"Using a differential test battery to illustrate a multi-dimensional theory of intelligence."

by Glen Deborah Fox, M.A. M.Sc. C.Psychol. AFBPsS

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy, May 1994

Table of Contents

Abstract

This study sets out to examine the premises of differential validity, or the use of score differences and patterns as predictors. This presupposes a view of ability or intellect as multidimensional, and therefore regards multidimensional patterns or profiles of scores, in addition to or irrespective of the actual levels of single test scores or weighted composites, as having predictive or classificatory uses.

The advantages of taking a multidimensional view of intellect as assessed by differential testing are contrasted with the advantages of a unitary approach to intelligent performance, which assumes that a test or weighted test composite must be used to create a single index of performance. The study also considers the possibility that psychometric testing, as it is commonly used in selection and development, overstresses levels of performance and under-utilises the amount of information that can be gained from studying patterns of test scores.

The differential test battery which is examined, the Morrisby Profile, is standardised and validated as part of this study by the author while working for the Morrisby Organisation, except where the assistance of others is specifically acknowledged. Methods for validating it both as a traditional and as a differential battery are examined, and various possible indices of differential efficiency are discussed, using multiple regression, discriminant function analysis and MANOVA. A further method is devised and presented for displaying the differential performance of a battery, using deviation scores.

As deviation scores have the effect of making the measure at least partly ipsative, some issues of ipsativity are addressed, and arguments presented to justify the use of statistical techniques with partially ipsative data.

Data is presented to show the relative effectiveness of different indices of validity, employing multiple regression, discriminant function analysis, MANOVA and the use of deviation scores. Although coefficients based on differentials alone rarely equal those based on score levels, combined coefficients are more effective than either, and their use is advocated. It is also argued that there may be real, if less easily quantifiable, advantages to the differential manner of presentation, with particular reference to groups commonly disadvantaged by traditional tests, especially in the field of development and guidance.

The data sets examined included occupational groups, (engineers, technicians, managers, careers guidance officers and teachers), school students with academic criteria, applicants for engineering technician posts, insurance salespeople and managers whose promotional ratings had been assessed. Both categorical and non-categorical criteria are used.

The battery is found to be an effective measure in both a traditional and a differential sense, although against the criteria available it is not possible to establish the absolute superiority of the differential approach in terms of predictive validity in the absence of all information relating to level of scores. It is shown that, when scalar and differential methods are combined, more of the variance is explained than when either method is used alone. In view of the possible disadvantages of traditional validation methods, it is suggested that there would be social advantages in utilising a differential method of testing The implications of differential testing in the context of current perceptions of human abilities are discussed, and possible developments for a differential approach are indicated.

Acknowledgements:

I should like to thank Professor Chris Brotherton for his kindness and encouragement throughout this study, Dr. David Clarke, who has made many helpful suggestions and comments during the course of the work, Tania Malkani, who has been of great assistance in finding references when I could not reach libraries, Dr. Theodora Scholarios, for lending me copies of her own work and that of her colleagues at George Washington University, and Malcolm Morrisby, who has provided much practical help, support and advice, as well as being most generous with both his time and that of his firm. I should also like to acknowledge the part played by the University of Ulster, whose Research Fellowship allowed me to complete this work, and, particularly, the debt owed to the late John Morrisby, without whom this thesis would certainly not have been written.

Chapter 1. Introduction

The intention of this research is to evaluate a differential approach to ability testing, using a differential test battery, the Morrisby Profile, for the purpose of assessing individuals in the context of job selection, promotion, development and career guidance, and to establish an index for assessing differential validity and differential performance. As the Morrisby Differential Test Battery, the original name of the Morrisby Profile, has not been altered or re-normed since the 1950s, this study also entails rewriting, standardising and validating the battery, which in its new form is called Morrisby Profile.

Once the battery has been assessed as reliable and valid according to current standards (BPS Review of Tests used in Vocational Guidance,1993, British Psychological Society), it should be possible to examine its ability to predict performance and classify groups by using it differentially; that is, by examining patterns of score differences and in some cases patterns of paired and grouped score differences. Differential performance can then be compared with traditional performance, using the same data sets, and conclusions drawn as to the contribution made to the selection and guidance process by including the differential approach. Various methods for devising a coefficient of validity for the performance of the battery as a differential, multidimensional measure can be considered, and preferred methods suggested. The usefulness of such a battery in selection and development can then be assessed.

Selection and development are not entirely natural partners in the area of psychometrics. The object of selection has always been, seen from the employers' point of view, to attract, find and retain the best possible individuals for the job in question. (Brotherton 1980). The implication is that individuals may already be ranked along some notional continuum of worth; that actual or potential abilities are differentially distributed among the candidates and there is a "best" candidate to choose. Psychometric tests have long been used to highlight those individuals in possession of desirable personal qualities, who are considered best able to perform the tasks considered important in the job.

The object of development, on the other hand, more comfortably combines the aims of the individual and those of the organisation or wider community, so that career guidance may be more readily included under the heading of development. In development, the intention is to maximise the strengths of the individual, whether for that individual's benefit and from her point of view, or for the benefit of the organisation or community at large. Psychometric tests, again, can isolate strengths and weaknesses and allow development to be tailored to individual need.

Theoretically, it would seem that no conflict need arise between the two areas. A "good" candidate will be selected on the basis of strength and potential strengths, and further development will build on the strengths already there and assist the individual to bring out untapped ability and compensate for weaknesses. Even in this seemingly ideal state of affairs, however, the organisation may need to highlight aspects of the individual she might prefer not to develop, to the detriment of areas which, in her own interest, she might with advantage extend. Current demands for relatively low-achieving arts candidates to fill empty engineering courses in our universities, for example, suggest an unequal relationship between vocational interests and developmental or educational goals.

In an examination of graduate admissions (Powers and Swinton, 1981), the researchers argued for a more multidimensional definition of academic talent, that might serve both the selection and guidance/development functions. They criticised traditional aptitude tests for providing only a limited description of students' academic strengths and weaknesses, and for failing to reflect either the diversity of their backgrounds and experiences or their differential development in other important cognitive areas. A more comprehensive measurement, they concluded, was required.

