as a validity study although in the opinion of this author the study belongs more to the twilight zone mentioned in chapter 1 between reliability and validity.

Validity

The traditional difficulty in distinguishing between reliability, i. e. the accuracy with which one measures whatever is being measured and validity,

i. e. the accuracy with which one measures what one intends to measure, is apparent in job analysis studies as in many other subjective rating situations. One type of study is reported by Prien and Ronan (1971) in which different factor studies of the same occupational group are compared. Prien and Ronan show for instance that studies at the occupational family level by Coombs and Salter (1949) and Orr (1960) have much in common. Within the clerical field a comparison of factors obtained by Prien (1965), Thomas (1952) and Chalupsky (1962) are also surprisingly similar. The authors also underline the well-known observation that the results of the factor analysis of questionnaire items are of course in the last analysis always dependent upon the variables chosen to be included in the questionnaire. Differences obtained between factor studies may to some extent reflect differences in the questionnaires used. A final study by Prien and Ronan (1971) with respect to a number of studies of managerial and executive functions also shows a surprising agreement between four different studies. The extent to which these measures may be regarded as validating the job analysis techniques applied is open to some doubt.

Many of the studies using multidimensional scaling which produce comparisons of factor congruence when different methods of data collection are used are also of relevance to the question of validity. Orr (1960) obtained a fair degree of factor congruence when using traditional multidimensional scaling and the A-technique; the study by Gonyea and Lunneborg (1963), comparing two applications of the A-technique, namely the method of non-serial matching and the method of triads, also showed factor congruence. Orr (1960) compared two different samples when applying the distance measure (although this should perhaps be classed as a split-half *reliability measure*). Further comparable studies not available to this author performed by McCormick and his colleagues in 1969 are reported in Prien and Ronan (1971).

These types of validity measures, while important indicators of satisfactory measurement, are not the central problem of the validation of job analysis techniques. The most interesting application, at least within the framework of the present study, is to study the extent to which the results obtained from job analyses can be used to provide a basis for selecting applicants and predicting success in the jobs in question. This is the type of validation to which further space will be devoted in chapters 4 and 5. There are of course other tasks for validation studies such as in a training situation, where future job performance for those whose training has been based on job descriptions and job analyses can be used as the criterion measure. Another approach is validation using attitude studies.

Within the selection procedure, the field has been almost entirely neglected. One study which has made a tentative attempt to fit in the role of job analysis as part of a total selection procedure has been provided by McCormick, Cunningham and Thornton (1967). A further examination of the issues involved will be postponed until chapter 5.

Sampling techniques and representativity

Sampling problems do not play an important part in this area, although a number of interesting questions do arise. In many of the studies reported on, no attempt has been made to cover a general field. Instead one job or area of jobs has been the subject of study. The 4.000 jobs used by the USES for ratings of trait requirements were selected from the 13.000 jobs described in the Dictionary of Occupational Titles, by means of sampling techniques aimed at providing a representative cross-section of the type of jobs to be found in the United States. Some of the studies reported on have attempted to select representative samples from these 4.000, while others have clearly stated that the actual jobs studied were *not* representative of the 4.000, or of the total population.

The fact that many of the studies have covered a specific job or job area implies that the results can scarcely be generalized. A comparison of studies of job occupation families, clerical duties and managerial variables shows that a certain number of these studies appear to have a general application extending beyond the sample of jobs studied. One feature in favour of generalization is that many of the studies have been performed on actual jobs and job situations, thus removing the objects studied from the artificial and sometimes suspect area of the laboratory. Many of the studies, especially those based on the USES data, have however used this material as a secondary source which tends to restrict the possibilities of interpreting the results in more general terms.

There has been a definite bias towards the study of white-collar workers. As the job content and trait requirements for these positions vary greatly from blue-collar jobs, where little work has been done to date, it must be concluded that the total body of research is not entirely representative of all of the main types of work.

A further question is the extent to which primarily American studies can be applied to other cultures. For linguistic reasons the ability of the author to follow scientific work closely is restricted to Scandinavia and the English-speaking countries. This study of the literature has shown these language areas to be dominated by American studies. Although there are many areas where one would hesitate to apply American studies to other cultures, this does not appear to be one of them. The pattern of jobs in any highly industrialized country will tend to be fairly closely comparable to any other one, assuming that the level of technological development is about the same. In Scandinavia there is for instance a Scandinavian vocational classification (*Nordisk yrkesklassificering*, 1966) which appears to be modeled on the American Dictionary of Occupational Titles. Results obtained in the United States may with a fair degree of confidence be considered applicable in other highly industrialized countries.

Aid to theory construction

Before the contribution to theory construction of the research reported on can be assessed, we must examine briefly what is known about work and the worker in modern life. In this field there is a tendency towards oversimplification which should be avoided. To be sure, work is of central importance to many adults but this is by no means always the case in industrial or other societies where the necessity to work varies. Furthermore, the importance of work will tend to vary over the life span of the individual. Intra- and interindividual differences should be taken into account in any theory construction concerning the psychological and physiological matching of individuals to jobs and vice-versa. An underlying value judgement is that people will be *happier* if they are performing tasks for which they are best *suited.* This then is the background and the task for research is obvious: to attempt to describe different types of jobs, to classify them in different ways, to determine what is required of job incumbents, to establish the attitudes and capabilities of these putative incumbents, and to provide a basis for rules and measurement instruments as to how the labour force might be classified with respect to their ability to perform different jobs. To some extent these goals have been achieved. In the Dictionary of Occupational Titles a very large number of jobs have been described and classified and trait ratings on the basis of the descriptions have been shown to have some practical value. Many other studies have, however, tended to complicate the picture rather than to elucidate. Studies by Fleishman (1967) have shown how difficult it is to establish a general taxonomy of occupations. Even when this is achieved at the psychological level there tends to appear a gap between the findings and practical applications. The many job description questionnaires, some of which are worker-oriented and others work-oriented, have tended to confirm the terminological confusion concerning these matters. If it is accepted that the primary aim of job analysis is in some way to relate individual traits to job requirements, it would seem that the most fruitful path to be pursued is that of the work-oriented questionnaire. Although the questionnaires of McCormick and his colleagues are a combination of worker- and work-oriented variables, the progress which they have made in attempting to derive general dimensions of work are the most promising to be found in the literature. In general, however, there is a lack of research concerning inter- and intraindividual differences. The methodological aspects concerning variable reduction, mathematical and statistical models have been dealt with in previous sections. What the present author feels to be lacking in the literature on job analysis is an attempt to fit the job analyses into a general theoretical structure, aiming at solving the general job/job incumbent classification problem. It is in fact this apparently unbridgeable gap between the technicalities of job analyses on the one hand and the problems involved in recruitment and selection on the other which was one of the reasons originally directing the author to this present field of study.

The usefulness of the research to society

Although some of the factor analytic and multidimensional scaling studies appear to be more centered on technicalities of data processing than on arriving at practically useful results, the majority of the studies is directly or indirectly applicable to practical problems. In the previous section mention has already been made of the part which job analysis can play in solving what might be called the general "social classification problem", i. e. in assigning people to jobs for which they are suited and, in general, matching job requirements and individuals in society as a whole. It may be maintained that this should be a central goal in any society; it is even inscribed in the constitution in countries such as the United States. Steps towards a "felicity calculus" of this type, whereby most people are fairly well suited to their present working duties, involve on the one hand the determination of job requirements and on the other hand techniques for measuring individual aptitudes. The consequences of being unable to determine job requirements at present are obvious. On the one hand people may be assigned to jobs for which they are manifestly unsuitable, with all of the social adjustment problems which that in the long run involves. On the other hand, inability to determine job requirements can mean that people fail to realize the extent to which certain jobs place entirely unreasonable demands on the employee, at least if the employee's long term capacity is considered. In Sweden there are for instance already signs that two labour markets are developing side by side, one of which is the open market, and the other a protected market where citizens who for one reason or another are unable or perhaps unwilling to meet the open market requirements are provided with less demanding jobs or a pension by the State. This suggests that a permanent gap is beginning to develop between the job requirements placed on people in Swedish industry and the total capacity of citizens as a whole. An improvement in the methods of job analysis could help to solve one side of this equation and contribute to a more equitable, and in the long run, more efficient solution to general labour market problems.

Job analysis models

It is now time to draw together the various threads of research on job analysis techniques and to present them within a more general framework. With the possible exception of McCormick and his associates this has scarcely been attempted previously. The majority of the researchers reviewed to date have restricted themselves to the testing or investigation of some specific limited technique, without attempting to discuss the practical applications or to place the research in a broader perspective.

Initially, it is possible to distinguish between a number of variables involved in job analysis. Firstly, there is the analysis level at which the study is to take place, in other words whether the study involves units such as *occupations, jobs* or *positions,* as well as the number of units to be studied

simultaneously. Secondly, there is the question of in what way the job is to be analysed. The previous discussion has shown that jobs may be studied in terms of *tasks, behaviour* or *traits* involved. Thirdly, is the study being performed by *one* or *more* raters? If the number of raters is greater than one, are the groups involved *homogeneous* or *heterogeneous,* i. e. do they belong to the same or different categories? Fourthly, the nature of the rating being performed. Assuming that the object of the exercise is some form of variable reduction, is the intention to obtain *clusters* of *tasks, behaviors, traits, occupations, jobs, positions* or *raters*? Fifthly, the nature of the response, i. e. whether it be *subject-centered, stimulus-(judgement-)centered* or *response-centered.* If judgement-centered, whether the technique be *comparative* (paired comparisons etc.) or *categorical* (such as sorting). If response-centered, which combination of *deterministic/probabilistic* and *categorical/ comparative* rating is applied? Sixthly, the method of data-processing the responses (to some extent determined by the nature of the response). The most important distinction to be made here is between *factor analysis* and *latent profile analysis.*

For the sake of clarity, the various alternatives are summed up in the table below:

variable involved	possible alternatives
1) level and number	occupation(s), job(s), position(s)
2) rating unit and number	task(s), behaviour(s), trait(s)
3) number of raters	one, homogeneous group, heterogeneous group
4) possible cluster of variables	tasks, behaviours, traits, occupations, jobs, positions, raters
5) nature of response	subject, judgement, response; comparative, categorical; probabilistic, deterministic
6) data reduction	factor analysis, LPA

We are now in a position to classify the research previously presented in terms of the variables in the above table. Only a limited number of all possible combinations has been studied, but this will serve to illustrate what research has attempted to achieve within the field and, more important, the problems which remain to be solved. It is perhaps most convenient to group research into the following eight basic models, each of which can be broken down into two types of cluster variables.

Model 1 This model consists of ratings of *a number of jobs,* n, with respect to *a number of tasks or behaviours,* p. The jobs are normally rated on a simple rating scale with respect to the importance of, or the presence of, the

tasks or work behaviours in question. The assumption appears to be that this is a *judgement-centered rating* with normally a *categorical* response pattern. The raters may be either *homogeneous* or *heterogeneous*; the fact that a judgement-centered rating is assumed suggests that for instance different employee groups are regarded as homogeneous with respect to the rating in question. The model may be represented graphically as follows:

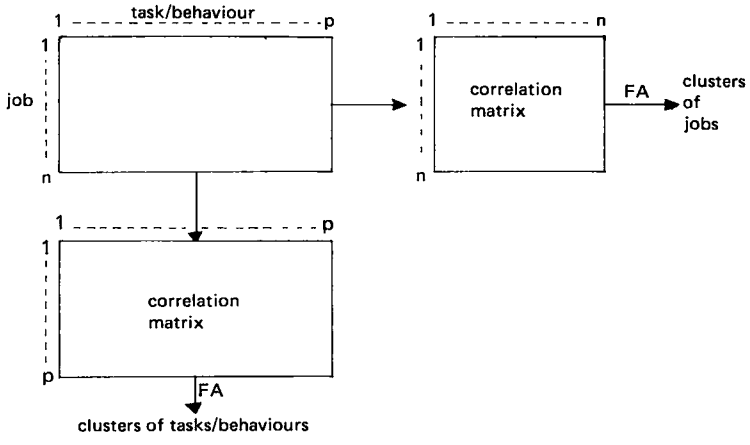


Fig. 3.3 Model 1: task/behaviour vs job

As may be seen from the figure, there are two different possibilities of processing this type of data. The first method involves taking for instance the average of the raw data in each cell, i. e. the ratings obtained from different raters and calculating the product-moment correlation *horisontally*, determining the agreement between task/behaviour patterns for different jobs. This produces an n x n correlation matrix which can then be analysed by some traditional method of *factor analysis* in order to provide clusters of jobs with various tasks/behaviours in common. This type of application has in fact seldom been performed, although one example is provided by Chalupsky (1962). The value of the application is obvious. Assuming that the sample of jobs included in the study is not entirely random, it would for instance be possible to use the job groupings in order to determine differentiated training programmes for the different categories, not to mention the uses in classification recruitment situations. Furthermore, the data could be used to provide job descriptions for each category in behavioural or task terms.

The other basic alternative for processing the data is to calculate the product-moment correlation *vertically*, thus showing the extent to which different tasks/behaviours tend to show similar patterns. This produces a p x p matrix which when factor analysed gives rise to clusters of tasks/behaviours. This type of study has been performed fairly often, see for instance Thomas

(1952), Thorndike *et al.* (1957), McCormick *et al.* (1967) and Gordon and McCormick (1963). The object of this type of study, to the extent that it has been at all specified, is to determine the *minimum dimensionality* in terms of tasks/behaviours which can be used to group jobs. These clusters of tasks/behaviours can in fact be transferred to clusters of jobs by assigning the jobs the appropriate weights for each cluster of tasks/behaviours, see McCormick *et al.* (1957).

One further problem which we will have occasion to return to repeatedly is the extent to which all of the units rating or being rated are *representative.* Firstly, with respect to the jobs involved, it is obvious that they will normally be selected from a relatively limited area, for instance within a given occupation or the jobs at a specific workplace. This seems eminently sensible for practical purposes; we need have no quarrel as to the representativity of the results: they are of course not representative and it is not so intended.

Secondly, selection of tasks/behaviours is another matter. In order to establish the items to be included it will be necessary to obtain a job description expressed in those terms. One method of doing this is to apply one of the many job analysis techniques briefly presented in previous sections. If the study is a preliminary one, it may be maintained that excessive detail is not necessary in view of the fact that one of the functions of the job analysis will be to give rise to clusters of tasks/behaviours which may then be used in future instead of the original list *in extenso.*

The third aspect of representativity refers to that of the *raters.* The usual procedure in studies of this type is to make use of a number of raters. This permits the determination of the reliability of the ratings by means of simple interrater comparisons. This has been done in a number of the above-mentioned studies, normally producing satisfactory results, at least with respect to within-rater categories, i. e. homogeneous groups. The problem arises, however, where heterogeneous groups are involved and where each group gives results which differ considerably from the others, perhaps even to the extent that completely different clusters of tasks, tasks or behaviours are obtained when factor analyses are performed for different rater groups. There is considerable evidence of this in a number of studies and the subject will be investigated in greater detail subsequently, especially the factors underlying these differences. For the moment it will suffice to pose the question: To what extent is it at all possible to perform any type of objective or fundamental measurement when repeated studies show that the factors obtained tend to reflect the perceptual structure of the heterogeneous groups performing the ratings?

Model 2 This model consists of a *single job,* described by means of a *number of tasks or behaviours,* p, by a *number of raters,* n. As before the rating refers to the extent to which the different tasks/behaviours are relevant or important. This also appears to be a *judgement-centered rating* using a *categorical* response pattern. The raters are *homogeneous,* although this is rarely made explicit. The graphic representation is as follows:

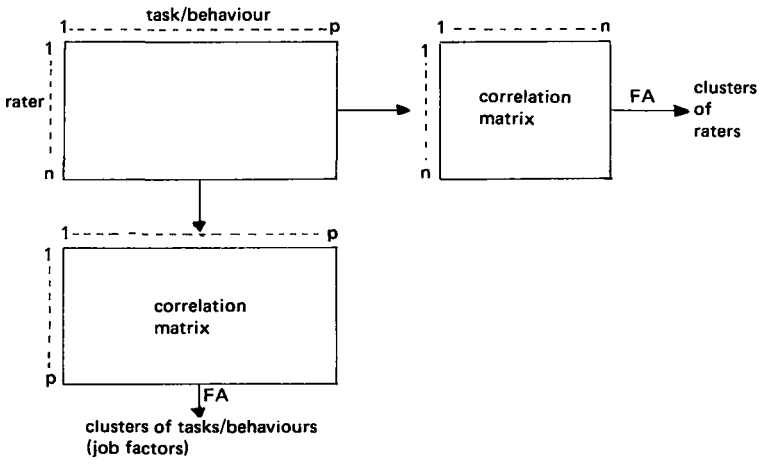


Fig. 3:4 Model 2: task/behaviour vs rater, one job

Once again there are two approaches to the processing of the original data matrix. On the one hand it is possible to obtain an n x n correlation matrix horisontally which, when factor analysed, will provide clusters of raters or, conversely, a measure of the extent to which the group of raters is in fact homogeneous. The author has been unable to find a single study of this type in the literature. The presumable explanation is that most practical researchers consider job analysis techniques to be complicated enough without questioning the implicit assumption that job analysis ratings are in fact judgement ratings and not response ratings.

The other alternative is to obtain a p x p correlation matrix which will reveal the extent to which the tasks/behaviours can be grouped together into a smaller number which may then be regarded as *job factors* for the job in question. Studies of this type have been performed by Saunders (1956), McQuitty *et al.* (1954), Peres (1962), Prien (1963), Brumbach and Vincent (1970) and Wofford (1970). Some of these researchers have adapted the basic technique by asking the raters to distinguish between a good and a bad job incumbent and to rate a stereotype from each of these categories with respect to the tasks/behaviours. Studies of this type have claimed to obtain significant differences between the two types, although in the view of the present author this must involve a certain amount of confusion between tasks, traits and behaviours.