A common problem with the types of cognitive measures used in selection is that, together with the lack of breadth mentioned by Powers and Swinton, they depend so heavily on score levels. This is made very evident to the majority of candidates for selection; those rejected on the grounds of low test scores. However, their development is also of importance, and depends on maximising their strengths and compensating for their weaknesses. The same tests used to select them may well have highlighted these areas for development, but in their case the weaknesses are used to rationalise their rejection. In some instances, those weaknesses may relate to their response to the assessment procedure, rather than predict any lack of performance in the job (Herriot 1984; Crawley, Pinder & Herriot 1990).

Careers officers "in partnership" with employers are well accustomed to this paradox. The employer asks for the top ten per cent of available clerical workers, and Fred, after assessment by the careers officer, is found to be in that top ten percent. In addition, Fred seems to be well suited to clerical work in terms both of interests and direction of abilities. Fred is sent along for interview, is appointed and acknowledges that testing has helped his selection and his career development. The paradox is dormant.

Henry, on the other hand, is at the twentieth percentile of clerical workers, although in terms of both interest and direction of abilities he would undoubtedly be a better clerical worker than he would be anything else. (He would, for example, be a very poor electrical engineer - he is at the second percentile). The careers officer is in a quandary. Should she, in the light of her duty to Henry, send him for interview for the job in which, although he would not excel, he would do better than he has a chance of doing anything else? Or should she, with her duty to the employer firmly in mind, deny Henry the chance of clerical work, and continue to send along for interview only the top ten per cent?

If there is a direct, linear, one to one relationship between test scores and performance, the dilemma is a moral and ethical one, and, as such, not readily subject to quantification. Fred would be a better candidate; Henry a worse one, despite his evident need, and there the matter would probably have to rest, with the triumph of common sense over compassion. However, it is the contention of this study that there is no such clear relationship.

The size of test scores may indeed indicate ability, although it might be argued that it might also be affected by self confidence or test familiarity. (Howe 1988). However, the relative sizes of test profiles - their "elevation" - is not the only ground on which they may be compared. Fred and Henry shared the "shape" or direction of their test profiles: that is to say, they were better at some things than at others, and Henry's lower profile paralleled the direction of Fred's higher one. Shape may well be a powerful indicator of performance. Parallel profiles at different elevations may have more in common than similarly high (or low) profiles with different "shapes".

If there is predictive power in profile shape, and if the sheer size of test scores may in some circumstances be suspect, there would seem to be commercial advantage, as well as social and ethical satisfaction, in considering Henry for clerical work if Fred is not available. If there is added power in the fact that the job matches his "shape", and if his poor test performance reflects his response to tests rather than his actual test taking ability, then the employer may well be losing a valuable employee by ignoring Henry in favour of higher scoring individuals whose pattern and direction of abilities may not even be as good as his own.

Psychometric test batteries have frequently been used for selection and development, and often their use is described as "differential"; that is, as indicating individual differences between candidates. (Anastasi 1979) These batteries implicitly - and, in some cases, explicitly - advocate the use of a single index of achievement in the form of a weighted test composite score. This implies totalling a candidate's weighted scores and comparing them with others' without taking account of the patterns of individual scores. The system is simple, computationally, to validate, as single indices of performance can be easily correlated with whatever criteria of performance are available, and it has common sense on its side. Interestingly, even those batteries, such as the DAT, which overtly describe themselves as differential, and recommend vocational guidance based on profile identification, only publish validation figures related to single scales, individually correlated with academic criteria. (Hambleton, 1984) Rarely are multiple correlations published, although the logic of the composite score would suggest the sense of such an approach, and no major test battery of ability measures has published truly differential validation figures based on the patterns or profiles they describe, although Cattell has done so in the case of personality profiles. (Cattell 1949,1969).

A differential battery would take both the levels of the scores and their differences into account in assessing a candidate's performance, allowing his strengths and weaknesses to be perceived in relation both to others and to himself.

The desirability of such a differential battery is not difficult to defend; maintaining its validity is rather harder. Much of this work is taken up with describing the processes undergone to test the validity and efficacy of the differential approach, as well as with describing the relevance and usefulness of the approach itself. It has been necessary to explore a number of statistical methods and to devise some new approaches in order to assess the differential approach, and to do this, it was necessary to ensure that the instrument examined conformed to acceptable standards.

As it was considerably out of date and in need of review, that entailed major rewriting and standardising of the measures, as well as carrying out traditional reliability, validity and fairness studies, before its value in differential terms could be assessed. Although this was all carried out by this author as an essential part of the preliminary work for this study, it also provided a very useful insight into the fallibility of traditional correlational methods, and the ease with which they may obscure irrelevance and compound inequalities.

The Morrisby Profile is based on certain assumptions. It presupposes that abilities are multidimensional, and that therefore no unitary scale exists along which individuals may be ranked; that these abilities are differentially distributed, and that individual patterns of ability make a contribution towards predicting successful job performance which is at least as useful as individual levels of ability. A complex system has evolved for the clinical interpretation of scores on the Morrisby Profile, which sets out deliberately to minimise their use as a simple single index of achievement on any criterion, and instead aims to provide a wide range of information on an individual's abilities and likely use of them, which can then be related to a selection, development or guidance decision.

The theoretical basis for these assumptions, and their implications for selection, development and career guidance, are discussed, and some suggestions made for further research.

This work is divided into five chapters following this introductory section. The second chapter is in two sections; the first section covers the literature relating to differential theories of intelligence, and this is followed by the second section; a review of the literature relating more specifically to methodological problems in differential test theory.

The third chapter deals with the work of John Morrisby, author of the Morrisby Differential Test Battery, the earliest version of Morrisby Profile, his conception of the use of differentials, and his presentation of the structure of human abilities. This chapter describes Morrisby's partial solution to the problem of the differential index of validity, and also briefly describes the clinical use of the battery.

The fourth chapter describes the process of standardising the Morrisby Profile as part of this study and establishing the reliability, validity and fairness of the separate tests using traditional methods.

The fifth chapter discusses the problems of validating the differential use of the test battery; traditional and non-traditional methods for validating the differential approach are considered and a new method is described. Data sets are analysed with reference to both levels and patterns of ability, and the results are compared. Different methods of establishing a coefficient of validity are described and the results compared.

Finally, in the sixth chapter, the implications of the work are considered and future lines of research indicated.

Chapter 2. Review of The Literature - 1.

Differential and multidimensional theories of intelligence

Despite the intractability of the problem of establishing the nature of intelligence, many theorists have attempted to define and model a theory of the construct. A discussion of the problems of definition and a number of definitions of the term may be found in the study by the editors of the Journal of Educational Psychology "Intelligence and its Measurement: A symposium (1921).

In this study, fourteen experts offered their definitions of the term. These included:

"The power of good responses from the point of view of truth or fact" (E.L.Thorndike)

"The ability to carry on abstract thinking" (L.M.Terman)

"Having learned or the ability to learn to adjust oneself to the environment"(S.S.Colvin)

"A biological mechanism by which the effects of a complexity of stimuli are brought together and given a somewhat unified effect in behaviour"(J.Peterson)

"The capacity for knowledge and knowledge possessed". (V.A.C. Henmon)

"The capacity to inhibit an instinctive adjustment, the capacity to redefine the inhibited instinctive adjustment in the light of imaginally experienced trials and error, and the volitional capacity to realise the modified instinctive adjustment into overt behaviour to the advantage of the individual as a social animal" (L.L.Thurstone)

"The capacity to acquire capacity"(H.Woodrow)

"The capacity to learn or profit by experience"(W.F.Dearborn).

These implicit theories of intelligence, or informed beliefs about intelligence, probably represent among them versions approximating to conceptions of intelligence held by most lay practitioners in the fields of selection and development. Among them are versions of intelligence which relate the construct to the pure acquisition of knowledge, to its practical application as "experience"; to operating effectively in given surroundings and to judgments relating to abstract absolutes such as right and wrong.

None of the experts describes intelligence as the ability to perform speeded test tasks divorced from context; each assumes it has a real function in the real world. Whether or not "experience", or "the environment" are explicitly mentioned, intelligence is most commonly assumed to be that attribute which enables a human being to function best and on most levels in her surroundings, and the question is begged as to what the surroundings and the levels may be.

The wide range of these implicit theories, coupled with the assumption that the construct exists, is clearly defined and represents a desirable criterion for such activities as job selection, explains the search for a more explicit theory of the structure of intellect.

The search for such explicit theories of intelligence, supported by experimental findings, has included many approaches, such as the physiological view (Hendrickson 1982) , the psychometric view (Spearman 1927), the Piagetian approach (Piaget 1921), the learning theory approach (Thorndike 1924) and the cognitive, or information-processing view (Sternberg 1977) (Irvine 1991), but it is the psychometric view that has predominated, even to the extent that it has been employed to validate most of the alternatives, commonly in the form of "IQ" tests.

The psychometric, or differential, view is based on the assumption that intelligence exists as a quantifiable construct, and that individuals may be ranked in terms of intelligence through the medium of testing according to their score differences. Intelligence is assumed either to underlie or to be identified by a set of abilities (e.g. numerical ability, reasoning ability) which may be identified through the technique of factor analysis. (The "differential" nature of the psychometric view here refers to its dependence on the individual differences between people, and not to the possibility that differences within individuals may be predictive of performance, which is the broad sense in which the term is used throughout this study, although different interpretations of the differential position are discussed later.)

The difference between all psychometric theorists has centred on the relationship and relative importance of "general intelligence" and more specific abilities, and the relationship and arrangement of those specific abilities to one another. Pretheorists, notably Alfred Binet, who devised the scales that later became known as the IQ scale, worked empirically to create a useful measurement of performance, and explicitly denied the intention to reify the construct of intelligence.

Binet insisted on three principles for the use of his tests: that there should be no reification of what was measured by the scales, which defined nothing innate or permanent; that the scales should be used to identify learning-impaired children, not to rank normal children; and that low scores should be used to identify those in need of special help, not brand them as congenitally incapable.(Binet 1909).

However, Binet's principles were largely ignored, and the mass testing that began in the first world war established a new principle; that all human beings could be ranked on an IQ scale and rated accordingly for their relative desirability as employees, immigrants or candidates for eugenic limitation. Psychologists seized on the atheoretical concept of measurement, endowed it with a pedigree and then used the assumptions of the new tool to argue its internal validity.

Three main types of theory have emerged: hierarchical theories attributing most importance to a general intelligence factor, but allowing for the existence of certain specific abilities, subsidiary to the main factor; horizontal theories attributing equal importance to a number of general abilities, but denying supremacy to a more general factor, and multidimensional or componential theories, which perceive intelligence in terms of the relationships of various groups of abilities.

The earliest hierarchical theories as to the nature and dimensionality of intelligence were those advocated by Spearman (1904) in his two-factor theory. Spearman argued that individual differences in performance on any test could be accounted for by one general factor which was common to all tests, and by one which was specific to that test (e.g. number recognition). The former he labelled "g", or general intelligence, and the latter he called "s", or specific ability. The general factor he characterised as "mental energy", which he saw as influencing performance on all the tasks used to assess intelligence.

Although Spearman perceived himself as opposed to "monistic" or single factor theories, and claimed that his two factor theory gave greater weight to aspects other than "g", in fact his theory greatly stressed the importance of "mental energy" in assessing performance on all intellectual tasks, and he was reluctant to concede that more than two factors, "g" and the relevant specific factor, might affect results on a single test.

Spearman was often accused by his critics of selecting his tests in such a way as to fit his two-factor theory, and eventually he conceded that other "group factors" might exist, together with two other specialised general factors which he classified as "perseveration" and "oscillation". However, he always asserted that the general factor is central and supreme in all tests of intelligence, and tentatively suggested that it might have demonstrable physiological correlates, anticipating some of the work of Eysenck and Hendrickson later in the century.

Spearman's theory gave rise to a number of more overtly hierarchical theories which attempted to account for the finding of more than one factor common to several tests in a battery.