The basic and most frequent application of this data-processing technique is that of establishing job descriptions of a more limited nature from a wide range of tasks/behaviours. In addition, the data obtained can also be utilised for instructional purposes and for obtaining variables which might be of special interest in a personnel appraisal system. Finally, the job descriptions may subsequently be used as the basic material for translating these descriptions in task or behavioural terms to traits.

In general this model is an extension of model 1. Instead of rating a whole series of jobs, one specific line in model 1 is selected, and is instead expanded to include a series of raters. Model 1 is in fact really a three-dimensional one, if the raters are regarded as heterogeneous, whereas model 2 reduces the situation to two dimensions, making differences between the raters an integral part of the study. This seems to be rather paradoxical. A certain amount of variation is required in order to process the data at all, while too much variation would suggest that completely different perceptual frameworks are involved on the part of the raters towards the jobs and that the basic job factors obtained would thus be worthless.

Model 3 This model is similar to model 1, the important difference being that the jobs are rated not with respect to tasks or behaviours, but instead with respect to the individual traits or attributes which are necessary to perform the jobs in question. The model is as follows:

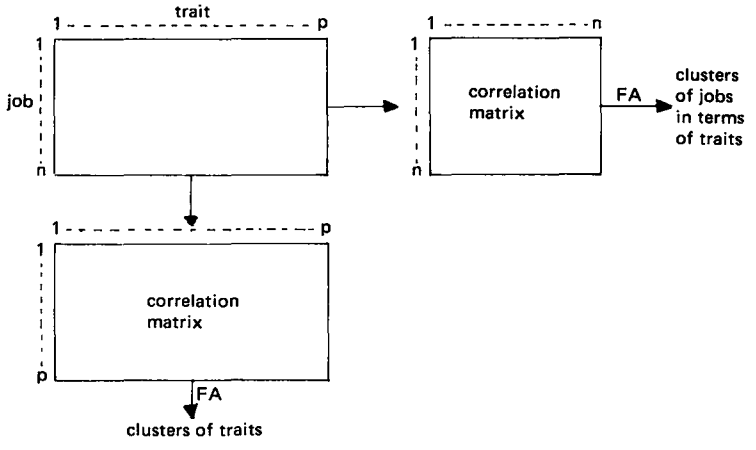


Fig. 3:5 Model 3: trait vs job

As in model 1 it is possible to obtain a correlation matrix horisontally or vertically, which can then be factor analysed. In the horisontal direction clusters of jobs with respect to traits will be obtained and in the vertical direction clusters of traits. This horisontal treatment is surprisingly uncommon in view of the obvious value of the method in connection with classification problems and career patterns. The vertical approach has been considerably more common, especially in situations where it is felt that the number of traits required to describe the jobs in question can be reduced. Examples of this type of study are Jaspen (1949), McCormick *et al.* (1957) and Norris (1956). As was the case in model 1, these clusters of traits can then be used to perform a subsequent grouping of jobs. This has in fact been done by McCormick *et al.* (1957).

The greatest significance of model 3 as opposed to the two previous models is that an attempt is made to distinguish between what is *done* or the

behaviour required to do it on the one hand, and the psychological *traits* which are required to perform the work on the other. The major application of this type of data is of course in connection with selection and classification problems. It is difficult to judge how successful applicants will be in their future duties, but if the attempt has been made to translate these duties into required traits, then the assessment situation immediately becomes more amenable to the standard methods applied to assess people in the selection or classification situation, i. e. school marks, references, interviews, medicals and psychological testing which are of course at present based on trait psychology.

Two questions arise in this respect: firstly, the extent to which people in general are able to rate traits on the basis of job descriptions, observations or their own experiences, and secondly, the extent to which they are bound by their own subjective frame of reference. Whereas it might be expected that there will be a fair measure of agreement among raters with respect to the tasks or even the behaviours involved, it is not at all certain that people from a group all having the same cognitive background will be in agreement as to the *traits* involved in performing these tasks or in manifesting that behaviour. It is even less likely that people belonging to different rater groups with widely different cognitive structures will tend to assess the required traits in the same way. Repeated studies, especially those referring to criterion data (looked into subsequently), have shown that superiors, peers and job incumbents themselves have widely differing ideas as to how the job in question should be performed, the attributes (traits) which are necessary and even the extent to which the job incumbent in question has been successful in performing his or her duties. A more general issue arising in model 3 is from whence the traits have been obtained. Should model 3 be regarded as the next stage of model 1, the traits having been derived from the behaviours and tasks in that model? Is not the trait approach a fundamentally unsound one in that job requirements are thereby fixed to a specific number in advance, precluding any other job requirement areas not included in the original model? There is much truth in these comments and they will be taken into account in the treatment of the criterion problem in chapter 5.

A further recurring problem arises in this model. It is of course important to determine that the traits which are selected for rating are in fact relevant and that they more or less cover the total range of different aspects of the job. One method of attempting to achieve this would be to ask different categories of raters to generate their own lists of important traits; these lists could then be combined into some form of common denominator which would attain a broad consensus of acceptance by the different categories. The present author has however been unable to trace any study of this type in the literature. The main source of material has in fact been the traits determined within the framework of the DOT/USES material.

Model 4 This model corresponds to model 2 in the same way as model 3 does to model 1.

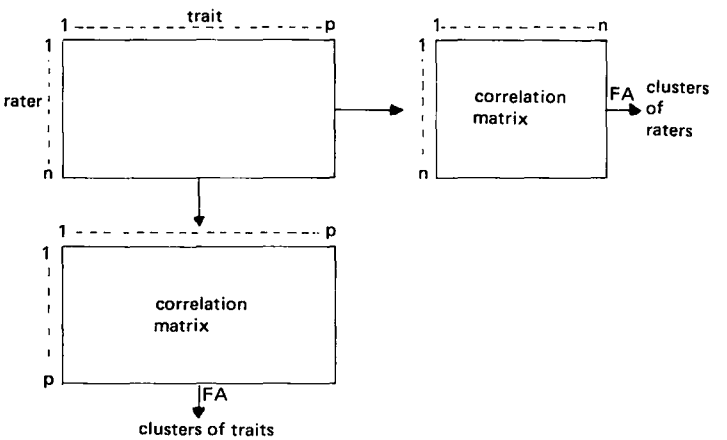


Fig. 3:6 Model 4: trait vs rater, one job

The difference is that the tasks or behaviours along the horisontal axis in model 2 have been replaced by traits in model 4. The vertical treatment thus gives rise to clusters of traits with respect to the specific job in question, while the horisontal treatment produces, at least in theory, clusters of raters. Many of the previous comments are also relevant here. The vertical treatment is one which has seldom been applied in practice, although this would appear to be the logical way to perform a trait analysis for the job in question. Studies of this type, where the design has been primarily that of a rating being made of the best and worst job incumbent known to the raters have been made by McQuitty *et al.* (1954), Peres (1962) and Wofford (1970). The horisontal treatment would appear to present an admirable experimental design for the study of the cognitive differences between rater groups, but no study of this type is known to the author.

These first four models have all comprised what may be regarded as the traditional, standard approach to data of this type. They all involve obtaining a matrix of data with respect to jobs, raters, and task/behaviour/trait descriptions, and using some form of factor analysis in order to study the dimensionality of the data and to draw some practical conclusions thereafter.

The remaining models, which are often restricted to one or a few studies, are attempts to extend this traditional approach *either* by means of applying some indirect method of scaling in order to obtain a correlation or covariance matrix which can then be factor analysed by traditional methods, *or* by means of considering the whole concept of factor analysis to be a restrictive one and applying some other form of data reduction, most notably latent profile analysis. It is to these extensions which we now turn.

Model 5 This model is similar to model 3 in that the basic structure is as follows:

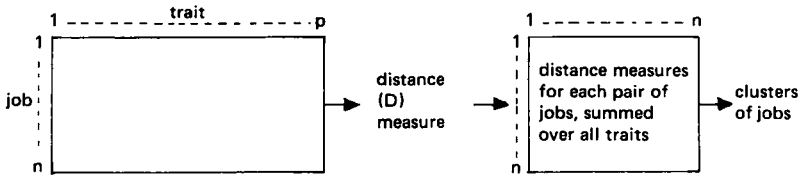


Fig. 3:7 Model 5: trait vs job – distance measure

The difference in methodology compared to model 3 is that the distance measure proposed by Thorndike (1953) is used instead of calculating the correlation between ratings for pairs of jobs. Expressed in the previous terminology this is a method of *comparative* judgement instead of *categorical* judgement. The distance measure applied in the study originating this model (Orr, 1960) consisted of stating for each job the extent to which the level of the trait required was possessed by different sectors of the working population as a whole. The distance between each pair of jobs with respect to each trait could then be determined and a measure of the total distance summed over all traits was obtained. The method proposed by Thorndike (1953) was then used to cluster the ratings into clusters of jobs.

A corresponding vertical data processing would be to have summed trait differences over jobs and to have applied the same cluster analysis to obtain groups of traits.

It is also worth pointing out that the distance measures could in fact have been used directly to provide a correlation matrix; this would however, hardly have substantiated Orr's claim that he was presenting an entirely new method for clustering jobs.

It is also obvious that model 5 could be used as a basis for generating a whole series of the types included in models 1–4, the only minor difference being in the method of data selection. A more fruitful methodological advance would have been to perform the distance ratings directly, instead of using a method of categorical judgement and then calculating comparisons between each pair.

Model 6 This model also involves a change in the method used to obtain the basic data, in this case similarity ratings being used as follows:

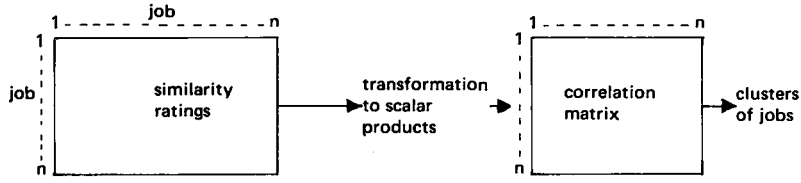


Fig. 3:8 Model 6: job vs job

This method involves comparative judgement of a direct type. A distinction may be made between comparing the jobs *globally* or with respect to specific *traits, behaviours* or *tasks*. The basic similarity ratings have to be transformed into scalar products for subsequent factor analysis. This will produce groups of jobs, either globally in terms of general similarity, or with respect to the specific variable(s) in question. The results in the case of global ratings would be similar to one of the possible methods of data processing presented in models 1, 3 and 5. The difference is that here there is a considerably lesser degree of structuring of the ratings. Instead of specifically stating with respect to which traits, behaviours or tasks the ratings are to be made, this is in the global case left to the rater himself. This is both the strength and the weakness of comparative judgement methods of this type. Their strength is that the rater is not constrained by the previous judgements of others as to what should and should not be taken into account in the course of a rating. The weakness is that they throw little light upon the cognitive processes of the rater or different rater categories. This makes it difficult to interpret the groupings of jobs obtained.

Furthermore, it is difficult to see how these global similarity groupings can have much practical value, other than as a very rough and ready measure. Without knowing the grounds for the similarity rating, which may of course differ widely from one rater to another, or from rater group to rater group, we have not really made much progress in analyzing the jobs in question. Studies of this type have been performed by Gonyea (1961), Gonyea and Lunneborg (1963) and Mårdberg (1972).

The other alternative is to make similarity ratings, for instance with respect to specific traits, and to compare jobs with respect to similarity on each specific trait. This will result in a whole series of clusters of jobs with respect to each trait and perhaps also with respect to each rater category. The opportunities of using data of this type in any meaningful way, other than as an exercise in methodology, appear to be very small indeed. Studies of this type have in fact been performed by Mårdberg (1971d) and by Mårdberg and Baneryd (1971), in both cases primarily for methodological purposes.

It is interesting to compare this model with the previous one. There the distance measures were summed over all traits before the jobs were clustered. In model 6, it is at least theoretically possible in a similar manner to sum the similarity measures over all variables, to perform a scalar transformation and to factor analyze the figures obtained.

A method which the author has been unable to find in the empirical studies presented in the literature, at least within the field of job analysis, would be to perform paired comparisons in order to determine on an interval scale where different jobs are located with respect to each given trait. The corresponding alternative, that of using the method of paired comparisons in order to rank traits with respect to a given job, is also an approach which would appear to provide considerably more information for practical use than the similarity measurements reported on here.

Model 7 For the sake of completeness it should be mentioned that one further study of this type, i. e. similarity ratings, is to be found in the literature, namely the following:

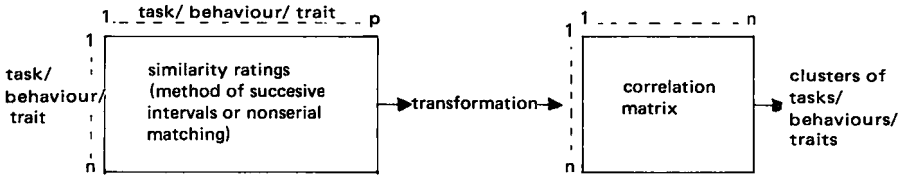


Fig. 3:9 Model 7: task/behaviour/trait vs task/behaviour/trait

In the actual empirical studies applying this approach the method of data collection has become more refined, but this is not a necessary requirement. In this model, only *one job* is the subject of investigation and the data are analyzed in order to determine the job dimensions involved. One such study by Brown (1967) produces clusters of *behaviours* and another by Schultz and Siegel (1964) clusters of *tasks*. The method produces more or less the same results as model 2, and it is difficult to see the advantages of this method other than as a methodological study for the comparison of different methods of data collection. One possible further use, similar to that of model 2, would be to use the technique in order to study the cognitive structure of different rater categories.

Model 8 The one new major attempt to provide a framework for the processing of job analysis data has been provided by Mårdberg (1971a, b, c). The basic model is similar to models 1 and 3, which are combined here in order to avoid unnecessary repetition. The model is as follows:

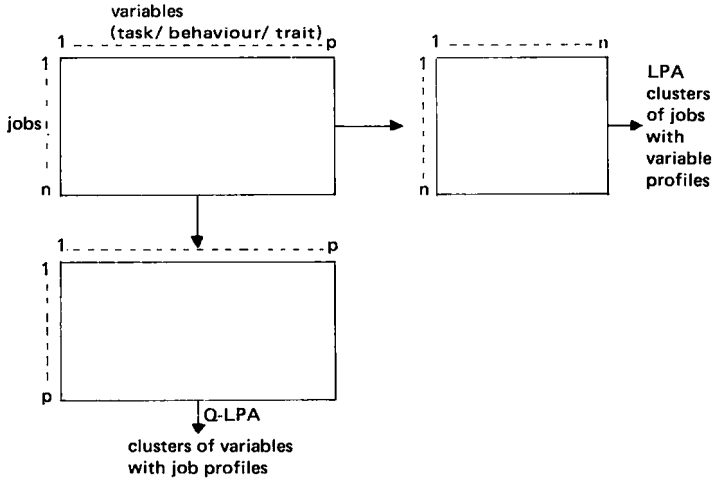


Fig. 3:10 Model 8: task/behaviour/trait vs job – LPA

The difference in this model compared with previous ones is in the method applied to reduce the dimensionality of the data. The number of raters may be one or more. When the number is greater than one, the model will assume a three-dimensional form, where it will also be possible to obtain clusters of raters.

A comparison of the models and an outline of applications

The table on page 63 showed various variables which may be used to structure job analyses. The models will now be examined against a background of these variables and conclusions will be drawn as to their applicability in various situations. In order to do so it will, however, be necessary to provide a brief outline of different practical fields where job analysis has been applied. These are presented in the form of different work hierarchy levels in the table below:

Work hierarchy level	Applied problems involving job analysis
Occupation	Vocational guidance, occupational therapy
Job	Job evaluation, classification, training analysis, personnel selection (blue collar)
Position	Personnel appraisal, organisational analysis, personnel selection (white collar)
Tasks (duties)	Organisational and method study
Subtasks, movements	Work study

Some of the applied problems (such as training analysis) arise at different levels, but they have only been mentioned once in the chart. Each level will now be examined in greater detail.

Occupation

The two major applications at this level are vocational guidance and occupational therapy; to this may also be added national manpower planning and the concomitant training and educational requirements. The basic application of job analysis at this level is to establish in the broadest possible terms the requirements made of job incumbents in different occupations, as well as the extent to which individuals who experience difficulty in obtaining jobs on the open market may be provided with some other form of meaningful occupation where the requirements are less stringent. Here the analysis is being performed at the occupation level and involves an almost unlimited number of possible occupations. With respect to the rating unit it seems reasonable that attempts should be made, in the first instance, to establish *tasks* and possibly thereafter to attempt to determine the corresponding *traits*. The group who are primarily interested in the occupational

structure as a whole are the Labour Market Authorities. As has been shown previously, much work has been done in the United States by the USES in their development of DOT. Similar work has been performed in Sweden through the Labour Market Board, although here the corresponding supportive research is conspicuous by its absence. In the United States, however, we have seen how both direct and indirect estimates of traits on the basis of task descriptions have provided satsifactory results. When performing analyses at the occupational level, it is important to have a number of raters who in different ways represent specialised knowledge. The task of reaching a consensus on the basis of their various assessments is, however, a complicated matter. With respect to possible clusters of variables, there are a number of different alternatives. The situation where individual occupations can be broken down into tasks and traits has already been mentioned. In addition it is possible to use a series of tasks, behaviours or traits in order to group together occupational families. Data of this type are valuable in manpower planning situations. The nature of the response will, as usual, hover somewhere between the judgement and response approaches; if the rating is regarded as being a type of response, there is no *prima facie* case for preferring any technique to any other. Similarly – with respect to data reduction – the choice of factor analysis, multivariate analysis or LPA will to some extent depend on the method of data collection, as well as on the assumptions as to the nature of the data and on the nature of one's concept of psychological measurement.

Returning to the models in the previous section, there is none which entirely fills the bill for measurement at the occupational level. The nearest ones in some respects are models 1 and 3, i. e. task, behaviour or trait vs job. In model 1 we are dealing with jobs, i. e. a lower level than occupations. This will of course simplify the ratings, and the use of task description and the processing using factor analysis to provide clusters of jobs would be the correct one. The replacement of occupation by job in model 3 makes the application of this model less satisfactory; we have also seen in the previous section how this step can be performed as a secondary stage in the ratings. Models 5 and 6, i. e. trait vs job or job vs job, have also some relevance, the rating units here being holistic ones of the respective occupations. It would be possible to use these models as a first step, perhaps together with different categories of raters, as a preliminary to establishing the most important variables when grouping occupations into families. Model 8 may of course also be used, applying LPA to the grouping of jobs (occupations).

There is one specific reason why detailed measurement work has seldom been performed at the macro-level: the possibility of obtaining raters who have a broad overview of a wide range of occupations is limited and, in addition, the very concept of an occupation is often vague. It is likely that raters tend to *conceptualise* in terms of specific jobs rather than by occupations as a whole. Furthermore, the occupation of being a building worker in general may in fact at that level involve a very high degree of overlapping job requirements with, say, a worker in the metal industry.

Occupations may not, in other words, have task or behaviour profiles which be clearly distinguished from other occupations and even if they have, these may disappear when arriving at the trait description level.

Finally, some mention should be made of *vocational therapy*. Here the spotlight tends to be focused more upon the individual who is experiencing difficulty in finding *any* occupation which might fit his or her specific talents or, perhaps more frequently, important restrictions with respect to these talents. In order to make a correct assessment at the individual level, it is necessary to know which traits the person in question possesses and in what degree, as well as the extent to which these trait requirements are present in different occupations and sometimes – more specifically – jobs. In this case we are interested in studying occupation by occupation at the trait level in order to compare the person in question with the specific trait requirements. This gives rise to an important question which will be dealt with in chapter 4: the extent to which it is possible to apply a *compensatory, conjunctive* or *disjunctive* model to the placement decision. As before, there is considerable scope for further research within this field.

Job

Once again job analysis has a number of important applications at this level. Of those mentioned in the above table, classification and personnel selection will be reserved until chapter 4. Here attention will be devoted primarily to *job evaluation*.

Briefly the purpose of job evaluation is to establish the *difficulty* of different jobs, primarily with the intention of relating difficulty to wage structures. These studies were originally performed for blue-collar jobs but they have also been applied to white-collar ones. Work is currently in progress in Sweden which aims at combining the two systems into a single one covering the whole of the labour market (Jonasson, 1970). The time schedule for this research (over a decade) indicates the complexity of the task. Elsewhere (Mabon, 1972), the author has made a critical review of some of the basic job evaluation techniques at present in use in Sweden. Applying the variables on page 63, this is of course measurement at the *job* level. When it comes to the rating units, the variables used in the leading white-collar and blue-collar systems are a disparate mixture of tasks, behaviours and traits, with environmental dimensions thrown in for good measure.

As job evaluations are often made by committees (representing employer and employee) the ratings will tend to be performed by *heterogeneous* groups. Research has shown that employers and employees sometimes have widely differing interpretations as to difficulty, even when the dimensions to be rated are predetermined (cf. Dahlbäck, 1966; Tiffin and McCormick, 1969).

The object of the rating is to cluster jobs into categories with respect to difficulty. This clustering presents a number of basic methodological problems. It is common to distinguish between the following possible alternatives:

	Total rating	Part rating
Direct (job to job) comparison	Ranking	Factor comparison
Indirect (job to scale) comparison	Classification	Points system

The method of *ranking* is in fact a version of model 6, where jobs are compared holistically with other jobs, in this case with respect to difficulty. The method has been little used, presumably because of the difficulties involved in arriving at a holistic definition of *difficulty*, especially when the rating groups may be expected to be cognitively heterogeneous.

The method of *classification* does not fit into the models presented in the previous section. While the method of ranking involves a method of *comparative* judgement, classification is a form of simple *categorical* judgement which permits little analysis. A difficulty scale is established, and jobs are simply placed in a specific category. Different categories of raters would be expected to have different ideas as to these placements, but in practice this will tend to be solved by negotiations between the two or more parties involved. The method of classification has in fact been widely used in the job evaluation of white collar workers, although the trend in Sweden is to attempt to encompass all job evaluations in a points system (see below).

The method of *factor comparison* is in fact an example of model 1 or model 8, depending on the method of data reduction, the grouping taking place with respect to jobs, not to duties. As part of the ratings involves traits of job incumbents, the model applied is also to some extent model 3. The interesting feature here is the method by which the data are obtained. The method used is one of comparative judgement, such as paired comparisons. This can be done by taking one variable at a time and by comparing pairs of jobs with respect to the extent to which one makes greater requirements of the job incumbent than the other. This leads to a ranking of jobs, at the interval scale level, with respect to the variable in question. The problem involved in this type of rating is that it is usual to have a number of different aspects of difficulty included in the rating. Each of these would have to be rated in the same manner, which would give a whole series of scales, with the jobs in question placed at different levels and in different rankings for different variables. Combining these into a total job evaluation would involve explicit or implicit assumptions about the relative importance of each of the variables. In the last instance this would mean that the system was being transformed into a points one.

The *points system* is the method which has won the greatest degree of acceptance at the blue-collar level. The distinction between the points system and factor comparison is the same as that between classification and ranking; the former is a *categorical* and the latter a *comparative* judgement. The points

system involves quite simply taking a set of difficulty variables as given, and for each job in question assigning a specific number of points for each difficulty variable. The range of points for each sub-variable is determined outside the rating system. This makes it difficult to perform any form of analysis of the value of the total rating, with the exception of studying inter-rater agreement and the contribution of each individual sub-variable to the total number of points obtained. Studies of this type have been performed (Dahlbäck, 1966; Tiffin and McCormick, 1969) and the results obtained have been singularly consistent. Employers and employees have different cognitive structures and the number of subvariables used has often been far in excess of what could be interpreted as the number required, the various sub-ratings presumably deriving from a massive *halo-effect.* One other possible interpretation is that most people are quite simply unable to split up an overall rating into a number of components in the manner required by a points system. This viewpoint is also relevant in chapter 5 where multidimensional criteria are discussed.

One of the basic problems in measuring job difficulty is the concept itself. An important field of study would be the dimensionality of this concept for different categories of raters, perhaps by applying some form of model 6. The present points systems make at least one basic distinction between difficulties arising with respect to the *duties* and difficulties arising because of the *environment.* A job may be difficult because a high degree of judgement and initiative may be required; it may also be difficult because it has to be performed in a temperature of 30°C or with a noise level of 90 dB. When working conditions improve, job "difficulty" is reduced, so that lower points will be awarded on these subvariables. Some of the job difficulty factors may be regarded as economic compensation for unsatisfactory job design. As jobs involving *low* demands with respect to duties are often identical to those with *high* environmental demands, while others with *high* duties demands have *low* environmental demands, this will tend to bring wages and salaries into closer parity. The vital factor is the weights which are attached to these main categories of job difficulty.

In conclusion, let us turn to the question of methods used for data reduction. Factor analysis can and has been used to obtain clusters of jobs. Mårdberg (1972) has, however, shown that the method of LPA provides profiles for different jobs which can very profitably be applied to job evaluations.

Training analysis: The purpose of this type of analysis is to determine the training requirements for a specific job, in the first instance with respect to specific individuals who are already working in the job, or who aspire to it. This is a straightforward application where the information required can normally be obtained by applying model 4 or perhaps model 7. By processing model 4 with respect to clusters of raters, it may be possible to draw conclusions as to expectations from different sectors of the job surroundings, and perhaps the extent to which these differences should be taken into account when carrying out the training. The testing of likely job incumbents

and the determining of their specific training needs is similar to the classification situation (see chapter 4) and also gives rise to problems involving compensatory models.

Position

Here we will be interested in two applications, namely *personnel appraisal* and *organisational analysis.*

The issue in personnel apparaisal is to assess how well one or more employees have succeeded in one or more jobs, often with a view to giving a salary rise or promotion. The basic measurement problem is one which has been encountered in chapter 2, i. e. whether a rating of this type should be work- or worker-oriented. Obviously it is the worker/employee who is being rated, but it is a matter of some dispute to what extent the rating should be related to specific job requirements or more general traits. Applying the variables on page 63, it will be seen that this is a rating at the position level for each individual, although to some extent it is also a job rating in that several people are often involved. As for the rating unit, many of the personnel appraisal systems in use today involve combinations of ratings with respect to traits and behaviours, and sometimes even ratings with respect to how well duties have been performed. The ratings are often performed by one or more superiors, i. e. usually a homogeneous group. The variable clusters are groups of individuals classified with respect to different rating patterns.

As to the nature of the response and the rating method applied, it is quite feasible to apply to personnel appraisal the same fourfold table as used when analysing job evaluation. The difference this time is that it is the *individual job incumbent* who is being rated and not the *difficulty of specific jobs.* The assessment is also more complicated in that the job incumbent rating is not a rating of the job incumbents *in general,* but instead with respect to how well they have succeeded in a *specific* job and position. This can mean that a person may be judged to be unsatisfactory because he is *too good* for the job in question, or because he is *not good enough.*

Let us once again examine the alternatives in the fourfold table one by one.

Firstly, the method of *ranking* is once again a version of model 6 where groups of individuals are arrived at by means of a method of comparative judgment, such as paired comparisons.

Secondly, the remarks previously referring to job evaluation are also relevant with respect to *classification.* The rating method is categorical and provides little scope for further analysis. The method of forced distribution involves the use of this method. The scaling level cannot rise above the ordinal level.

Thirdly, the method of *factor comparison* involves rating the job incumbents with respect to a number of variables. The author is not aware of this method actually being applied in practice: if applied, the result would be a rating for each individual on each specific trait, behaviour or duty. Matters

relating to weighting and aggregation are as relevant here as in the section on job evaluation.

Fourthly, the *points system* has also been widely applied in personnel appraisal. Once again there are problems arising from the choice of variables, the scaling level of the categorical judgements and the weighting of sub-variables required to provide a total score.

Organisational analysis is regarded here as the analysis of individual positions and – to some extent – jobs in order to determine job content for interdependent employees. The scope of this study does not permit a detailed analysis of the uses of job analysis in this respect. Suffice to say that the systems approach presented in chapter 2 has its obvious applications. The central models have also a number of uses in organisational analysis. Model 1 may be used to give clusters of duties which should suitably be included in specific positions. Similar considerations apply to model 3. At the macro-level the application of models 5 and 6 could provide valuable information as to the basic principles according to which the company should be organised. As was pointed out on page 31 the concept of a job or positions is a relative one, to some extent dependent on surrounding jobs and positions, and on present job incumbents. The results obtained from the data processing outlined in the models above could in the present situation give rise to a discussion as to the possible realignment of job or position boundaries.

Tasks (duties)

We are now rapidly approaching the micro-level of job analysis applications. Model 2 is primarily involved here (accepting that each specific job or position has been clearly defined at a higher level). The clustering of duties may be of some use for arriving at methods of performing the whole series of duties in the most efficient manner.

Subtasks, movements

The last major application of job analysis is in connection with work study, the purpose of which is to determine the time taken for different types of subtasks or movements. These times are then used to provide a basis for determining piece-work wages. This is a subject which can only be outlined here. The technique cannot be analysed within the framework of the above models, but has nevertheless been included for the sake of completeness and because of its connection with other previous issues. One of these issues is the extent to which there exists the "the one best way to do a job". The dynamic systems approach mentioned in chapter 2 suggests that this is not necessarily the case. Furthermore, the assumption that the *sequence* of tasks will not affect the time taken for each sub-task is a questionable one. In addition, work study systems ignore individual variation; a 60-year old and a 30-year old may take the same time to perform some routine manual task, but the

exertion of the latter may be very much greater. In conclusion, although the people who tend to perform these ratings are homogeneous, i. e. work study engineers, the inter-rater reliability of sub-task times is surprisingly low (cf. Abruzzi, 1956).

Chapter 4 Job analysis and the selection procedure

In chapter 2 and 3 an analysis has been made of many different ways in which jobs and job requirements have been described. One of the aims of this study is to investigate the extent to which this research has been of value as an introductory stage in the selection procedure.

Briefly, the traditional selection model consists of the six stages shown in figure 4:1.

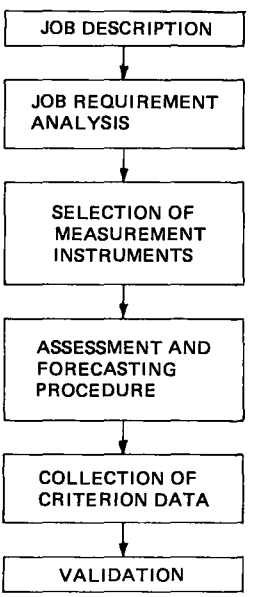


Fig. 4:1 Stages in a selection procedure

It will be shown in this chapter and in the following one that there are a number of interesting measurement problems arising at each stage of this procedure. The first four levels will be dealt with in this chapter and the two remaining ones will be treated in chapter 5.

Job description

This subject has been dealt with extensively in chapters 2 and 3. It was established that in normal circumstances it is advisable to describe the duties involved in a job before attempting to estimate the traits required to perform

this job. A great number of studies, many of them specific to one job or occupational category, has been performed over the years. As yet there is no generally agreed framework within which different types of jobs may be described. Between the physiological level of Fleishman (1967) and the clerical and management questionnaires of McCormick and Prien there is a large gap. Restricting ourselves purely to this first level, the measurement problem of greatest interest is the extent to which *reliable* job descriptions can be obtained using different methods. A considerable amount of research into this matter has been presented in chapter 2 and although different methods have been shown to vary in reliability, many of the specially derived check lists have been shown to have satisfactory job description reliability, at least when categories of raters are considered separately.

Job requirement analysis

Once again a large amount of research on this matter has been presented in previous chapters. The measurement problems are, however, worthy of further consideration. If the job requirement analysis is based on a job description, do reliable job description methods provide a suitable basis for assessing the necessary traits? What is the reliability of trait ratings made on the basis of what is to be hoped are reliable job descriptions? Relatively few studies have examined this question and those which have compared different rater categories such as peers and superiors have shown that the trait ratings, while attaining a fair degree of reliability *within* each category, tend to differ widely *between* categories, for reasons which have not been studied systematically in the literature. We shall return to this question in chapter 5. The question of the number of traits to be included is also a difficult one. If the trait ratings are based on satisfactory, extensive job descriptions, there should be every opportunity of determining a comprehensive list of relevant traits. The actual process whereby this is done is, in the opinion of the author, an interesting one which has never been studied in detail. If, on the other hand, an attempt is made to determine job trait requirements directly, without the benefit of a job description, as was done in a number of factor analytic studies reported in chapter 2, there is also the difficulty of determining the traits to be included in the study. The structure of a factor analysis cannot extend beyond the variables used in the analysis.

A further question related to the later stages of the selection procedure is the extent to which it is meaningful to attempt to determine requirements on traits which in general are exceedingly difficult to measure and which would be impossible to rate with any satisfactory reliability, if some of the more primitive selection instruments were used. It is all very well to establish that a job requires considerable mental resilience, but this will be of little use if the selection procedure is based on an interview performed by a relatively unqualified personnel employee, who has to make an informal assessment of this intangible variable. A more general issue arises concerning the rating of traits. Many researchers have maintained that statistical methods tend to

conserve theories; new theories require and are in some cases generated by these new methods. This is true of trait psychology in general, which has to a large extent been based on the method of factor analysis. If latent profile analysis is preferred, then instead of determining specific traits independently of each other, the consequence would be that job requirement analysis would be based upon *profiles* of traits rather than individual traits. Once again there is considerable scope for research in this area. Mårdberg states emphatically that the trait approach is more or less worthless, which is nothing less than the death sentence for traditional selection procedures. In the opinion of the present author, Mårdberg has not as yet entirely proved his case: there does not at present appear to be any *a priori* method of establishing that the theories deriving from LPA are better than those deriving from factor analysis. A final difficulty with respect to the rating of traits will be expanded upon in the final section of this chapter; in many cases it is not sufficient to determine trait requirements, one must also attempt to judge the extent to which success in the job situation varies with these traits. In some circumstances this will increase *monotonically*, while in other cases success on the job may be restricted to a limited range along the trait.

Summing up these two first stages in the selection procedure, a complete research programme on measurement in the selection procedure should aim at studying the reliability of job descriptions and the reliability with which these descriptions can be translated into traits. It is not necessary to consider validity at this stage.

Selection of measurement instruments

Having determined the job requirements, the third stage of the selection procedure is the choice of instruments which will be used to measure the applicants on the traits in question. There is an enormous literature dealing with this subject (see for instance Miner, 1969, and Mabon, 1972), which for reasons of space must be dealt with selectively. The measures most commonly available are: school and further education marks, job and other references, anamnestic data, the interview, observations in group situations and psychological testing.

These first five methods will not be dealt with here – only the last one. Let it suffice to state that a number of studies and commentators have recorded grave doubts as to the reliability and validity of these various assessment techniques.

Before examining *psychological testing*, it is perhaps worthwhile to briefly recapitulate the purpose of this whole section. The first two stages in the selection procedure are concerned with job descriptions and job analyses. The outcome of these two stages has traditionally been a number of traits believed important for job incumbents. The purpose of the third stage is to determine the extent to which the personnel administrator has techniques available to assess these traits for job applicants. The five methods previously mentioned all make some sort of direct or indirect contribution to the determination of applicants' traits. The final method, that of psychological testing, is obviously

of a completely different nature to those previously presented, in that in this case we are specifically able to measure directly in trait terms as opposed to making direct or indirect assessments of these traits. The question now arising is the extent to which available psychological testing methods can be used to measure the traits shown to be important for a specific job in the course of a job analysis.