The hierarchical theory of general abilities put forward by Burt (1940) and Vernon (1950) maintains that individual differences may best be accounted for in terms of one general factor, common to all tests, and a series of group factors (e.g. numerical ability), each of which has positive loadings on some tests and zero, or near zero, loadings on others, in addition to the specific factors mentioned by Spearman. The number of group factors depends upon the number and variety of tests used. In work on Army and Navy conscript recruits, Vernon (1947) found that 'g' covered more than twice as much variance as all group factors combined.

In an analysis of thirteen tests given to 1000 Army recruits, comprising progressive matrices, dominoes (a non-verbal test), a group test, squares, assembly, the Bennett mechanical test, verbal dictation, ATS spelling, instructions, and various arithmetic tests, after the removal of "g", the tests fell into two groups; verbal/numerical /educational, described as the v:ed factor, and practical /mechanical /spatial /physical, referred to as the k:m factor.

This theory is supported by methods of factor analysis which maximise the total amount of variance which is due to the common factor, and it is an extension of Spearman's two-factor theory rather than a rebuttal of it. It has been defended by Burt (1955) on the grounds that it produces solutions which are closest to the way the term "intelligence" is generally used. He concludes that the large number of factors claimed by multiple factor theorists such as Guilford, and the lack of agreement among researchers, are partly due to certain apparently irrelevant aspects of test performance such as test sophistication.

However, test sophistication might instead equally well be seen, by proponents of multiple factor theories, as the apparent general factor underlying all test success, in contrast to Burt's view, which perceives it as responsible for throwing up spurious factors.

The claim that theories of intelligence are validated because they correlate with implicit theories of intelligence, or lay beliefs about the structure of intelligence, is not new. Hereditarians and racial supremacists have for some time been prone to seize on corroborative theories of intellect. This has not been a tendency confined to lay practitioners; Jensen, for example, used Cyril Burt's allegedly fraudulent twin data (but see Joynson 1989 for a rebuttal of the accusations made against Burt) in supporting his theories of white superiority.

Both hierarchical theories and multiple factor theories have appealed to lay perceptions of intelligence for support, as if semantic definition were synonymous with evidential proof. Preference for a congenial model of intelligence has activated practitioners as well as lay "experts", and it is socially and historically interesting that the egalitarian multiple factor theory came from the democratic United States, whereas it was in Britain that the hierarchical theories evolved.

Thurstone (1924,1940) and Guilford have been among the most influential proponents of the multiple factor, or horizontal, view of abilities. Thurstone developed a different application of factor analysis from that used by Spearman, and in 1931 he presented a new model for factor analysis that generalised Spearman's two-factor model to multiple factors.

His monograph "Primary Mental Abilities" (1938) identified at least seven independent factors of mental ability. Thurstone employed a criterion for selecting a factor solution which he called "simple structure", by which he meant a factor solution which maximises the number of zero loadings. Ideally, each factor would have high loadings on some tests and zero loadings on others.

Although Thurstone maintained that such a solution made most psychological sense, it necessarily minimised the likelihood of finding a general factor, and some of his colleagues were dismissive of his claim to have established the existence of more specific factors. "In 1935 he (Thurstone) wrote ... about a factor analysis he had done with a battery of 30 tests in which he extracted 15 factors. Later he and his colleagues found about 30 more. They were about as stable as soap-bubbles and I expect they are all forgotten now."(Vincent. D.(1980) pers. comm to C. Brotherton.)

Since tests in fact tend to correlate positively with one another, it is difficult to achieve "simple structure", and so Thurstone was forced to utilise oblique factors, and generate a "second order" general factor to account for the correlation between the primary factors. His theory thus conceives of intelligence as a loosely related group of "primary abilities" which are labelled V (verbal) P (perceptual speed) I (inductive reasoning) N(number) M (Rote memory) D (deductive Reasoning) W(word fluency) and S (space or visualisation).

Spearman countered this theory by pointing out that, as all Thurstone's tests inter-correlated positively, they could equally well be analysed to yield a large general factor, similar to "g", and smaller group factors.

Holzinger, Harman and Eysenck carried out just such a study in 1939, and obtained a "g" factor and a number of group factors whose content corresponded closely to that of Thurstone's primary factors.

It has been argued (Carroll 1982) that Thurstone's multiple factor theory may now probably be perceived as interchangeable with the hierarchical view of Vernon; any covariance found among obliquely rotated primary factors by Thurstone's methods can be redistributed to the general and broad group factors of Vernon's model, using the Schmidt-Leiman technique (1957), or the orthogonalization technique of Cattell and White (1965). However, principal components analysis can as easily be accused of forcing a general factor on data. Both are only techniques for organising data, not manifestations of hidden truth.

Guilford (1967) also argued for the existence of parallel abilities. He claimed that even the multiple factor theory offers only a simplified concept of intelligence and fails to do justice to its richness and variety. Accordingly, he constructed a complex theory which first distinguished five types of mental operations: thinking, remembering, divergent production (problem solving which leads to unexpected and original solutions), convergent production (problem solving which leads to the one correct solution), and evaluating.

Then he sub-divided each of these operations into six products: units, classes, relations, systems, transformations and implications; and, finally, he identified four types of content upon which the operations are performed: figural, symbolic, semantic and behavioural.

These classifications generated 120 distinguishable "abilities", and Guilford has constructed test batteries designed to yield factors generally corresponding to these abilities, most of which have been defined factorially.

However, Guilford's work has been criticised on both methodological and logical grounds. Horn points out that he accepts factors which appear to fit the data on the basis of subjective judgments, ignoring correlations which may be due to chance. (Horn 1967). Carroll argues that there is nothing in Guilford's data that requires or even positively indicates the classificatory constructs that he postulates, since the factors are claimed to be orthogonal and independent. (Carroll 1972).

Nevertheless, Guilford's research did encourage the growth of interest in multifactorial batteries. The Guilford-Zimmerman Aptitude Survey (1947-56); the Differential Aptitude Test (1947) and, more recently, the International Primary Factors Battery (Horn 1973) and the Comprehensive Abilities Battery (Hakstian & Cattell 1976) are all based on the research identifying multiple factors of mental ability, although they do not necessarily rely on Guilford's implicit assumption that the general factor of intelligence has been entirely swallowed up by the "operations".