A detailed examination of this question would lead us well beyond the scope of the present work. In Sweden, as in a number of other countries, tests have been developed for the measurement of, *inter alia*, ability, aptitudes and personality. The ability tests are normally concerned with the measurement of specific traits such as *word fluency* or *planning ability* and can to that extent be said to provide measures of traits derived from job analyses. The measurement of abilities in Sweden and elsewhere has been greatly influenced by the work of Thurstone and his factor analytic studies, giving rise to seven basic ability factors (1938). The vast majority of tests available in Sweden are for instance concerned with the measurement of one or other of these factors. With respect to intelligence, it may be said that the measurement methods available fulfill normal reliability requirements (i. e. at least 0.90). This assumes that one accepts the Thurstone trait approach as a correct one. As has been mentioned previously, Mårdberg and others have been extremely sceptical of the trait approach, which they regard as a direct consequence of the application of factor analysis. As latent profile analysis is often a better method for processing data within the field of the behavioural sciences because of less demanding data requirements, a whole new school of testing could develop around the idea of ability *profiles* and other types of profiles, instead of around the idea of individual traits. Mårdberg is aiming in that direction but this mammoth task is as yet far from completion.

Within the framework of a trait approach, little criticism need be made of ability testing with respect to its test-retest reliability. Even in the case of practical aptitudes the established assessment procedures for testing motor skills may be said to have reasonable reliability.

The problems of psychological testing are greater in the area of personality tests. One of the basic problems is that the responses to statements or other stimuli constituting the test tend not to be a measure of the respondent with respect to certain personality traits. Rather they are a measure of his idea as to how one *should* respond in order to create a good general impression on the basis of general norms as to ethical behaviour, values and attitudes, or possibly the respondent's impression as to how an employee in the job for which he himself is seeking employment would tend to respond to the stimuli in the personality test. Faking on personality testing in this manner does not necessarily mean that the application of personality testing will be unreliable. That validity cannot be high under these circumstances is obvious.

The whole issue of personality testing has recently been the subject of public enquiry in Sweden, see Ekvall (1971), and the findings point to the fact that for ethical and other reasons the application of personality testing within the public sector in Sweden will be restrictive.

Assessment and forecasting procedure

The situation after the three first stages of the selection procedure is as follows. On the one hand information has been obtained about job requirements from the job analysis, on the other hand a variety of techniques have been available for the assessment of applicants on the corresponding variables. The following variables determine the forecasting model to be adopted.

- the number of jobs analysed, J_i, i = 1 . . . . . n
- the number of traits/requirements rated for each job, R_j, j = 1 . . . . . p
- the number of applicants available, A_m, m = 1 . . . . . r
- requirement level or score on R_j, R_{jk}, k = 1 . . . . . q
- thus, $A_m(R_{jk})$ = the score k obtained by applicant m on trait j. $J_i(R_{jk})$ = the level k on trait j is required in order to perform job i.

The possible combinations of all of these variables give rise to hundreds if not thousands of possible decision situations. The aim of the following analysis is to present a selection of the central problems involved. The alternatives will be built starting from the simplest ones.

a) *One* job to be analysed, *one* position available, *one* global requirement rated for each individual. We may commence with the trivial case of a single applicant. Assuming that the regression of criterion variables (CV) on forecasting variables (FV) is linear and monotonically increasing, the decision rule is simple.

If $A(R_{1k}) > J(R_{1a})$ select A otherwise reject. (The second subscript a, b, c etc refers to specific values of the requirement in question.)

If, however, the relationship between criterion and forecasting instruments is known or assumed to be curvilinear, the decision rule will be

If $J(R_{1a}) \leqslant A(R_{1k}) \leqslant J(R_{1b})$, select A, otherwise reject. Both of these situations can be represented with the aid of simple illustrations.

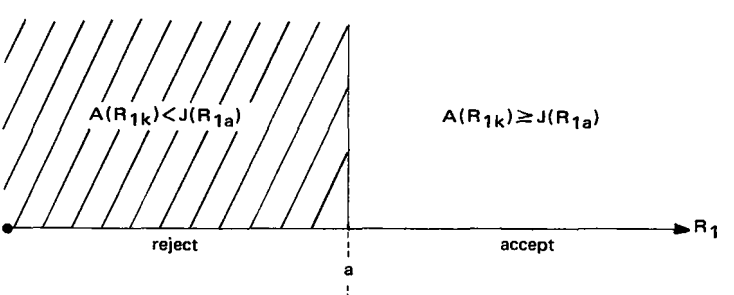


Fig. 4:2 The simplest selection situation

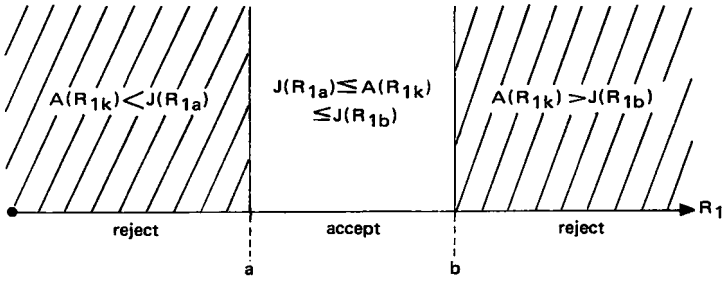


Fig. 4:3 Selection assuming a curvilinear relationship between CV and FV

Let us now extend the case to more than one applicant while still referring to only one position. In the monotonic case the applicant with the highest score on R_1 is selected subject to the proviso that this score is higher than the minimum level for acceptance. The picture becomes more complicated for situations of the type shown in fig. 4:3. Here the decision rule will be the random selection of any A with R-scores between R_{1a} and R_{1b}. There is no way of determining which of the applicants in this category is most suitable.

Let us now gradually remove the requirements one at a time with respect to the various variables involved:

b) Assume that there are a number of jobs to be analysed and that one position is available for each job. There is still only one total global requirement to be rated. Let us also assume that $r > n$, i. e. the number of applicants is greater than the number of jobs.

Here the situation rapidly becomes extremely complicated. Let us initially assume that the job requirements are mutually exclusive, i. e. that those who are suitable for one job will not be suitable for any other. The situation is shown in figure 4:4.

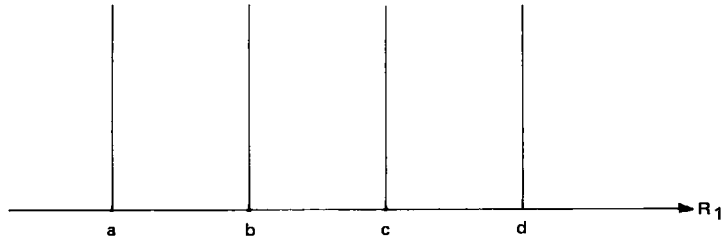


Fig. 4:4 A classification model based on unidimensional data

The decision rule in this case is to select at random one applicant from each area along the requirement dimension, in other words one applicant where $A(R_{1k}) < J(R_{1a})$, one where $J(R_{1a}) < A(R_{1k}) < J(R_{1b})$, one where $J(R_{1b}) < A(R_{1k}) < J(R_{1c})$, and so on. There is no rule for determining which of the applicants conforming to each of these requirements should be selected. The solution becomes even more problematical if there is some degree of overlapping with respect to

job requirements. Some jobs may need employees where $A(R_{1k}) \leqslant R_{1c}$, others $R_{1a} \leqslant A(R_{1k}) \leqslant R_{1d}$, or $A(R_{1k}) \geqslant R_{1c}$. In this case we can proceed along familiar lines, assigning one individual at random who fulfills the specific requirements to each job. One further proviso: individuals who could be assigned to more than one job, should not be so assigned until it is established that there is at least one applicant for each category. This rule could also be expanded so as to ensure that individuals belonging to this category would have the same random chance as others to be assigned a job.

These final considerations become more important when the number of positions for each type of job available is greater than one. In the simplest model with no overlapping, the situation will be this: if there are more applicants than positions in each specific job, then the required number should be selected randomly from those suitable; if there are less than the required number then all should be assigned to that job.

If the model is an overlapping one, the decision rules would be: Starting with the greatest demands, are there a sufficient number of applicants who satisfy *only* these requirements? If yes, these should be assigned to the number of positions available for the job in question. If no, can the numbers be made up by adding those who are also suitable for some other job? If the answer is yes, do so and continue to the next job requirement group. Can the number of positions available here be filled by those who are only suitable for this position *and* those who are also suitable for more qualified positions but who have not been assigned to them? If yes, select randomly from these two groups; if no, can the numbers be made up by those who are qualified for this job as well as for less qualified ones? By continuing in this way along the whole length of the continuum, justice will be done to all of the applicants as well as it being ensured, as far as is possible, that each quota, i. e. the number of positions available for each job, is filled.

c) So far we have restricted ourselves to a single global assessment of the job requirements and the extent to which individual applicants meet these requirements. Within this framework it has been possible to build up fairly simple models indicating how personnel administrative decisions can be made. Much of what has been studied in previous chapters has concentrated upon the extent to which jobs and individuals can be analysed and described using a number of different traits, or possibly profiles, and not merely a global assessment. It is to this area which we now turn.

Let us once again begin with the simplest model of this type, i. e. *one* job to be analysed, *one* position available, and initially, in order to permit a graphic presentation, two requirements to be rated. If we have only one applicant the position is as follows: there are a number of possible ways of formulating the requirements to be satisfied by the applicant; the deterministic models for categorical data presented in chapter 1 distinguish between *conjunctive, disjunctive* and *compensatory* models.

The first two, which are mathematically equivalent, can be formulated as follows: if $A(R_{1k}) \geqslant J(R_{1a})$ and if $A(R_{2k}) \geqslant J(R_{2b})$ then select A, otherwise not.This may be shown graphically in the following figure:

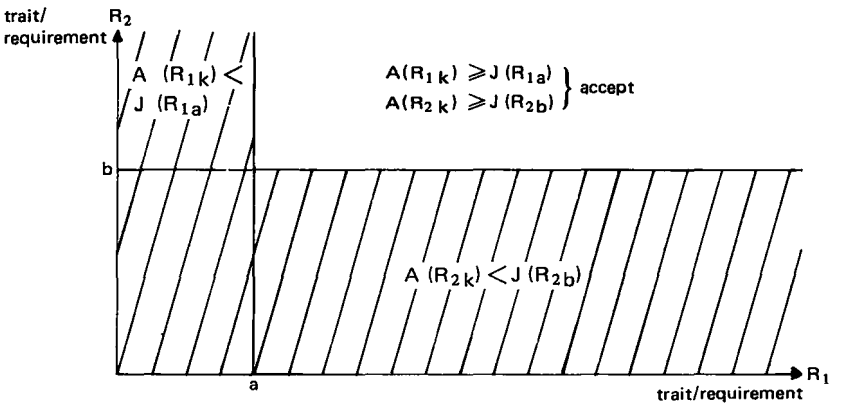


Fig. 4:5 A conjunctive selection model

In a compensatory model, an especially good performance on one of the variables may compensate for a slightly poorer performance on the other one. In mathematical terms this is $A(R_{1k} + R_{2k}) \geqslant J(R_{1a} + R_{2b})$. This can be illustrated graphically (fig. 4:6):

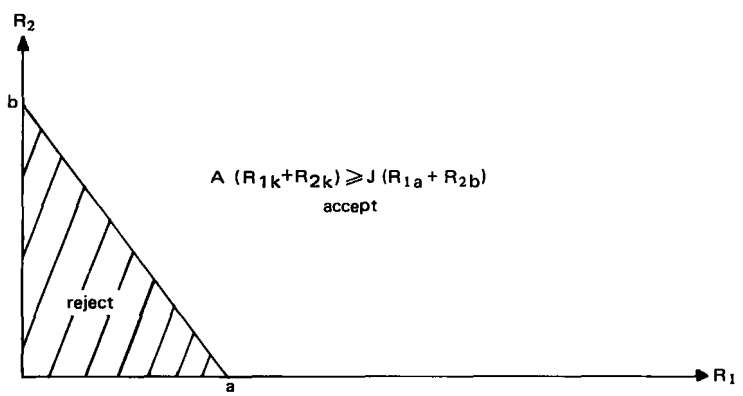


Fig. 4:6 A compensatory selection model

It is of course possible for the compensatory model to be formulated in more complicated curvilinear terms, but this will not be done here. A further extension of the model is to assume a curvilinear relationship between performance on the forecasting variables and later criterion variables, as was done previously in the case of the unidimensional models. This could give the following decision model (the prefix J is omitted):

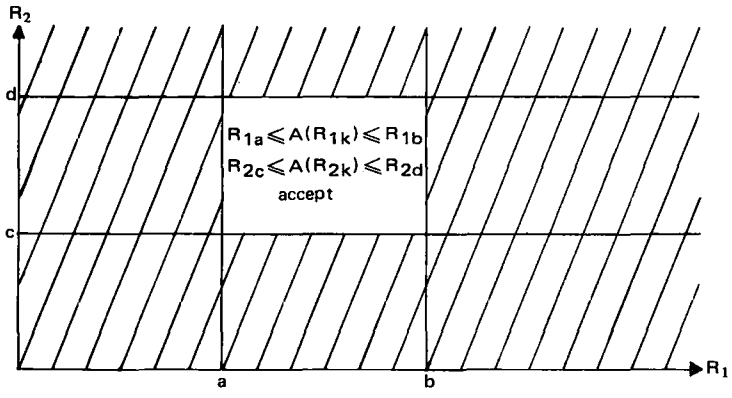


Fig. 4:7 A decision model assuming curvilinear relationships between PV and CV

Models of the type shown in figure 4:7 are in fact beginning to approach what might be called profile thinking with respect to the job in question. This point must be developed as it will be of considerable importance for the further analysis of the various recruitment models.

We have seen in chapter 2 how traditional methods for performing job analysis normally give rise to a number of factors or traits which are assumed to be more or less independent of each other. The requirements which the job in question places on the incumbent can then be determined. One way is to test already successful job incumbents with respect to the traits in question, and perhaps to compare these results with those obtained by less successful job incumbents. Studies of this type can then provide a basis for deciding the extent to which the various models sketched above can suitably be applied. The data obtained from present job incumbents provide in other words a form of criterion data.

The position when *latent profile analysis* is used is a different one. Here a number of profiles with respect to the traits required are obtained and different jobs are assigned to different profiles on the basis of probability calculations. As has been noted on page 56, this assignment and calculation of probabilities is based on the total number of traits measured, irrespective of the possible varying relative importance of these different traits. A job may thus have one or two extremely important traits in the vicinity of one trait profile, but the job itself could be assigned to another different profile, on the basis of a number of rather unimportant traits being located near the the latter profile. Ways have been determined to establish the goodness of fit of any specific profile solution, but much work remains to be done before these will be entirely satisfactory, statistically speaking.

Let us now assume that job analysis data have been obtained by means of latent profile analysis. Ideally, a preliminary validation of the model would be obtained by testing to what extent job incumbents who are successful conform to the profile in question, and to what extent they are poorer if they

deviate from the profile with respect to one or more traits. Studies of this type are at present unknown in industrial psychology and a theoretical discussion will have to suffice at present.

For the sake of permitting a graphic presentation we will continue with the two-dimensional model, the "profile" of the traits for the specific job in question then being reduced to a point. The actual values obtained for the job in question before it was assigned to the profile in question may in fact be slightly different, and these have been denoted in the figure by means of an oblique dash above the letters, see figure 4:8.

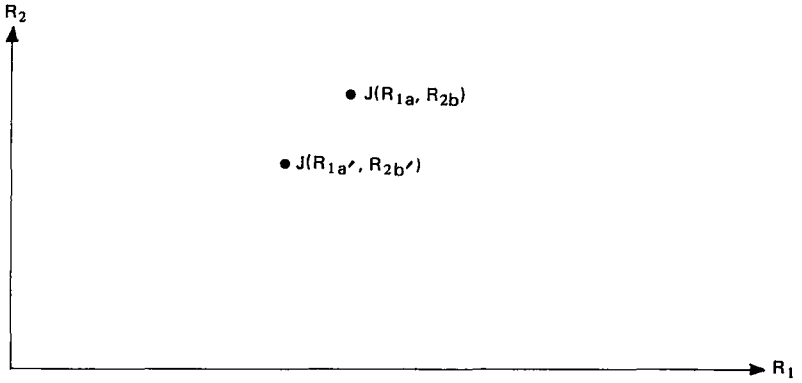


Fig. 4:8 Job requirements according to latent profile data

The question now arises as to how latent profile data of this type should be used in a selection decision process. To what extent is an applicant allowed to deviate from the exact values of the profile, or perhaps even the actual obtained values ($R_{1a'}$, $R_{2b'}$), and still be regarded as suitable? It would seem that the profile approach excludes the possibility of a compensatory model and there is considerable doubt as to how a conjunctive or disjunctive model would be applied. Mårdberg's work, which to date has concentrated upon the job analysis side of latent profile analysis, has little or nothing to say about practical applications of this type in the recruitment situation. These problems become aggravated when our present models are extended to the classification situation with a number of jobs involved.

Up to now we have assumed only one applicant and only two dimensions. The number of dimensions can be extended without substantially changing the nature of the problems involved and the types of decisions with which the administrator is faced. Additional problems arise when we have a number of applicants and a smaller number of positions available. In the conjunctive and compensatory models it is sufficient to establish the number of applicants lying within the "accept" area. If this is less than the number of positions available, then all of these applicants should be selected. If the number of applicants meeting the requirements is greater, then further decision criteria may or may not be added. The simplest rule would be to select randomly

from all those found to be suitable; assuming monotonic relationships, those who are much better than the minimum requirements may be regarded as more valuable than others. In both the compensatory and the conjunctive model the rule would then take the form of selecting the largest $I(A_1 + A_2)$, and then continuing to the next largest until the quota is filled. In the case of the curvilinear model shown in fig. 4:7, it is unlikely that any such model would be of value, and here it would be a question of resorting to some form of random selection among those found to be suitable.

When the latent profile model is applied, it is at the present stage of development difficult to present any sort of meaningful decision rule. If one assumes that each profile trait is equally important, which is one of the basic tenets of LPA, the decision rule could be to begin by selecting those with a minimum distance from the ideal profile (ignoring signs) and to continue until the quota is filled. Note that there is nothing in the model showing when a deviation from a single profile may be regarded as disqualifying for the job in question.

d) Let us now remove the final restriction, that of the number of jobs being investigated simultaneously. This brings us to the model normally known as *multivariate classification.*

The issues involved can be shown most clearly by means of a two-dimensional model, in this case possible assignment to two or more different jobs. This is in fact the multidimensional version of figures 4:3 and 4:4. It is also similar to the models in fig. 4:5 and 4:6, with the important difference that some overlapping may take place.

Assuming that there is no overlapping with respect to job requirements we have the situation shown in fig. 4:9.

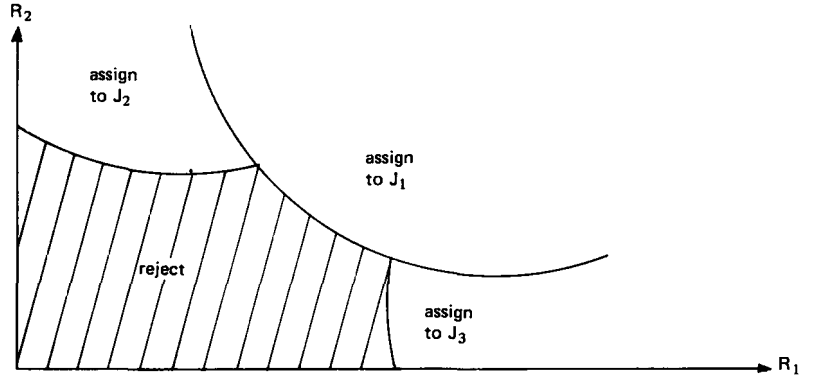


Fig. 4:9 Multivariate classification, no overlapping

Assuming that the number of positions is less than the number of suitable applicants within each area, it will be necessary to devise some system of random or other quotas in order to obtain the required number. Models of this type are common in clinical psychology, where different types of

treatment are under consideration, and in educational psychology where pupils are assigned to different courses and streams.

The situation may also be more complicated, in that a certain amount of overlapping may take place as in figure 4:10.

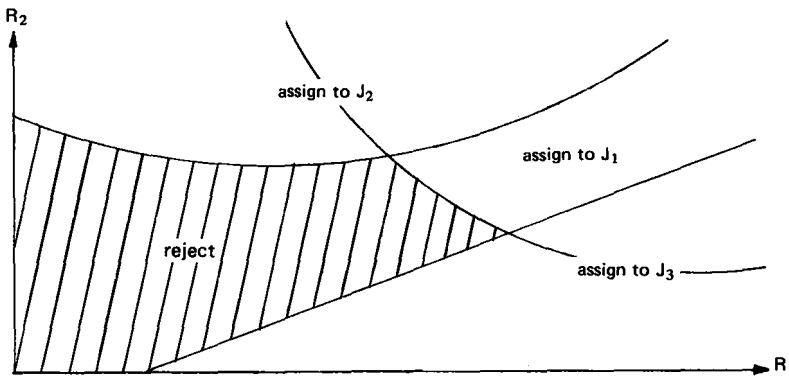


Fig. 4:10 Multivariate classification, with overlapping

The priority rules sketched in previous sections are also applicable in this case and will not be expanded upon here.

The analysis so far in subsection d) is based upon the traditional trait approach. Turning to the alternative, that of LPA, we are once again faced with a number of problems which have not as yet been solved satisfactorily. Here we will have a number of different points representing two-dimensional profiles as in figure 4:11.

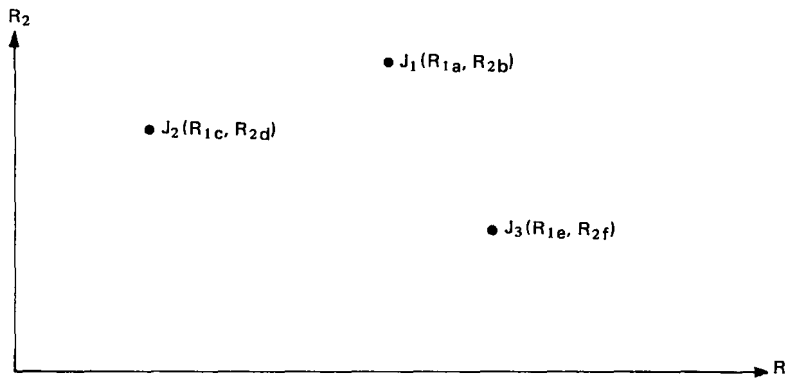


Fig. 4:11 Jobs with two-dimensional requirement profiles

Once again we may proceed as before. The decision rule would be that each individual is assigned to the nearest ideal profile. As before we have no rule saying how far one may be from a profile and still be able to perform the job. In the multidimensional model we have the further complication that the

number of positions available for each job is assumed to be limited. This will mean that some applicants who are very close to one profile may be assigned to some other profile (and job) because the quota for the first job is filled. As yet no one has attempted to discuss the consequences of applying LPA in situations such as this.

This brief review has served to illustrate some of the more basic measurement problems which arise when an attempt is made to match job analysis data and measurement of individual traits. Returning to the various job analysis models presented in chapter 3 it will be seen that not all of them are directly applicable to the problems dealt with above. The models shown in this chapter have for instance assumed that it is possible to make fairly accurate assessments of what traits are really required in a specific job. As will be shown in chapter 5 when dealing with the criterion problem, the divergence among different rater groups is very great. The models which are of greatest value in the actual recruitment situation are model 3, where a basic matrix is obtained for n jobs and p worker-oriented traits; model 4, assuming that the ratings are homogeneous; to some extent model 7 and finally the latent profile model 8. Some of the others can be used to provide preliminary job description data, while others are mainly used for the analysis of cognitive job perceptions.

Chapter 5 Criterion Measurement and Validation

The two concluding stages of a selection procedure are the collection of criterion data and the validation of the whole procedure. We will commence this chapter with a general study of the measurement problems arising in connection with criteria and continue with a study of how one may validate the selection procedure. It is the contention of the author that this validation procedure is even more complicated than is often supposed.

Criterion measurement

Before examining the question of criteria, let us review the stage reached in this study. Job descriptions have been performed and, if necessary, these have been translated into job requirement traits. It has then been possible to measure individuals on the same traits and, with the aid of one or other of the selection models dealt with in chapter 4, it has been possible to assign applicants to one or more jobs. The final stage is then to establish – at a later date – the extent to which the applicants have been successful in the job or jobs to which they have been assigned. The determination of the extent to which these applicants have been successful is the basic criterion measurement problem. It is by no means clear which dimensions should be used to assess this success. Tiffin and McCormick (1963) distinguish between four basic types of criteria: *performance criteria*, such as the amount of work produced; *subjective criteria*, normally taking the form of ratings made by superiors; *physiological criteria,* such as heart rate and blood pressure; and finally *accidents,* this final criterion throwing light also on job design and on the extent to which some people are less accident prone than others.

Thorndike (1949) distinguishes between *immediate, intermediate* and *ultimate* criteria. Immediate and intermediate criteria are those which are normally readily available and measurable, such as production figures, sales or ratings; the ultimate criterion is the final goal of the organization and it is as a rule impossible to relate individual performances to goals of this nature.

A further distinction is made by Weitz (1961) who suggests that criteria may vary with respect to the *time* at which the criterion is measured, the level of *difficulty* of the criterion and the *type* of criterion such as ratings, errors etc. Finally, Astin (1964) makes a distinction between *conceptual criteria, criterion performance* and the *actual criterion measurement.* In Sweden, Ekvall (1971) has made a corresponding distinction between *verbal–conceptual criteria* and *operational criteria.*

It is difficult to determine the extent to which these various criterion

measurements (as proposed for instance by Tiffin and McCormick) are used in practice or for research purposes. Ekvall (1971) quotes a study of the two largest English speaking psychology journals during the five year period 1950–1955, which showed that 81 % of all criteria reported consisted of subjective ratings. In the same study Ekvall presents an exhaustive review of criterion research in Sweden, comprising 33 studies with some claim to scientific value. It is worth noting that only *one* of these involves a performance measurement (production), 8–10 use results from training programmes as one criterion, while the rest of the studies use subjective ratings by superiors as the criterion measurement. For reasons which will become apparent later, this is tantamount to stating that criterion research in Sweden is almost non-existent.

Should criterion data be global or differentiated?

For the sake of simplicity, it would obviously be most satisfactory if it were possible to obtain single, global measures or ratings to be used as the criterion of job success. This approach will appeal equally to the practitioner who in general will tend to think in these terms, as well as to the test methodologist whose validation measurements will thereby be simplified. In a celebrated review article, Dunnette (1963) states emphatically that there is no such thing as *the* criterion and that multiple dimensions of criteria must be accepted. This gives rise to the seeming paradox that if these multiple dimensions have high intercorrelations, then any one would have been sufficient, whereas if they do not it becomes considerably more difficult to determine the structure of job success. One is then faced with a number of statistical problems dealing with whether or not an attempt should be made to combine these independent measures into a weighted general criterion measure. As Dunnette (1963) states: "It is in this area that the dilemma becomes even more clear cut. If various measures correlate highly, an investigator gains some confidence that he is measuring the *'core'* of job success; yet, the necessity of their being weighted to form a single composite becomes less. Conversely, if the measures show low correlations a researcher may be gratified that he is tapping rather independent dimensions of job success or he may feel discontented because of the apparent lack of unity in this job success construct". This problem, originally examined by Thorndike (1949) has been investigated by a number of researchers such as Ghiselli (1956), Guion (1961) and Peres (1962). A certain amount of empirical evidence has been obtained. A study by Seashore, Indik and Georgopoulos (1960) provided a wide range of criterion measurements for delivery men. These involved performance measures, ratings, accidents, absences and errors and the correlation between the measures was for no single pair of criteria greater than 0.32.

Granted that many jobs involve multiple criteria, there are a number of schools of thought as to how this should be used in the validation of a selection procedure. One school has maintained that the dimensions should be kept separate and that the efficiency of the various forecasting instruments

should be assessed in relation to each criterion measurement. Another school maintains that attempts should be made to provide each of these independent dimensions with some weighting to produce a general composite criterion. A strong argument against the later approach is presented by Tiffin and McCormick (1963); they stress the fact that composite weighted criteria may give the same total score for two different applicants who in fact perform the job entirely differently and who have different performance structures on the original measurement instruments used in the selection procedure. There is also the complicated question of the extent to which *conjunctive* or *compensatory* models can be used in the actual criterion. We shall return to a number of these questions in the section on validation in this chapter.

A number of later studies has also thrown light on these problems: Fogli, Hulin and Blood (1971) established a number of behavioural job criteria useful for grocery retail staff as opposed to the application of a single, global dimension. Pallett and Hoyt (1967) collected 23 specific behaviour characteristics ratings and two overall ratings made by superiors for university graduates 5–10 years out of college. A multiple regression analysis showed that 13 of the ratings were sufficient to account for the entire set of variances and covariances. Eight of these 23 ratings made independent contributions to the prediction of overall ratings and could thus be regarded as independent criteria dimensions of success in general business. Turner (1960) performed a factor analysis of 11 objective criterion measures and 9 criterion ratings, including a global one. He obtained 4 factors, 3 of which were related to the objective measures of performance and one to the ratings. An interesting feature of this experiment was the attempt to determine factor relevance weights for each of the 4 factors with respect to two different plants. These relevance weights obtained from raters showed that only two of the factors had any weightings and that the two remaining factors should not in fact be included in any composite criterion.

Are criteria relating to individual workers static or dynamic?

In our initial attempts to specify the dimensions of criteria it was apparent that many of the measurement criteria made are short term, operational ones, whereas it is impossible to determine the long term ones, which are more connected with company goals in general. One may also remark that a further long term goal and criterion which is seldom considered in the field of industrial psychology is that of determining the extent to which job incumbents contribute to the financial situation of the company. This is a subject which Cronbach and Gleser (1963) revolutionized, after pioneering work by Brogden and Taylor (1950), when they tried to apply cost accounting concepts to criterion construction. Their approach was that money can be used to act as a common denominator for different criterion measures, at least in some cases. One illustration of this is that there may be two criterion measurements, production performance and number of errors respectively.

By careful study it may be possible to determine some form of simple relationship between the time taken to produce each unit and the time taken to rectify each error. It is obvious that this type of criterion measurement is only suitable for certain types of work.

The point of time at which the criterion should be measured has been studied in detail by Levine and Weitz (1971). Weitz had shown in a number of previous studies that there is a relationship between the difficulty of a task and the time at which the criterion should be measured. Weitz (1961, 1964, 1966) maintains that "If a task is easy, maximal differentiation on a dependent variable will appear relatively early in the task; if a task is hard, maximal differences will appear late in the task". The study by Levine and Weitz (1971) confirms this finding when it comes to group performance measures.

A number of other studies have also examined the dynamic nature of criteria, the contention being primarily that changes in skill or motivation will result in changing relationships between predictors and performance measures over time. Changes over time in the tasks involved in the job will also produce similar results. Studies of this type include Fleishman (1957),Fleishman and Fruchter (1960) as well as Fleishman and Hempel (1954). A study by Ghiselli and Haire (1960) showed that changes in performance (often of a positive nature) were still taking place ten years after initial employment for a group of investment salesmen. The criterion of the *rate* of change in performance was also shown to be a better measure during the first 18 weeks of employment for a group of taxi-cab drivers. Guion (1961) also showed that first-year sales and sales after five years of service produced different results and that predictors of early success were different from those of later success. We must therefore conclude, as suggested by the theoretical analysis of Ghiselli (1956), that not only are criteria multi-dimensional, but also subject to change over time.

A further feature of the dynamic aspects of criteria is presented by Thompson (1965) in which he takes up an idea from Super (1963), who states that the *occupational* model should be extended to a career model when determining how individuals succeed in various jobs. Looked at from the viewpoint of the individual, and perhaps also from the company, the important feature of a career is not the extent to which one is successful in specific positions, jobs or occupations but instead how successful one is in building up a career which may extend over a number of decades.

Super (1963) has studied these developments with respect to the trial and error attempts of young people to establish a *niche* for themselves in working life and a comparable study has been performed by Thorndike and Hagen (1959) when they studied the career patterns, both vertically and laterally, of 10.000 airforce officers.

What should the criterion measure and who should make the measurements? – The multitrait/multirater approach

In the previous section we have studied criterion measurement changes over time. The reason for these changes, as mentioned in that section, is – for one thing – that the individual job incumbents change in various ways. In fact, both the worker and the work may be expected to change over a period of time, and the complicated interaction between these two components has led some researchers to maintain that it is a fiction to attempt to distinguish between the worker and the work. The study by Levine and Weitz (1971) also stressed that features of the actual job situation may give rise to variability in the criterion measurements. Situational factors may in fact often be of considerable importance when obtaining criterion measurements.

Let us now return to the distinction made by Tiffin and McCormick between performance criteria, subjective criteria, physiological criteria and accident criteria. Performance criteria may of course vary in the short term or long term due to individual variations. We must however also expect the production technology or other situational factors to be important for performance criteria. In a job where the production rate and quality is entirely controlled by machines, it is of course impossible to establish any difference between the operators, other than by using indirect measures such as absenteeism, labour turnover or accident rates. If, however, a specific job is subject to considerable variations in the surroundings, such as irregularity in the supply of raw materials for production, then production measures taken at any specific time may be expected to have a very low degree of reliability. One study of this type reported by Tiffin and McCormick (1963) showed that the correlation of production records for odd and even weeks during a ten week period for the job of unit wirer gave a correlation of 0.87, which must be regarded as a very high degree of stability of work performance over a short period. This is however only one specific job and it is impossible to generalize this result. One possible factor influencing this stability is that the workers quite simply have arrived at a formal or informal agreement among themselves about how much they intend to produce.

In a similar manner, it is to be expected that the situational factors such as production technology will also give rise to considerable variations in physiological criteria and even accident criteria. It is obvious that the use of subjective criteria – such as ratings – by superiors introduces a further dimension of uncertainty into the picture. Here we may expect variation in situational variables, individual performance and individual ratings to produce subjective criteria measurements of a highly suspect nature. And yet it is worth remembering that it is this type of criterion measurement which dominates in the US and in Sweden, according to Ekvall (1971). This perseverance with the traditional rating method may be due to the fact that most raters quite simply do not realise how bad their ratings are. A further possible explanation is that many ratings experience a feeling of parsimony in the rating situation which makes them unwilling to embark upon complicated multivariable ratings.

The primary purpose of this section is, however, to examine a further source of uncertainty and variation when it comes to subjective criteria, namely the extent to which different rater categories have different cognitive structures with respect to the assessment of individual job performance. The traditional approach has been for one or more superiors to rate job incumbents. In recent years a large volume of research has compared these ratings with the ratings of subordinates, peers, and self-ratings. The differences between these rater categories cast even more doubt upon the value of superiors' ratings, as well as renewing the whole question of determining job performance. The method most commonly adopted when examining this problem is an adaptation of the Campbell and Fiske (1951) model for the *multitrait/multimethod analysis* of measurement presented in chapter 1. This has normally been transformed into a *multitrait/multirater* analysis where the ratings of different traits by different rater categories such as superiors, peers, subordinates and self-ratings have been compared.