A further contribution made by Guilford to the study of human abilities was his three-dimensional presentation of their inter-relationships. The concept that abilities may be structured other than hierarchically is one that underlies attempts to redefine human abilities in terms of shape rather than size, which will be a recurring theme in the discussion of Morrisby's portrayal of the structure of abilities.

Guilford's cubic presentation of abilities also influenced Morrisby in his own cubic presentation of the personality factors, or modal profile, which he includes in his profile of abilities, as Sternberg was to do later in his triarchic theory. Although Guilford did not include any overt "personality" factors in his theory, some of his operations, such as divergent thinking, may touch on the fuzzy area of thinking style where intellect and personality merge.

Personality factors were given prominence in Cattell's approach, which is also closely related to hierarchical thinking. In 1943, Cattell had distinguished two group factors standing alongside 'g' or a general factor. The group factors were seen as conceptually distinct but correlated; labelled fluid and crystallised intelligence.

Crystallised intelligence is utilised in those tests which require learned habits of thinking, and fluid intelligence is employed in new situations where successful adaptation cannot be achieved by the individual's existing repertoire of cognitive skills. Both factors had associated with them a larger number of factors representing abilities similar to Thurstone's primary mental abilities.

Cattell argues that fluid intelligence has a greater association with genotypic difference than crystallised intelligence, which may be acquired rather than inherited, and predicts for fluid intelligence a different growth curve, which reaches maximum level earlier than the growth curve for crystallised intelligence.

Cattell claims that earlier factor studies failed to isolate the nature of general intelligence definitively because they have only used cognitive tests. He argues that it is necessary to include personality measures in the battery to isolate intelligence, and, using the Hakstian/Cattell Comprehensive Abilities Battery, which included such measures, he managed to extract the two correlated factors demanded by his theory.

In Cattell's model, larger numbers of factors depended from fluid and crystallised intelligence. The analysis of his model (Hakstian and Cattell (1978), analysed 20 primary factors, and derived them from 6 second level factors (fluid intelligence; crystallised intelligence; visualisation capacity; general retrieval capacity; general perceptual speed, and general memory capacity). These in turn were accounted for by three oblique third order factors; original fluid intelligence; capacity to concentrate, and school culture. Such a model implied hierarchy; in a hierarchy, the terminal branches of a tree will proliferate.

There are very few theorists within the psychometric tradition who fall properly outside the hierarchical or the horizontal camps, and may be perceived as multidimensional theorists. The earliest was probably Guttman, with his radex theory (1954,1965), although this theory contains many of the concepts central to hierarchical thinking.

This "radial expansion of complexity" combined the concepts of differing abilities and differing orders of complexity between and within those abilities. Thus there may be a hierarchy of verbal tasks, ranging from simple through to complex, but this hierarchy is itself placed within the radex structure on a point more or less distant from "true intelligence" than, say, a hierarchy of reasoning tasks.

This radex theory anticipated the possibility that intelligence may in itself be indefinable and untestable, and that it may only be approached by means of a structure of lesser abilities.

A more recent theory of this nature, although it also owes something to that of Cattell, is propounded by Carroll (1986), who states that abilities differ in the degree of generality with which they appear to relate to tasks. He claims three orders of abilities; general intelligence, general abilities and specific abilities, rather than the general and specific orders of Spearman and the versions of his theory which followed.

General intelligence, in Carroll's theory, is not measurable as an entity, but may be inferred from the tendency of second order abilities to cluster together.

The seven general clusters of abilities which make up Carroll's "general abilities" are classified by him as: fluid and crystallised intelligence; visual and auditory perception; mental speed; idea production or fluency, and memory capacity. Most of these aspects of intelligence were included in the Morrisby Profile, to be described later.

Irvine (1990) comments that the importance of Carroll's theory lies in the capacity given by a limited number of factors "to express relationships parsimoniously ...... general intelligence can coexist with a limited number of constructs that help to define it scientifically, and whose tests define it operationally.....we are unlikely to need a great many tests to encompass the major dimensions of human ability."

However, Carroll's theory may be even more important in its readiness to accept that the aim of a theorist working in this field is not to isolate and then test general intelligence, as if its existence as a theoretical construct made it as accessible and definable as height or weight, but rather to consider what separate aspects of intelligence make for success in different tasks or fields.

This approach parallels Sternberg's sub-theory of the contextual framework for intelligence, arguing that a large part of intelligence consists of the capacity to respond to one's environment, whatever it may be. This idea was echoed by Frederiksen (1986) who argued for a model of intelligence that was no longer static, but varied as subjects change and circumstances alter. In tests simulating real-world problem situations he showed that the cognitive processes involved in taking a test are influenced not only by test format but also by the situation or setting in which the test is administered, and by such personal characteristics as the examiner's level of expertise. This finding has been replicated several times, as in the unusual study by Philip Saigh. In this experiment, matched groups of Catholic students were administered the WISC by different examiners. Those whose examiner wore a large cross performed significantly better on all but one of the WISC tests than students whose examiner wore an equally large Star of David. (Saigh 1986).

This more pragmatic approach to intelligence echoes the recent challenge to the very existence of the construct, which has also called in question the usefulness of the whole psychometric tradition.

In a comprehensive review of intelligence measurement (Carroll 1988), Carroll questions the assumptions underlying the search for a general factor of intellectual ability. He points out that correlations between two performances may arise for reasons other than their joint dependence on some such capacity as "general intelligence".

The responses involved might have been learned together for some quite arbitrary reason, or learning one of the performances might be prerequisite to the learning of the other performance. He also observes that the kinds of materials that go to make up a test battery are usually limited in their diversity, even though they may superficially appear wide in content, and suggests that even if a general factor is found from such a battery, it may be somewhat less "general" than is claimed.

Possibly, that general factor merely reflects the extent to which test candidates have mastered and retained the skills required to perform test tasks - language skills, skills of attention and concentration, problem-solving algorithms, facts about the number system, vocabulary and so on.