One of the first studies of this type was performed by Hicks and Stone (1962). They compared ratings of overall performance, promotability and versatility performed by peers and superiors with test results in the selection procedure. The correlations between test battery and peer ratings were significant with respect to all three rating variables but were significantly correlated with the supervisors' evaluation only when it came to the variable *versatility*. Although the Campbell and Fiske terminology is not used by the authors, this study clearly shows the considerable differences in rating procedures between peers and superiors. A complicated study by Albrecht, Glaser and Marks (1964) also throws light upon this area. A wide range of subjective and objective rating methods were used to predict job performance for a number of industrial managers. Three types of performance were then selected as criterion data and a fourth variable, overall performance, was also included. Four different categories of raters then rated job success. Two groups of senior managers as well as peers *ranked* the job incumbents with respect to four criterion variables. The immediate superiors *rated* the job incumbents with respect to the same variables, and a consultant predicted performance on all four criterion variables on the basis of the prediction instruments. This design can be regarded as consisting of *five method or rating variables*, i.e. the four sets of criterion judgements and the set of predictor judgements performed by the consultant, and *four trait variables*, i.e. the three specific performance dimensions and the overall one. The product-moment intercorrelations between performance rankings and predictions were computed and the pattern of correlation coefficients examined. The first question asked was whether or not the trait ratings were independent. This was shown not to be the case. In other words there is a very considerable *halo-effect* which may or may not be based upon genuine clusters of good and bad traits. A second question was whether the ratings were valid. One method of determining this is to investigate whether measures of the *same* trait by *different* raters are more closely related than measures of *different* traits by *different* raters. As the authors point out, this comparison

is seldom made in criterion studies. The validity requirement was, however, met in this study, the median combined validity correlation for each trait being 0.38, while the median combined heterotrait correlation was 0.17. The main point of this study was to investigate the extent to which the consultants were able to predict how well the managers would perform their jobs. The correlations between these predictions with respect to the four traits produced 9 significant correlations out of 12 when the predictions were compared with criterion data obtained from one group of higher managers, the immediate superiors and peers. None of the predictions was significant when the other senior managers' criterion data were used.

Evans (1969) made a study of the convergent and discriminant validities of the Cornell job descriptive index and the goal attainment section of the need-satisfaction measure developed by Porter. It was shown that the extremely rigorous requirements presented by Campbell and Fiske (1959) were met in at least one of the samples. A methodological advance is the use of the *coefficient of concordance* in order to measure the fourth criterion as formulated by Campbell and Fiske.

The studies which have most carefully examined the issues of the multitrait/multimethod analysis of ratings and its significance for the criterion problem have been performed by Lawler (1967) and Kavanagh, MacKinney and Wolins (1971). Both are concerned with the measurement of managerial job performance. The only method available for obtaining criterion data is that of ratings, as opposed to the situation at the blue collar level, where it is often possible to obtain performance data. Lawler maintains that the scope of management criterion data can be greatly enriched by using a number of rater categories such as superiors, peers, subordinates and self-ratings. He considers that each of these categories has an adequate view of the manager's performance, although he considers that their views will tend to differ slightly. Some of the reasons for these differences are suggested by Mabon (1972). Lawler points out that subordinate ratings are important, since this determines the impact of the superior on the "human resources of the organization", to use Likert's terminology. This view is endorsed, *inter alios*, by Lennerlöf (1966), who has made similar studies using the ratings of subordinates as supervisory criteria. Lawler also makes the point that job performance is a function not only of ability but also of motivation. The fact that job performance ratings for a number of employees do not correlate highly with ability results from a test battery in an earlier selection procedure does not necessarily mean that the ratings are invalid; the discrepancy may for instance be due to motivational changes in the job incumbent. Lawler also takes up the distinctions made by Campbell and Fiske (1959), slightly reformulating the discussion for the situation in question.

One consequence of this approach is that if only one global criterion rating is obtained it is impossible to determine the validity of the criterion in any meaningful way. This is also the case where a single prediction rating is validated against a multidimensional criterion. Lawler also makes the important point that agreement between different rater categories does not

necessarily mean a valid rating; it may simply mean that both groups are rating the job incumbent in question in the same incorrect or biased fashion.

Lawler then reports a study concerning middle and top level managers in which superior, peer and self-ratings were obtained with respect to quality of job performance, ability to perform the job and effort put forth on the job. His study showed that the superior and average peer ratings had good convergent validity and that two of the three requirements for discriminant validity were met. What is perhaps most interesting in Lawler's study is that the *self-rating data* provide neither convergent nor discriminant validity, as they appear to bear little relationship to the ratings made by superiors and peers.

It is difficult to know how to interpret these results. The fact that they differ considerably from the two other rating categories may suggest that it is only the job incumbent himself who is really able to appreciate what has been achieved; a more likely interpretation is that defense mechanisms adopted by everyone in- and outside working life may tend to distort the self-insight of the job incumbents themselves. Lawler has performed some interesting comparisons with job satisfaction measures and considers that discrepancies between self-ratings and the ratings of superiors will in some cases give rise to lower job satisfaction, in that the employees may feel that they are not appreciated. One application of the multirater method which springs to mind in this situation is obvious, i. e. using the discrepancies with respect to the ratings between different categories in order to discuss how well the organization is functioning. It is in essence techniques of this type which are applied within the fields of organizational development and process consultation (see for instance French and Bell, 1973).

The methodological and data processing aspects of the multitrait/multirater method are examined in detail by Kavanagh, MacKinney and Wolins (1971). They show that data of this type can be studied by means of analysis of variance and that the rating is a function of the effect of the *manager*, the *trait* and the *source*, i. e. the rater category. The three types of variation can then be analyzed to determine *halo-effects, discriminant validity* and *convergent validity*. They then present a study where three different hierarchically ordered managers are rated with respect to twenty variables, consisting of three major sections: functions of the job, subjects of the job and personal traits. The middle manager was the basic unit of analysis and each unit consisted of his superior and his best and worst subordinate. The results showed good convergent validity, but there was a large halo-effect. Discriminant validity was rather weak, which suggests a reduction of the number of rating dimensions. A partitioning of the data into 6 trait dimensions and 14 performance dimensions showed that there is more halo for personality traits than for performance traits. This did not however lead to more differentiation of managers but rather more bias within raters and more agreement between raters.

One of the most interesting results of this study is the extent to which personality traits have been shown to have some value in this situation. They

provide an increase in convergent validity for the total matrix, the drawback being the increase in halo and the limited increase in discriminant validity. The study suggests that personality trait ratings performed by managers are not entirely without value, as has been suggested or implied in many other studies.

The final area of interest in the present section, is the type of data which should be used in the criterion variables. The previous studies of job analysis have shown that a distinction can be made between describing a job in terms of tasks, of job behaviour and of trait requirements. The vast majority of studies reported here and elsewhere have tended to phrase criterion measures in terms of *behaviour.* The study reported above by Kavanagh *et al.* has shown that trait assessments also have some value. A useful comparison of the three approaches to criterion measurement is provided by Bell, Hoff and Hoyt (1963). Selection data based on the general aptitude test battery were available for 1,711 workers in 16 occupations. The criterion rating A was job oriented and was obtained by means of a "critical incident"-approach to interviewing immediate supervisors of each employee with respect to essential job elements. When these elements had been listed and defined, the supervisor was asked to rate each of his employees in the sample on each element. The criterion rating B was also completed by the worker's immediate supervisor and consisted of 25 pairs of positive and negative behavioural descriptions. The criterion rating C was a rating of each worker on 10 traits made by second-line supervisors. Correlation coefficients were in general low, seldom rising above 0.30. Nine per cent of the correlations were *negative* for ratings A and C, and as many as 24 % were *negative* for the criterion rating B, i. e. the behavioural descriptions. Rating C produced the highest correlation with the aptitude in question in 62 instances, i. e. 62 different jobs, rating A in 52 instances and rating B in 27 instances. The authors consider that the *trait-oriented* rating is the best one, especially when the jobs involved are blue-collar ones. Rating A, i. e. the job element or task approach to criterion measurement, is appropriate for white-collar jobs where the predictors have already been identified. These results are thought-provoking, as most criterion measurements, at least when it comes to ratings, tend to be type B, i. e. behavioural ratings. Finally, it is worth noting that the various correlation measures do not correlate particularly highly with each other, which also suggests that the criteria are measuring different aspects of job performance.

A final study with far-reaching implications is that performed by Campion (1972) of the tasks and behaviour involved for maintenance mechanics. Thirty-four maintenance mechanics were assessed on the basis of work sample measures, as well as of paper and pencil test results and as criterion data three supervisory evaluations of job success were used. The work sample predictions proved to correlate highly with all three criterion variables, whereas traditional paper and pencil testing normally considered of value as predictive instruments in cases such as this, were found to have correlations close to zero with the criterion variable. Care must be taken in interpreting a concurrent validation study of this type, but there is much to suggest that

attempting to forecast working behaviour on the basis of a behavioural consistency model as proposed by Wernimont and Campbell (1968), and to some extent confirmed in the study by Campion, is a better measure than that of establishing tests for trait measurements thought to be important in the performance of the job. Contrary to the previous study by Bell *et al.*, (1963), this suggests that the behavioural approach has considerable value after all.

Validation of a selection procedure

The aim of the present study is not to present a detailed analysis of testing theory and its application to the selection situation. There are however a number of technical aspects of validation which have been touched upon in chapter 1 which are of relevance for our continued discussion of the applications of job analysis to the selection situation. They will be elaborated upon briefly below.

Predictive validity

In its simplest form predictive validity is simply the correlation between a global prediction score and a global criterion rating. A number of practical difficulties arise, in the application of this measurement. The relationship between predictive validity measurement and the utility of the selection procedure is a subject which has been debated for decades and which is analyzed in detail in Cronbach and Gleser (1963). Some researchers have considered that the *coefficient of determination* (r^2), i. e. the square of the validity coefficient, is in fact a better measure of the value of the selection procedure. Others have favoured an index of forecasting efficiency (E) defined by

$$E=1-\sqrt{1-r^2}$$

A further problem which arises is the *restriction of range* on the criterion variable, due to the fact that not all of the applicants are assigned criterion scores. This will in fact mean that the validity coefficients obtained between forecasting variables and criterion variables will be *lower* than that which would have been obtained if all of the applicants had been assigned scores. The work of Cronbach and Gleser (1963) has also served to underline the doubtful value of the validity coefficient *per se*. As they and others have shown it is not the validity coefficient alone which determines the value of a selection procedure but also the *selection ratio* and the *base ratio*, i. e. the proportion of applicants selected and the proportion of applicants who would be suitable. The tables developed by Taylor and Russell (1939) show that selection procedures with low predictive validity can still be of considerable value for certain combinations of selection ratio and base ratio. Some examples of this are given in Mabon (1972), where attempts are also made to relate selection decisions to financial criteria.

So far, we have only considered the simplest case, that of a *single* predictor and a *single* criterion. In the case of multiple predictors and a single criterion the validity of each individual predictor can be computed with respect to the criterion. It is then a simple matter to determine with the aid of multiple regression analysis the weights which should be assigned to each predictor in future applications of the selection procedure. The measurement problem is a similar one if the criterion measurements are multidimensional and combined to form a composite criterion measurement. In the present context this is comparable to a single global measurement. The arguments against this procedure have been presented in a previous section. If both predictors and criteria are multidimensional, it will be possible to compute a validity matrix for the predictors and criteria. The development of multiple regression analysis used to study the structure of data of this type is canonical analysis (cf. Mårdberg, 1969, Andersson, 1972).

One problem arising in this connection is the classic controversy concerning *clinical vs. statistical prediction.* Those in favour of statistical prediction consider that the weightings assigned to different predictors and the extent to which various predictors are applied should be based entirely on the statistical analysis of the correlations between different predictors and criteria. A study of the pattern of validities in the matrix mentioned above will demonstrate the value of various predictors and provide information as to how they should be combined to maximize predictive validity. On the other hand, the clinical approach is supported by psychologists who maintain that the psychologist or rater himself should be given the freedom of judgement to combine, exclude or add to the various predictive measures, on the basis of his or her own intuitive feeling of how the data best can be utilized in each individual case. A classis review of this evidence is provided by Meehl (1954). The basic thesis in that book is summed up in the preface "There is no convincing reason to assume that explicitly formalized mathematical rules and clinician's creativity are equally suited for any given kind of task or that their comparative effectiveness is the same for different tasks. Current clinical practice should be much more critically examined with this in mind than it has been" (Meehl, 1954, page 6). Holt (1958) claimed that the clinical approach has something to add to actuarial techniques with large quantities of data and this plea is echoed by Trankell (1959). In this study he demonstrates the value of the clinical approach as an instrument for predicting success in the selection of pilots.

The problem of predictive validity becomes even more complicated when we leave selection for a single job and pass on to the more complicated situations of *classification* and *placement.* A topological approach to this subject is provided by Rulon, Tiedeman, Tatsuoka and Langmuir (1967). Much work on multiple absolute predictions has also been performed by Horst (1962), but this is not within the scope of the present study.

Absolute profile prediction

We have on a number of occasions had occasion to refer to the work of Mårdberg, especially with respect to his attempts to apply latent profile analysis to the selection situation. Much of this work was originally derived from Mårdberg's dissatisfaction with traditional validation methods. In an introductory work (Mårdberg, 1966, page 74 *et seq.*) he criticizes some of the simpler validation models described above and he maintains that neither Cronbach and Gleser (1965) nor Horst (1962) have been entirely successful in solving the many problems arising in the study of multi-dimensionality, of the profile approach and of decision theoretical techniques.

Mårdberg presents a model in which he assumes a requirement space of *n*-dimensions; any relevant job can be assigned to a point in this space. Latent profile analysis can be used to classify jobs into job classes, these classes having a given level on each requirement dimension. The problem of selection or classification then becomes a simple one, according to Mårdberg. All that is necessary is to measure applicants with respect to the various requirements; those who are closest to each specific job class profile are the most suitable for the jobs in question. We thus have a test space whose dimensions are identical to those in the requirement space. This technique of assigning applicants to job class profiles is called *absolute profile prediction* and the distance of an applicant to the class in question is measured using the generalized distance formula, Mahalanobis'd^2 (1936). In a selection procedure the applicants are chosen who are closest to the profile and in classification each individual is assigned to the class profile to which he is closest. Advantages of this method are that identical elements are used in the selection and requirement dimensions. In addition, the concepts of validity and reliability can be combined into one of *generalizeability*. The validity of the test thus consists of determining the extent to which predicted outcomes have been confirmed, i. e. whether the performances on the job correspond to the predicted profile in the requirement space. According to Mårdberg the criterion measurements should be based on *behavioural* categories. To be sure, we do not know whether other areas of the requirement space would have been equally successful in the jobs in question, but Mårdberg considers this problem insoluble at present. The greatest difficulties arising in the model are the determination of job elements and requirement information and this is considered by Mårdberg to be a central area for job analysis work. A second problem is the construction of equivalent measurement instruments so that the test instruments directly correspond to the requirement space. This means that a considerable degree of the interest at present devoted to predictive validity will be transferred to other forms such as synthetic and construct validity. This is all for the best, in view of the fact that restriction of range tends to make nonsense of many predictive validity measurements. Mårdberg considers that his model is a simplification which is in itself a weakness, especially in view of the fact that predictions in the individual case will have a considerable degree of uncertainty. Problems of this type arise in many areas of psychological measurement and need not necessarily disqualify this

approach in this specific situation. Furthermore, one must expect a certain degree of overlapping between different classes of jobs. This implies that absolute profile prediction will tend to be based on probability rather than certainty. As this is also the case for latent profile analysis, the concept of probability in individual assignment has been made an integral part of the model. In a recent study (Mårdberg, 1973), he returns to the theoretical problems involved in validation. Here he stresses the complications arising from the multidimensionality of the criterion variables and from the fact that the relationships between predictors and criteria may be non-linear. One method of getting to grips with these problems is to use latent profile analysis.

The validation model presented by Mårdberg in 1966 gives rise to a number of further complications over and above those which he himself has indicated. First of all it is not clear how the requirement variables are to be obtained. In some places he appears to assume that the job descriptions will involve specific job elements and that there will be a one-to-one relationship between job elements and job requirements. This is of course not the case, as an extremely interesting study by McCormick, Cunningham and Thornton (1967) shows (see p. 107). Mårdberg does not distinguish between the Swedish equivalents of *jobs* and *positions*, as defined in chapter 2. In addition, a number of the problems presented in the section in chapter 4 on selection models once again present themselves. Here we are assuming a *conjunctive model*, but until some form of validation has been performed it is impossible to determine where the cut-off lines should be drawn. Furthermore, we are assuming that each trait is of equal importance, which is almost certainly not the case. This fact will make it extremely difficult to assign individuals to specific job class profiles when they do not closely fit any specific pattern. In conclusion, Mårdberg considers that the procedure can be validated by means of behavioural criterion data. It is, however, not at all clear in what way these behavioural data will demonstrate the extent to which the individual job incumbent meets job requirements. The studies previously reported by Bell *et al.* have shown that behavioural criterion data are often suspect and that trait or job element data are at least in that specific situation better. It is in fact not clear why Mårdberg does not take the step of recommending *trait* criterion data directly corresponding to the requirements and trait spaces. This would presumably be contrary to his basic rejection of the trait approach.

A further study which throws light upon these issues was completed in 1971 (Mårdberg and Baneryd). This study has been reviewed previously, but from a somewhat different angle. In it the authors attempt to classify jobs in accordance with the previous profile model. To do so they use a number of different types of data for grouping jobs. The first method, which is a job description, uses more or less objective data dealing with environmental conditions, training requirements and job characteristics such as responsibility and length of the working cycle. LPA, based on a number of variables of this type, is then used to group 38 jobs within the foundry branch. The solution

selected has 13 profiles. The authors then select, somewhat arbitrarily, 8 psychological job requirements which can be used to classify the 38 jobs. This LPA gives rise to a number of solutions one of which containing 7 profiles was selected. Finally LPA was used to classify the 38 jobs with respect to a number of medical variables. This produced 9 profiles. These studies are all of considerable interest showing how LPA can be applied to data of this type and in providing a basis for discussions as to the practical value of this type of data processing. What they fail to do is to fit into a coherent validation model in accordance with that presented previously by Mårdberg (1966) himself. No attempt has been made in the study to combine on the one hand the job description data and on the other hand the various psychological and medical data, i. e. the job requirement data. What is necessary to develop the previous model is to establish some sort of relationship between them, especially from the point of view of showing how the job description data can be translated into job requirement data, which can then be measured with the aid of standard psychological and medical methods. The job descriptions provided by Mårdberg and Baneryd (1971) are very extensive and carefully thought out. If it is established for instance that a certain job involves working in an environment where the temperature is 30° and the noise level is 110 decibels should the industrial psychologist then attempt to translate these environmental requirements to individual requirements or should he state that a working situation of that type should not be permitted at all?

It appears as if Mårdberg has not yet solved some of the basic problems in his *absolute profile prediction model*, while others such as the criterion problem have not yet been touched.

Synthetic validity

The crucial problem of attempting to relate job description data to job requirement data has, as far as the author is aware, been examined only by McCormick and his colleagues. An important article (McCormick, Cunningham and Thornton, 1967), which indirectly also shows an awareness of the fact that combinations of syndromes or attributes may be more valuable than unitary attributes, deals with the problems of job characteristics and corresponding human attributes. The study aims at developing the idea of *synthetic validity*, as propounded by Lawshe (1952). With some slight adjustment of the terminology in the article the authors maintain that the basic elements necessary in order to test synthetic validity are: an extensive job description covering essential elements in the job, the corresponding attributes (i. e. traits) which each such job element requires and appropriate measures of these attributes. Thereafter some synthetic technique can be applied in order to combine these attribute measures in a manner which reflects the characteristics of the jobs in question. The authors realise that this may present certain difficulties in determining what is meant by the *true* attributes of a job and they continue: "Rather, it might be necessary to build a bridge between each job characteristic and the 'measure' or predictor that is

valid for predicting satisfactory performance in the aspect of the job that corresponds to each job characteristic" (*op. cit.* p. 433). It is this problem which Mårdberg's model has not yet solved.

The basic material used by the authors to study these questions consisted of the 162 elements of the Worker Activity Profile (WAP) developed by them previously and 42 human attributes primarily obtained from the material used in the Dictionary of Occupational Titles. These attributes had been identified with the aid of factor analysis. As a first step, each job element, i, was rated with respect to each of the 42 attributes, j, i. e. an 8-point scale was used to establish the extent to which each job element required the attribute in question. This produced a 162 x 42 matrix of attribute scores, x_{ij}. Each individual job as described in the USES material was then analysed with respect to the 162 elements and given an importance rating on each element, y_i. These scores for each element could then be used to multiply the attribute scores previously established for each job element and, summing the cross products for each attribute, 42 attribute scores could then be obtained for each job. Expressed in matrix terms, this involves calculating $y_i x_{ij}$. In other words the job description has been translated into a job requirement trait analysis. In a first sample, 401 job descriptions were obtained from the USES. This material, as was stated in chapter 2, also consists of quantified trait requirements, although only for 37 of the 42 traits in question here. These synthetic attribute scores, i. e. $y_i x_{ij}$, could then be compared with those available *a priori* from the original USES data. This may be regarded as an example of both *concurrent* and *synthetic validity*, the USES data being regarded as criterion data. The two different attribute scores, i. e. the synthetic ones and the USES data were correlated for all 401 jobs: 32 of the 37 variables showed correlations significant at the 0.01 level. Aptitudes showed the highest correlations, while temperament and interest attributes tended to be lower, a not unexpected result.

A second experiment within the framework of the same general study involved 43 naval jobs for which synthetic attribute scores had been obtained using the same method as before. In addition, 8 subtests had been applied to these jobs for selection purposes and a total of 125 validation studies had previously been performed. Each test was then equated to one of the attributes. Each naval job thus had an attribute score corresponding to the test in question and for each test there were 125 validity scores. These attribute scores were then correlated with the validity scores for each of the tests. Similar comparisons were also made when the individual job elements in the WAP were reduced by means of factor analysis to job dimensions. The vast majority of the correlation coefficients so obtained were significant. The correlations at the job dimension level were somewhat higher than at the WAP level, which caused the authors to conclude that there may be some possibility of applying more parsimonious techniques in future practical applications.

This study has been presented at length as in the view of the present author it takes up a number of points of basic importance. The synthetic

validation is, to be sure, subject to previous predictive validations with all the weaknesses which they contain: nevertheless, the fact that the correlations were almost all significant suggests that the synthetic approach is of great value and worthy of future research.

Concluding discussion

Let us now review some of the themes from the discussion in chapters 4 & 5, and relate them to the methodological problems dealt with in this and previous chapters. Traditional research within this field seems to be in agreement that the selection model should consist of job description → job requirement analysis → applicants measured on job requirements → applicants assigned to positions or rejected → criterion data collected and used to validate the study. The models presented in chapter 3 have consisted of those which have bypassed the job requirement analysis (models 1 and 2), while others have not based the job requirement analysis on a previous job description (models 3, 4, 5, 7 and 8). Apart from the study just presented by McCormick *et al.*, no researcher has attempted to bridge the gap between description and requirement analysis.

The McCormick model also presents a number of difficulties.

Assuming that we have tests or other predictive instruments to measure attributes, we will be able to compare applicants and requirements in the traditional manner. The McCormick model will not, however, provide much information as to whether or not a compensatory or conjunctive model should be applied or in what way the attributes should be weighted, if at all. This is presumably an area which could be studied with the aid of traditional validation techniques, although the purpose of this approach is, in the long run, to do away with these techniques. Mårdberg's work at the present stage consists of studies of job descriptions on the one hand and job analysis data on the other, with little connection between the description and the analysis.

There are also differences of opinion among researchers when it comes to the processing of the data, whether they be descriptive or analytic. The method which has dominated to date has been factor analysis, but non-linear methods such as LPA have gained ground in recent years.

Turning to the problem of criterion data, a number of alternatives seem open. The traditional method of comparing a global, composite or multi-dimensional criterion with predictive instruments has fallen into disrepute for reasons presented above. Some further difficulties in the predictive model should also be apparent from the above reasoning. Apart from standard difficulties such as restriction of range, one may ask what in fact is the significance of a low correlation coefficient? It may, for instance, mean that a poor and misleading job description has been used in the selection procedure. If the description is more or less satisfactory, then it is possible that the (intuitive) translation of the description into job requirement terms is wrong or had deficiencies. A third possible source of error is that the measurement instruments used for prediction purposes are unsatisfactory and do not

directly correspond to the job requirements. In addition there are all the problems of the reliability of the actual tests or other selection procedures which are normally dealt with in books on selection theory. It is the contention of the author that the traditional validity coefficient is more or less worthless, as it throws little or no light upon the processes involved in the determination of job requirements and the assignment of people to the jobs in question.

If this model is not regarded as satisfactory, what other type of validation procedure should be applied? Mårdberg has suggested *absolutee profile prediction* using LPA, based on the conviction that it is patterns of attributes which are important, not individual ones. His model from 1966 which is based on the idea of synthetic validity would be: selection based on the correspondence between patterns of job requirements and individual trait patterns. The extent to which the selection has been successful is then determined by studying the extent to which the selected applicants "behave" in a way to be expected by their trait patterns. As has been mentioned previously, it would appear to be more correct to collect criterion data along the same dimensions as the requirement patterns, i. e. to use synthetic criteria. Conversely, it can of course be maintained that it is only by studying the relationship between actual behaviour and presumed important patterns of traits that some idea of the value of these trait patterns for the job in question can be obtained.

A second attempt to solve this problem is that of McCormick. Here it would be possible to "validate" the selection in several ways. One type of validation would be to establish the extent to which the original job description data is still relevant. This could be done by criterion raters going through the WAP or a dimensional abbreviation, and the results obtained then being compared with the original job description. Differences here could of course explain low validity, using traditional predictive methods. A second form of validation would be for the raters to reappraise the attributes required for the job. This would give a measure of the extent to which the previously mentioned "transformation matrix", i. e. $y_i x_{ij}$, is in fact correct for the job in question. A third type of validation would be to rate the job incumbents' actual performance on the job. Low validity here, assuming that the previous two types of validity are high, would be a sign that the faults, apart from deficiencies in the collection of criterion data, are to be found in the actual instruments used to measure the attributes in question. Thus, in the opinion of the author, the build-up of a pattern of synthetic validity which step by step covers different aspects of the selection procedure is greatly to be preferred to the predictive model.

A completely different model for the "validation" of a selection procedure is suggested by the work of Campion (1972). This researcher has cut the Gordian knot of translating back and forward from descriptive data to attribute data by simply ignoring the hypothetical constructs. His recipe is to perform a detailed job description which results in a number of job elements. Instead of translating these elements into psychological attributes, the

applicants are instead rated directly on their ability to perform the job elements in question. Once again there are some problems involving the possible weighting of different elements, but the relative importance of these components can be more easily established in cases such as this rather than when assessing the relative importance of different patterns of psychological traits. The "validation" procedure in this case is quite simple. It can be checked whether or not the job elements have changed and if this is not the case any deviation from the performance of the applicants on the selection job sample must be due to errors in the ratings or changes on the part of the job incumbents. Assuming that the work sample was carefully thought out, that the work is unchanged and that the criterion scores are more or less reliable, the whole problem of the validation of the procedure is then reduced to the question of the extent to which individuals vary in their job performance. This occurs of course in all selection procedures, but is considerably harder to establish when more complicated psychological models are applied. Campion's model is an attractive one, which is likely to be applied in the future. Its weakness is of course that it can normally only be applied to jobs where it is relatively easy to break down the job into a number of clearly distinguishable elements, each of which can be investigated in the work sample selection situation. As this type of job is gradually decreasing in importance, due to new job designs and automation, the scope of the model may be reduced in the long run.

In concentrating upon the relationship between the criterion and previous stages in the selection procedure, we have consciously ignored the problems presented by the criterion itself, not least the difficulty of using rating data. It has been shown in this chapter that the impression of how well someone has succeeded in his or her job differs widely when the opinions of groups such as superiors, peers, self-ratings and subordinates (at least in theory) are compared. It is obvious that the cognitive aspects of the situation must be considered and that each role within the organisation will tend to create a set producing different evaluations from different categories (and even to some extent within the categories). The final question in this chapter is how this variation should be taken into account in the models presented and discussed in chapter 3.

The fact that different groups of people in the job environment have widely differing ideas as to how well a job incumbent is performing, is of course a serious blow to all validation discussions. Some of the possible consequences should be readily apparent. Assume that a superior has participated throughout in a selection procedure. This could mean that he has described the job, determined the job requirements and then rates the job incumbents on the extent to which they have fullfilled these requirements. (This hypothetical situation which introduces *criterion contamination* into the picture widens the issue still further.) The point which the present author is trying to make is that a high or low predictive validity in this case will reveal very little other than the extent to which the behaviour of the job incumbent in question corresponds to the cognitive framework of the

superior in question. Many of the models in chapter 3 are suitable for studying these cognitive aspects of job descriptions and job requirements. It is, in the first instance, models 2 and 4 which permit the study of these different cognitive frameworks. The multitrait/mulitrater method provides similar information with respect to criterion data. This area is an interesting field for research which, as far as the author is aware, has not been undertaken to date. Models 2 or 4 can be used to distinguish between the impressions of a specific job for different rater categories, either with respect to the job description or the job requirements or both. It would be interesting to relate these differences to possible differences with respect to the rating of different job incumbents. It is possible that each category of raters is rating consistently within the framework of a basic approach to the job and its specific requirements. This possible agreement would, however, tend to become contaminated for personal reasons, especially in the case of self-ratings. The study of different cognitive views of jobs and job incumbents is however of considerable importance.

This concludes the present study of "the criterion problem" and of the extent to which various types of job analysis techniques can contribute towards increasing our understanding of the processes underlying the assignment of "the right man to the right job". In the opinion of the author the most fruitful approaches are those of the transformation matrix devised by McCormick and, within certain areas, the work sample technique of Campion. Mårdberg's work also promises to provide new insights into job analysis and the criterion problem, although much practical and theoretical work still remains to be done within the field of LPA.

References

(English translation of foreign titles in brackets)

Ableson, R. P. (1954) A technique and a model for multidimensional attitude scaling. Amer. Psychologist. 9, 319 (abs.).

Abruzzi, A. (1956) Work, workers and work measurement. New York: Columbia University Press.

Albrecht, P. A., Glaser, E. M. & Marks, J. (1964) Validation of a multiple-assessment procedure for managerial personnel J. Appl. Psychol., 48, 351–360.

Almroth, H. & Jonsson, B. (1971) Några systemteoretiska organisationsmodellers empiriska värde (The empirical value of some organisation models based on system theory). Mimeo. Department of Business Economics. Stockholm University.

American Institute for Research (1951) The development of job analysis procedures. Amer. Inst. Res. Note No. 4.

Andersson, G. (1972) Kanonisk analys och multipel regression vid personalurval genom psykologisk lämplighetsprövning (Canonical analysis and multiple regression in personnel selection using psychological suitability testing). Report No. 4, Department of Psychology, Gothenburg University.

Andersson, B.-E. & Nilsson, S.-G. (1966) Arbets- och utbildningsanalyser med hjälp av critical incident metoden (Job and training analysis using the critical incident method). Uddevalla: Akademiförlaget.

Andrews, T. G. & Ray, W. S. (1957) Multidimensional psychophysics: A method for perceptual analysis. J. Psychol., 44, 133–144.

Astin, A. W. (1964) Criterion-centered research. Educ. & Psychol. Measurement, 24, 807–822.

Attneave F. (1949) A method of graded dichotomies for the scaling of judgement. Psychol. Rev., 56, 334–340.

Barnard, C. I. (1938) The functions of the executive. Cambridge, Mass.: Harvard University Press.

Bell, F.O., Hoff, A.L. & Hoyt, K.B. (1963) A comparison of three approaches to criterion measurement. J. Appl. Psychol., 47, 416–418.

Benjamin, A. C. (1955) Operationism. Springfield, Ill.: Charles C. Thomas.

Benson, J. & Hellberg, U. (1970) Arbetsbeskrivning sedd i systemteoretisk belysning (Job descriptions and the systems approach). Mimeo. Department of Business Economics, Stockholm University.

Björkman, M. & Ekman, G. (1957) Experimentpsykologiska metoder (Methods in experimental psychology). Stockholm: Almqvist & Wiksell.

Boalt, G. (1965) Profetgloria och forskarmöda (Prophetic haloes and research exertions). Stockholm: Almqvist & Wiksell.

Boling, J. & Fine, S. A. (1959) Cues used by raters in the rating of temperament of jobs. J. Appl. Psychol., 43, 102–108.

Bridgman, P.W. (1927) The logic of modern physics. New York: Macmillan.

Brogden, H. E. & Taylor, E. K. (1950) The dollar criterion – applying the cost accounting concept to criterion construction. Pers. Psych., 3, 133–154.

Brown, K. R. (1967) Job analysis by multidimensional scaling. J. Appl. Psychol., 51, 469–475.

Brumback, G. B. & Vincent, J. W. (1970) Factor analysis of work-performed data for a sample of administrative professional and scientific positions. Pers. Psych., 3, 101–107.

Buckley, W. (1967) Sociology and modern system theory. New Jersey: Prentice-Hall.

Cambell, D.T. & Fiske, D. W. (1959) Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol. Bull., 56, 81–105.

Campion, J.E. (1972) Work sampling for personnel selection. J. Appl. Psychol., 56, 40–44.

Carnap, R. (1950) Logical foundations of probability. Chicago: University of Chicago Press.

Carroll, S. J. & Taylor, W. H. (1969) Validity of estimates by clerical personnel of job time proportions. J. Appl. Psychol., 53, 164–166.

Chalupsky, A. B. (1962) Comparative factor analysis of clerical jobs. J. Appl. Psychol., 46, 52–55.

Christensen, J. M. (1950) A sampling technique for use in activity analysis. Pers. Psych., 3, 361–367.

Coombs, C. H. (1950) Psychological scaling without a unit of measurement. Psychol. Rev., 57, 145–158.

Coombs, C. H. (1952) A theory of psychological scaling. Eng. Res. Bull. No. 34. Ann Arbor: University of Michigan Press.

Coombs, C. H. & Salter, G. A. (1949) A Factorial Approach to job families. Psychometrica, 14, 33–42.

Cronbach, L.J. & Meehl, P.E. (1955) Construct validity in psychological tests. Psychological Bulletin, 52, 281–302.

Cronbach, L. J. & Gleser, G. C. (1965) Psychological tests and personnel decisions. Urbana, Ill. University of Illinois Press.

Dahlbäck, B. (1966) Arbetsvärdering med poängsystem (Job evaluation using the points method). Mimeographed Report No. 43. Stockholm: Swedish Council for Pers. Adm.

Das, R. S. (1960) Comparison of worker analysis ratings based on job descriptions and motion-time study. Occup. Psych., 34, 141–147.

Dictionary of Occupational Titles (1949) United States Employment Service, Washington DC. US Government Printing Office.

Dodd, S. C. (1943) Operational definitions operationally defined. Amer, J. Sociol., 48, 482–489.

Dunnette, M. D. (1963) A note on *the* criterion. J. Appl. Psychol., 47, 251–254.

Ekvall, G. (1971) Personurval med hjälp av psykologiska undersökningar (Personnel selection using psychological testing). Swedish Government White Paper No. 1971:47, Stockholm.

Evans, M. G. (1969) Convergent and discriminant validities between the Cornell Job Descriptive Index and a measure of goal attainment. J. Appl. Psychol., 53, 102–106.

Fine, S. A. (1957) USES occupational classification and Minnesota Occupational Rating Scales (MORS). J. Counsel. Psychol., 57, 4, 218–223.

Fine, S. A. & Heinz, C. A. (1957) The estimates of worker trait requirements for 4 000 jobs. Pers. Guid. J., 36, 168–174.

Fleishman, E. A. (1957) A comparative study of aptitude patterns in unskilled and skilled psychomotor performance. J. Appl. Psychol., 41, 263–272.