If this is so, the general factor may then only be a measure of total learning, partly attributable to such situational variables as differential exposure to education, styles of child rearing and so on. The general factor may therefore measure intellectual experience rather than capacity.

The close relationship between learning and tested intelligence suggests that when students have become accustomed to new types of tests and examinations, they become more adept at recognising unsuspected clues, in reading the passages carefully, at guessing wisely, apportioning their time carefully and so on. (Estes 82). This would enhance correlations between tests of different subjects on different functions, and for this reason such tests have been regarded as unhelpful in distinguishing between abilities because they inter-correlate highly with most aspects of ability. (Quereshi 1972; Kline 1993).

It could be argued that, if a test does in fact correlate highly with most aspects of ability, there would seem to be prima facie grounds for regarding it as a fair measure of the major factor underlying ability. This is probably a simplistic argument. That major factor may not be intelligence. It may reflect education, class, handwriting skills, test stamina or good eyesight. It may simply represent being good at test-taking.

Howe suggests that the correlations of more meaningful learning tasks with intelligence can be accounted for by the shared presence of some unspecified combination of any of a variety of qualities that are extrinsic to the learning process as such: for example, previously acquired knowledge, strategies and skills. (Howe 1976). He concludes that high measured intelligence merely indicates the presence of unspecified causes that make a person intelligent.

He admits that acquired knowledge and skills, that make an intelligent person good at performing meaningful learning tasks, are all aspects of human intelligence, but concludes that this is only to argue that success in meaningful learning tasks, by people who score highly in intelligence tests, results because success in the tasks depends on the possession of the qualities which make up "intelligence" - a circular argument.

More recently (1988) Howe argued that intelligence is used as a descriptive rather than an explanatory concept, and he denied there exists any evidence that basic abilities to reason or to solve problems in general are involved in intelligence. He lists ten conceivable states of affairs, not necessarily mutually exclusive, which he claims would help to explain success at tests of mental ability, if proved to contribute to or to be indicated by measured intelligence.

These ten hypothetical situations are that levels of intelligence are reliably related to: 1) observable physiological variables; 2) variability in basic mental processing mechanisms; 3) the capacity to learn or remember; 4) fundamental thinking skills, by which he means basic ability to solve problems; 5) the ability to reason abstractly; 6) the complexity of a person's cognitive functioning; 7) mental flexibility; 8) executive controlling functions, such as planning and organisation; 9) biological mechanisms and 10) specific patterns of ability.

In what he sees to be the absence of reliable connection between test scores and one or more of the above, Howe argues that the different mental skills of a person are largely independent of one another, and that apparent constructs such as abstract thinking are in fact strongly situationally dependent.

Sternberg, who might be expected to agree with the last contention as echoing his triarchic claims for the part played by the individual's response to experience and to the external world, defends the psychometric tradition by countering (1988) that a substantial body of literature to the contrary does in fact exist and has been reviewed in several places: (Johnson-Laird 1984; Pellegrino 1984).

Sternberg himself, before formulating the triarchic theory, found high correlations between conventional intelligence test scores and both inductive and deductive reasoning (Guyote & Sternberg 1981; Sternberg 1980; Sternberg & Gardner 1981).

Although Sternberg agrees with Howe that knowledge plays a part in these correlations, he argues that that role is not exclusive, as the correlations occur without regard to content, and he points out that the knowledge is unlikely to have been there upon birth, but "rather was acquired and retained through the application of the processes of intelligence." The multidimensional approach depends on perceiving the inter-relationship of experience and capacity, and developing measures that describe the individual in broader and more meaningful terms than can be allowed for by a single index describing a quantity of knowledge acquired, the speed of a response or an IQ quotient.

Campione, Brown and Bryant (1984), in their review of the literature on learning, memory and intelligence, pointed out that substantial correlations may be found between learning and intelligence whenever the tasks are meaningful and ecologically valid, although no such correlation is found when trivial tasks are employed (Estes 1982).

Sternberg and Powell suggested that the reason that tests of the products of learning, such as vocabulary tests, predict overall psychometric intelligence so well is that they represent a proxy for meaningful learning of concepts in context.

In their research, they found high correlations ($r = 0.6$) between learning from context scores and psychometric intelligence test performance. Their work, both in the areas of intelligence theory and in psychometric validation studies, suggests it is a little early to abandon all hope of reflecting some form of intellectual functioning by using tests. The question is what model of intellectual functioning would best fit present need, and what form of psychometric mapping best represents it.

However, recent developments in theories of abilities measured by mental tests suggest a less comfortable association between the cognitive, psychometric and physiological schools. This owes something to parallel processing theories derived from information technology. The performance of a mental task is perceived as involving parallel or sequential operation of mental processes upon inputs from external stimuli or from internal memory stores, governed by learned executive routines, resulting in further operations or responses involving either internal memory stores or peripheral motor systems.

This approach has permitted a move away from the psychometric test as the unit of analysis. Instead of looking for a test variable related to "g", researchers have isolated the average evoked potential (Eysenck & Barrett 1985), the electroencephalogram (Pollock et al. 1989), glucose metabolism in the brain as expressed by PET scan (Haier et al., 1988) and reaction time (RT) on elementary cognitive tasks (Jensen 1987).

Hunt, Frost and Lunneborg (1973) assumed that these mental processes rely on parameters such as speed, capacity of memory stores, probability of transfer between different memory stores, and so on, and correlated performance on mental tests with performance on tasks specifically designed to test these parameters.

Jensen's view was that individual differences in mental ability are essentially related to speed of information processing, and he directed his attention to the question of whether a general mental speed factor exists, and whether elementary cognitive tasks are inter-correlated. (Jensen 1987). He concluded that they do share a degree of variance (roughly 30%) but that group factors also exist. He agreed with Detterman (1987) that various mental tests correlate highly with one another because each draws on many of the same elemental processes, although he claimed that the correlations of individual differences in conceptually distinct processes (as measured by elementary cognitive tasks) argued the presence of a fundamental, possibly neurological, level of processing.(Kranzler & Jensen 1991)

In a lengthy debate, (Carroll 1991a 1991b) (Kranzler & Jensen 1991a 1991b)Carroll questioned Kranzler and Jensen's findings on the non-unitary nature of "g", arguing that they confounded their test variable with what it sought to measure. In a factorial re-analysis of their data, he found many elementary cognitive tasks have substantial loadings on a second order general factor, which was orthogonal to a factor representing general speed of movement. He argued that "g" might well be interpreted in terms of general speed and efficiency in information processing. Kranzler and Jensen agreed the central role of speed and efficiency in "g", but maintained it depended on a number of independent processes.

In the last decade, tests of divided attention, emphasising working memory capacity, have been produced, which can be delivered and scored on microcomputers.(Irvine, Dann & Anderson 1990) These tests rely on the association described by Carroll between general intelligence and attributes such as speed of encoding, auditory perception, mental speed and memory capacity.

Irvine et al. claimed that "IQ test performance may be explained by individual differences in encoding, comparison and reconstructive memory processes". As may be seen from the careful wording of this claim, the emphasis has moved from attempts to define intelligence to attempts to define the nature of test performance.

This is more than a merely semantic distinction. If theories of intelligence can only be verified by tests, which themselves can only show test performance, then test performance is the subject of the theorists, not intelligence itself. Theories of test performance, on the other hand, are much more easily verified, and any attempt to substantiate a theory relating to mental ability must assume that, even if such a construct as intelligence exists, the technology, perhaps fortunately, does not yet exist to isolate it infallibly.

A positive theory of intellectual development, which argues in much more detail both the need for a broader, less "g" oriented approach and for a closer marriage between the psychometric and cognitive traditions of intelligence research, has been that of Sternberg (1985), with his triarchic model of intelligence. This model depends on a complex patterning of two triads and a duet within a quartet.

The outer quartet relates to the individual's internal world, the external world, the individual's experience and the individual's personality. The first triad, the internal world, consists of the three internal components; of human internal intellectual functioning, which he calls metacomponents, performance components and knowledge acquisition components. These may be somewhat oversimplified as problem-solving, dealing with inferences and analogies, and learning and applying what is learned.

The second triad, the external world of the individual, is portrayed as her/his practical response to the environment, and it can be described as adaptation to, selection of or shaping that environment. The duet, the experience of the individual, consists of the way the individual copes with novelty and how quickly he/she develops automatizing processes for new situations.

The whole triarchic pattern is modified by the fourth factor in the quartet; personality, which consists of twenty features, including motivation, perseveration, confidence, initiative, follow-through and the ability to distinguish the whole from its component parts. It is notable that many of the personality features, including all of those mentioned above, are explicitly measured in the Morrisby Profile, discussed later, and there are the other strong parallels between the triarchic model and the Morrisby Profile, developed independently by Morrisby over thirty years before Sternberg published his work on the triarchic model.

Although Sternberg has developed a number of separate tests to assess aspects of his model, he has not yet turned his attention to a way of assessing the full model, or to devising an index of efficiency for such a measure. It is interesting to note that, several decades before the conception of the triarchic model, Morrisby had begun to devise a measure assessing many of the aspects of the later model, and had started to address the problem of its validation.

A comparison of the two models shows their similarity, especially in the ability of the Morrisby model to allow for fluid intelligence. This is perceived primarily as understanding or the ability to deal with new concepts, the capacity to process information in several modes, to understand and generate new ideas, to acquire, process and present knowledge. to deal with the practical world, to use strategic or tactical planning styles, and to use abilities in the light of the particular personality structure of the individual.

Figure 1a. Sternberg's triarchic theory

1. **The internal world of the individual:**
   - Metacomponents
   - Performance components
   - Knowledge acquisition
2. **The experience of the individual**
   - Coping with novelty
   - Automatisation
3. **The environment of the individual**
   - Adaptation
   - Selection
   - Shaping
4. **(Negative) /Personality Factors**.
   - Lack of motivation
   - Lack of impulse control
   - Lack of perseverance, or perseveration
   - Capitalisation on the wrong abilities
   - Inability to translate thought into action
   - Lack of product orientation
   - Task completion problems & lack of follow-through
   - Failure to initiate
   - Fear of failure
   - Procrastination
   - Misattribution of blame
   - Excessive self-pity
   - Excessive dependency
   - Wallowing in personal difficulties
   - Distractibility & lack of concentration
   - Spreading oneself too thick or too thin
   - Inability to delay gratification
   - Inability/unwillingness to see the forest from the trees
   - Lack of balance between critical analytic thinking & creative synthetic thinking
   - Too little/too much self-confidence

Fig 2. Morrisby Profile structure

1. **Abstract Reasoning**
    - Understanding concepts
    - Coping with novel problems
2. **General abilities**
    - Information processing
    - Verbal,numerical, perceptual media
3. **Practical intelligence**
    - Spatial
    - Mechanical
    - Planning style:
        - *Strategic - wood*
        - *Tactical - trees*
4. **Problem-solving**
    - Inventiveness
    - Learned methods and solutions
    - Trial and error
5. **"Modal" profile**
    - Flexibility
    - Perseveration
    - Awareness
    - Motivation
    - Inner conviction
    - Outward confidence
    - Initiative
6. **Manual dexterity**
    - Speed
    - Skill

Most of the horizontal and hierarchical theories relating to the structure of intellect have assumed that intelligence, and those general abilities which are associated with it, may be measurable, and the instrument of choice has been the normative psychometric test, standardised against a reference group. Three possible objections to this approach suggest themselves. The first is a practical difficulty; intelligence and general abilities may vary between individuals in relation to other factors not subject to measurement. For example, people may function very much more intelligently in familiar than in unfamiliar surroundings, and familiarity of surroundings is a factor difficult to keep constant with heterogeneous groups of candidates.

The second is a conceptual difficulty; relating score levels to a reference group, even one which is occupationally distinct, implies that performance is distinguished by size, rather than direction, of score.

The third is a social, moral or ethical difficulty; is it right to label individuals as potential successes or failures on the basis of tests?