Fleishman, E. A. (1967) Individual differences and motor learning. In Gagné R.M. (Ed.): Learning and individual differences. Columbus, Ohio: Merrill.

Fleishman, E. A. & Fruchter, B. (1954) Factor structure and predictability of successive stages of learning Morse code. J. Appl. Psychol.. 44, 96–101.

Fleishman, E. A. & Hempel, W. E. (1960) Changes in factor structure of a complex psychomotor test as a function of practice. Psychometrika. 18, 239–252.

Flanagan, J. C. (1954) The critical incident technique. Psychol. Bull., 51, 327–358.

Fogli, L., Hulin, C. L. & Blood, M. R. (1971) Development of first-level behavioral job criteria. J. Appl. Psychol., 55, 3–8.

French, W. L. & Bell, C. H. (1973) Organization development. Englewood Cliffs, N. J.: Prentice Hall.

Gagné, R. M. (ed) (1962) Psychological principles in system development. New York: Holt, Rinehart and Winston.

Ghiselli, E. E. (1956) The placement of workers: concepts and problems. Pers. Psych., 9, 1–16.

Ghiselli, E. E. & Brown, C. W. (1948) Personnel and industrial psychology. New York: McGraw-Hill.

Ghiselli, E. E. & Haire, M. (1960) The validation of selection tests in the light of the dynamic character of criteria. Pers. Psych., 13, 225–231.

Gibson, W. A. (1959) Three multivariate models: Factor analysis, latent structure analysis and latent profile analysis. Psychometrika, 24, 229–252.

Gibson, W. A. (1962) Class assignment in the profile model. J. Appl. Psychol., 46, 399–400.

Gonyea, G. G. 1961) Dimensions of job perceptions. J. Counsel Psychol., 8, 305–312.

Gonyea, G. G. & Lunneborg, C. E. (1963) A factor analytic study of perceived occupational similarity. J. Appl. Psychol., 47, 166–172.

Gordon, G. C. & McCormick, E. J. (1963) The identification, measurement and factor analyses of "worker-oriented" job variables. Occupational Research Center, Purdue University.

Gragg, D. B. (1962) An occupational survey of an airman career ladder: Supply warehousing inspection. USAF PRL tech. docum. Rep., No. 62–69.

Green, B. F. (1952) Latent structure analysis and its relation to factor analysis. J. Amer. Statist. Ass., 43, 71–76.

Guilford, J. P. (1954) Psychometric Methods (Second Edition). New York: McGraw-Hill.

Guion, R. M. (1961) Criterion measurement and personnel judgements. Pers. Psych., 14, 141–149.

Guttman, L.: (1941) The quantification of a class of attributes: A theory and method of scale construction. In Horst, P. *et al.*: The prediction of personal adjustment. New York: Social Science Research Council.

Guttman, L. (1944) A basis for scaling qualitative data. Amer. Sociol. Rev., 9, 139–150.

Guttman, L. (1950) Chapters 2, 3, 6, 8 and 9 in Stouffer, S. A. *et al.* (Eds.) Measurement and prediction. Princeton, N.J.: Princeton University Press.

Harman, H. H. (1967) Modern factor analysis (Revised edition). Chicago, University of Chicago Press.

Harman, H. H. (1968) "Factor analysis" in Whitla, D.K. (Ed.): Handbook of measurement and assessment in behavioral sciences. Reading, Massachusetts: Addison-Wesley.

Herzberg, F.: (1959) The motivation to work. New York: Wiley.

Hicks, J. A. & Stone, J. B. (1962) The identification of traits related to managerial success. J. Appl. Psychol., 46, 428–432.

Hinrichs, J. R. (1970) Psyhology of men at work. Ann. Rev. Psychol., 519–554.

Holt, R. R. (1958) Clinical and statistical prediction: A reformulation and some new data. J. Abnorm. Soc. Psychol., 56, 1–12.

Horst, P. (1962) The logic of personnel selection and classification. In R. M. Gagné (Ed.): Psychological principles in system development, pp. 231–271. New York: Holt, Rinehart & Winston.

Jackson, J. M. (1949) A simple and more rigorous technique for scale analysis: A manual of scale analysis. Part II (Mimeographed). Montreal: McGill University.

Jaspen, N. (1949) A factor study of worker characteristics, strength intelligence, inspection. J. Appl. Psychol., 33, 449–459.

Jonasson, S. (1970) Svensk sysselsättningsnomenklatur (Swedish occupational terminology). Internal Mimeographed Report, National Central Bureau of Statistics, Stockholm.

Jones, E. I. *et al.* (1953) A survey of the literature on job analysis of technical positions. Pers. Psych., 6, 173–194.

Jones, J. J. & de Cotiis, T. A. (1969) Job analysis: national survey findings. Personnel Journal, 48, 10, 805–809.

Kavanagh, M. J., MacKinney, A. C. & Wolins, L. (1971) Issues in managerial performance: Multitrait-multimethod analyses of ratings. Psychol. Bull., 75, 34–49.

Kershner, A. M. (1955) A report on job analysis. Washington DC, Office of Naval Research, Department of the Navy, 1, 728.

Klingberg, F. L. (1941) Studies in measurement of the relations between sovereign states. Psychometrika, 6, 335–352.

Lazarsfeld, P. F. (1950) The logic and mathematical foundation of latent structure analysis, in Stouffer, S. A. *et al.* (Eds.) Measurement and prediction, pp. 362–412. Princeton, N. J.: Princeton University Press.

Lazarsfeld, P. F. (1959) Latent structure analysis. In Koch, S. (Ed.) Psychology: A study of a science. Vol. 3, pp. 476–543. New York: McGraw-Hill.

Lazarsfeld, P. F. & Henry, N. W. (1968) Latent structure analysis. Boston: Houghton & Mifflin.

Lawler, E. E. (1967) The multitrait-multirater approach to measuring managerial job performance.. J. Appl. Psychol., 51, 369–381.

Lawshe, C. H. (1952) Employee selection. Pers. Psych., 5, 31–34.

Lennerlöf, L.: (1966) Supervisory criteria? Stockholm: Swedish Council for Pers. Adm.

Levine, E. L. & Weitz, J. (1971) Relationship between task difficulty and the criterion: Should we measure early or late? J. Appl. Psychol., 55, 512–520.

Lord, F. M. (1954) Scaling. Rev. Educ. Research, 24, 375–393.

Mabon, H. (1968a) Den statliga personaladminstrativa verksamheten med avseende på delområdet – arbetsbeskrivning och arbetsanalys (Personnel administration within the public administration – job description and job analysis). Mimeo. Report to the Swedish State Rationalization Department (Statskontoret), Stockholm.

Mabon, H. (1968b) Arbetsanalysen – en kritisk granskning (Job analysis – a critical examination.) Internal mimeographed report, Swedish Council for Pers. Adm., Stockholm.

Mabon, H. (1972) Personaladministrativa mätmetoder (Measurement methods in personnel administration) (Second edition). Stockholm: M & B Fackboksförlaget.

Mabon, H. (1973a) Beteendevetenskap för ekonomer (Behavioural science for economists) (Third edition). Stockholm: M & B Fackboksförlaget.

Mabon, H. (1973b) Organisationslärans utveckling (The development of organisation theory) (Fourth edition) Stockholm: M & B Fackboksförlaget.

Madden, J. M. Hazel, J. T. & Christal, R. E. (1964) Worker and supervisor agreement concerning the worker's job description. Pers. Res. Lab., Aerospace Med. Div., Air Force Systems Command.

Magnusson, D. 1967) Test theory. Reading, Mass.: Addison-Wesley.

Mahalanobis, P. C. (1936) On the generalized distance in statistics. Proceedings of the National Institute of Science, India, XII, 49–55.

McCormick, E. J. (1964) The development, analysis and experimental application of worker-oriented job variables. Office of Naval Research Report. Department of the Navy.

McCormick, E. J., Cunningham, J. W. & Gordon, G. C. (1967) Job dimensions based on factorial analyses of worker-oriented job variables. Pers. Psych., 20, 417–430.

McCormick, E. J., Cunningham, J. W. & Thornton, G. C. (1967) The prediction of job requirements by a structured job analysis procedure. Pers. Psych., 20, 431–440.

McCormick, E. J. Finn, R. H. & Scheips, C. D. (1957) Patterns of job requirements. J. Appl. Psychol., 41, 358–365.

McQuitty, L. L. (1954) An approach to isolating dimensions of job success.
Wrigley, C. J. Appl. Psychol., 38, 227–232.
& Gaier, E. L.

Meehl, P. E. (1954) Clinical versus statistical prediction.
Minneapolis: University Minnesota Press.

Messick, S. J. (1956) An empirical evaluation of multidimensional successive intervals.
Psychometrika, 52, 57–66.

Messick, S. J. (1954) The perception of attitude relationships: A multidimensional scaling approach to the structuring of social attitudes.
Ph. D. Thesis, Princeton University.

Militaerpsykologisk arbejdsgruppe (1956) Betænkning vedrørende jobanalyse og manskapsfordelning (Report on job analysis and the classification of military personnel).
Fredriksberg Slot, Denmark.

Miller, R. B. (1962) Task description and analysis. In Gagné, R. M. (Ed.): Psychological principles in system development.
New York: Holt, Rinehart and Winston.

Miner, J. B. (1969) Personnel psychology.
London: Collier-McMillan.

Morsh, J. E. (1962) Job analysis bibliography. USAF PRL tech. docum. Rep., No. 62-2 (IV).

Morsh, J. E. (1964) Job analysis in the United States Air Force.
Pers. Psych., 17, 7–17.

Morsh, J. E., Madden, J. M. & Christal, R. E. (1961) Job analysis in the United States Air Force. USAF WADD tech. Rep., No. 61–113, VII.

Mosel, J. N., Fine, S. A. & Boling, J. (1960) The scalability of estimated worker requirements. J. Appl. Psychol., 44, 156–160.

Mårdberg, B. (1966) Arbetspsykologisk klassificering (Classification in industrial psychology). Mimeographed.
Stockholm: Swedish Council for Pers. Admin.

Mårdberg, B. (1969) Urval och klassificering av personal (Selection and classification of personnel). Mimeographed.
Stockholm: Swedish Council for Pers. Admin.

Mårdberg, B. (1971a) Clustering jobs by latent profile analysis. Preliminary mimeographed report.
Stockholm: Swedish Council for Pers. Admin.

Mårdberg, B. (1971b) Latent profile analysis as a non-linear alternative to factor analysis. Preliminary mimeographed report.
Stockholm: Swedish Council for Pers. Admin.

Mårdberg, B. (1971c) A study of dimensions of job description data by the Q-technique of latent profile analysis. Preliminary mimeographed report.
Stockholm: Swedish Council for Pers. Admin.

Mårdberg, B. (1971d) Job perception at different personnel levels. Mimeographed. Report No. 333 from the Psychol. Laboratories, University of Stockholm.

Mårdberg, B. (1972) Stability in job evaluation. Mimeographed. Report No. 0044.
Stockholm: Swedish Council for Pers. Admin.

Mårdberg, B. (1973) A model for selection and classification in industrial psychology. Mimeographed.
Report supplement 19 from the Psychol. Laboratories, University of Stockholm.

Mårdberg, B. & Baneryd, K. (1971) Psykologisk befattningsanalys (Psychological position analysis). Mimeographed.
Stockholm: Swedish Council for Pers. Admin.

Newman, S. H. & Fine, S. A. (1957) Validity of job descriptions for physical requirements and work condition information. Pers. Psych., 10, 181–189.

Nordisk yrkesklassificering (1966) (Scandinavian occupational classification). Swedish Royal Labour Market Board.

Norris, R. C. (1956) I Rating Dimensions. USAF Personn. Train. Res. Cent. Res. Rep., No. 5–63, VI.

Office of Naval Research (1959) Tri-service Conference on Job Qualifications Analysis. Dept. of the Navy, Washington, DC.

O'Neill, H. E. & Kubany, A. J. (1959) Observational methodology and supervisory behavior. Pers. Psych., 12, 85–95.

Orr, D. B. (1960) A new method of clustering jobs. J. Appl. Psychol., 44, 44–49.

Pallett, J. E. & Hoyt, D. P. (1967) An empirical approach to criterion specification in general business. J. Appl. Psychol., 51, 174–180.

Palmer, G. J. & McCormick, E. J. (1961) A factor analysis of job activities. J. Appl. Psychol., 45, 289–294.

Peres, S. H. (1962) Performance dimensions of supervisory positions. Pers. Psych., 15, 405–410.

Porter, L. W. (1961) Perceived trait requirements in bottom and middle management jobs. J. Appl. Psychol., 45, 232–236.

Prien, E. P. (1963) Development of a supervisor position description questionnaire (SPQD). J. Appl. Psychol., 47, 10–14.

Prien, E. P. (1965) Development of a clerical position description questionnaire. Pers. Psych., 18, 91–98

Prien, E. P. & Ronan, W. W. (1971) Job analysis: a review of research findings. Pers. Psych., 24, 371–396.

Richardson, M. W. (1938) Multidimensional psychophysics. Psychol. Bull., 35, 659–660.

Rulon, P. J., Tiedeman, D. V., Tatsuoka, M. M. & Langmuir, C. R. Multivariate statistics for personnel classification. New York: Wiley.

Rupe, J. C. (1955) Research into basic methods and techniques of Air Force job analysis I, III. USAF Personn. Train. Res. Cent. Res. Rep.

Rupe, J. C. (1956) Research into basic methods and techniques of Air Force job analysis IV. No. AFPTRC-TN-56-51, IX.

Rupe, J. C. & Westen, R. J. (1955) Research into basic methods and techniques of Air Force job analysis II. No. AFPTRC-TN-55-51, VIII.

Saunders, D. R. (1956) Objective description and classification of engineering jobs II: Factor analysis of the key group data form. Educational Testing Service, Princeton, New Jersey.

Schultz, D. G. & Siegel, A. I. (1964) The analysis of job performance by multidimensional scaling technique. J. Appl. Psychol., 48, 329–335.

Seashore, S. S., Indik, B. P. & Georgopoulos, B. (1960) Relationships among criteria of job performance. J. Appl. Psychol., 44, 195–202.

Sheppard, R. N.	(1962)	The analysis of proximities. Multidimensional scaling with an unknown distance function, Parts I & II. Psychometrika, 27, 125–140 & 219–246.
Simon, H. A.	(1947)	Administrative behavior. New York: Collier-McMillan.
Stevens, S. S.	(1951)	Handbook of Experimental Psychology. New York: Wiley.
Studdiford, W. S.	(1953)	New occupational classification structure. Emplmt. Secur. Rev., 20 (9), 36–39.
Super, D. E.	(1963)	The definition and measurement of early career behavior: a first formulation. Personnel and Guidance Journal, 41, 775–780.
Taylor, H. C. & Russell, J. T.	(1939)	The relationship of validity coefficients to the practical effectiveness of tests in selection. J. Appl. Psychol., 23, 565–578.
Thomas, L. L.	(1952)	A cluster analysis of office operations. J. Appl. Psychol., 36, 3, 238–242.
Thompson, A. S.	(1965)	The criterion problem in selection and guidance. Part 2. Occupational Psychology, 39, 83–88.
Thorndike, R. L.	(1949)	Personnel selection. New York: Wiley.
Thorndike, R. L.	(1953)	Who belongs in the family? Psychometrika, 18, 267–276.
Thorndike, R. L., Hagen, E. P. Orr, D. B. & Rosner, D.	(1957)	An empirical approach to the determination of Air Force Job Families. USAF Pers. Train. Res. Cent. Tech. Rep., No. 57-5, VI.
Thorndike, R. L. & Hagen, E.P.	(1959)	Ten thousand careers. New York: Wiley.
Thurstone, L. L.	(1948)	Psychophysical methods. In Andrews, T. G. (Ed.) Methods of Psychology. New York: Wiley.
Tiffin, J. & McCormick, E.	(1965)	Industrial psychology. (5th ed.) London: Allen & Unwin.
Toops, H. A.	(1945)	Some concepts of job families and their importance in placement. Educ. and Psychol. Measmt., 5, 195–216.
Torgerson, W. S.	(1951)	A theoretical and empirical investigation of multidimensional scaling. Ph. D. Thesis, Princeton University.
Torgerson, W. S.	(1952)	Multidimensional scaling: I. theory and method. Psychometrika, 17, 401–419.
Torgerson, W. S.	(1958)	Theory and methods of scaling. New York: Wiley.
Trankell, A.	(1959)	The psychologist as an instrument of prediction. J. Appl. Psychol., 43, 170–175.
Trattner, M. H., Fine, S. A. & Kubis, J. R.	(1955)	A comparison of worker requirement ratings made by reading job descriptions and by direct job observation. Per. Psych., 8, 183–194.
Triandis, H. C.	(1959)	Differential perception of certain jobs and people by managers, clerks and workers in industry. J. Appl. Psychol., 43, 221–225.
Turner, W. W.	(1960)	Dimensions of foreman performance: A factor analysis of criterion measures. J. Appl. Psychol., 44, 216–223.

United States Employment Service (USES) (1956) United States Department of Labor Estimates of worker trait requirements for 4 000 jobs as defined in the Dictionary of Occupational Titles. Washington: Government Printing Office.

Weitz, J. (1961) Criteria for criteria. Amer. Psychologist, 16, 228–231.

Weitz, J. (1964) The use of criterion measures. Psychological Reports, 14, 803–817.

Weitz, J. (1966) Criteria and transfer of training. Psychological Reports, 19, 195–210.

Wofford, J. C. (1970) Factor analysis of managerial behavior variables. J. Appl. Psychol., 54, 169–173.

Wofford, J. C. (1967) Behavior styles and performance effectiveness. Pers. Psych., 20, 461–496.

Young, G. & Householder, A. S. (1938) Discussion of a set of points in terms of their mutual distances. Psychometrika, 3, 19–22.

Index of figures

Name index

Subject index