Although none of these objections can be fully met, some partial answers are available from those subscribing to the "multidimensional" theories, rather than the horizontal or hierarchical ones. These theories take into account an individual's capacity to respond to novelty; they attempt to build a picture of the individual's whole pattern of intellect, rather than finding her place on a single index of performance, and, in many cases, they have been developed specifically with the needs of education and development in mind. Sternberg, for example, has attempted to build a whole educational system on his triarchic theory, which he sees as the foundation of education rather than assessment. The Morrisby system, too, has a place in educational and training establishments, in the stress it lays on the value of diversity of abilities. If psychometrics - the measurement of mind - is not to be perceived as an oxymoron, concepts of measurement may need to be less rigorously quantitative, so that the discipline can allow itself to be used positively and constructively.

Review of the Literature - 2.

Differential validity

General cognitive ability theory assumes that a single measure of general cognitive ability, or "g", will be most effective in predicting the performance of individuals in a wide variety of areas of intellectual performance. Like the psychometric view, on which it depends, its use would therefore be particularly convenient for employers and selectors in that its predictive validity may be generalised across situations so that they need not seek to validate different predictors for each job.

This general cognitive ability appears to account for the largest proportion of criterion variance when compared with other predictors, and it has therefore been assumed that there is little practical gain in terms of incremental validity in including any other predictors in a test battery (Schmidt 1988; Jensen 1986)

However, this theory focuses on the factor structure of the predictor space, while ignoring the relevance of the joint predictor-criterion space. It presupposes that the goal must always be to maximise incremental overall predictive validity, ignoring the different needs of job assignment and classification and favouring predictors which provide the maximum average correlation with the separate performance criteria, taking into account the effects of all previously selected predictors (Schmidt, Hunter and Larson 1988; Ree and Earles 1991)

When the intention is to maximise performance of individuals in different jobs, it becomes apparent that several factors may differentially predict performance in various situations.

The original concept of differential prediction can be traced back to Kelley, who published a method for determining what proportion of the differences between scores on a pair of tests could not be accounted for by chance,(Kelley 1923). In 1934 Segel followed Kelley's method in the experimental work he describes in "Differential diagnosis" (Segel 1934), and in 1947 Bennett and Doppelt used his method on which to base their work on the Differential Aptitude Tests.

The assumption on which differential prediction is based, using the phrase in the sense in which it was used in the 1930s and 1940s, rests on the nature of the association between a set of test scores and a number of criteria. Although it is common for a composite score to be used to predict performance on a single criterion, a composite score is unlikely to be as useful as a differentially weighted score where multiple criteria exist.

> "In many situations, it would be possible to hire from the same population of applicants for a number of possible jobs, and the problem of deciding the best job in which to place each applicant becomes important. It is possible in such a situation to develop a single composite measure having the highest average validity for all jobs, or to develop a battery of tests with differential weighting for each job.
>
> In choosing between these two alternative approaches, it is usually recognised that differential weighting is likely to result in higher validities, on the average, than those obtained when the same weighted composite is employed for all jobs" (Brogden 1951)

In other words, differential weighting provided for the likelihood that different patterns of performance would be likely to predict differences between high performers in different fields of work.

Most test batteries have been devised in the belief that each of the separate subtests would measure a separate aspect of the criterion. It has always been perceived as preferable, therefore, if the subtests inter-correlate as little as possible, so as to avoid redundancy, and it has been presumed that either a simple composite score, or a weighted composite score, should be measured against the criterion to give an indication of the validity of the test battery.

However, there is every likelihood that in fact tests will intercorrelate. If the traditional, hierarchical view of intelligence prevails, and there is a general factor of intelligence underlying all mental abilities and separate groups of abilities, all of these will intercorrelate to some degree because of the general factor of intelligence which they share, and some of these will intercorrelate to an even greater degree because of their shared membership of a sub-group of abilities. Even if the hierarchical view proves not to be correct, and the general factor usually isolated bears little relationship to intelligence, and relates to test-taking skills, or some other construct, its existence would still cause intercorrelations between tests, as they are caused by the existence, not the intellectual nature, of that general factor.

If tests are used to differentiate between performance on several criteria, rather than to predict performance on a single criterion, Brogden showed that, as the degree of correlation between differential predictors increases, the reduction in gains from the use of differential predictors is itself reduced; a finding replicated in experiments described in the fourth section of this work. Brogden also showed that the curve of the loss of selective efficiency has marked positive acceleration, so that approximately one half of the gain resulting from the use of differential predictors is still obtained with a correlation between the predictors of 0.8.

He argued that this showed too much emphasis need not be placed on a search for subtests of a differential prediction battery with low inter-correlations; a radical departure from the conventional demand for uncorrelated subtests.

Although the early work on differential theory stressed the importance of specific abilities, the preferred tool in the 1940s and 50s for establishing the validity of differentiating tests was still multiple regression. Once a test battery had been validated against a set of criteria, using multiple regression, equations were developed, based on the beta weights for each variable when measured against the criteria, which were usually specific job performance measures, and these were then used to compute a composite score for each candidate for each job or group of jobs. It should be stressed that, at this time, the differentiation was regarded as between criteria rather than between predictors. Multiple regression seemed, therefore, an obvious tool to use, in view of its acceptance of the need for several predictor variables, but its linear nature and the assumptions implicit in the calculation of a composite score were not wholly "differential" in nature in the sense in which the term is used later in the study, to refer to profile patterns and shape rather than to scores alone.

This was compounded by the fact that the underlying linear assumptions of multiple regression and true differential validity, in the sense in which it is used in this research, are not identical. Nevertheless, in the 1940s and 1950s most of the work on differential testing centred on methods of calculating differential validity coefficients. Linearity was assumed, which was why multiple regression was the preferred method, as it allowed differentiating Beta weightings to be used as weightings for multiples for subtests once a multiple correlation with the criteria had been established.

John French, in a study which otherwise recommended multiple regression as the method of choice, pointed out that discriminant analysis was a useful method to employ in cases where criterion scores were unavailable, or excessively restricted in range. (French 1955) Now that there is much greater awareness of the fallibility of criterion measures, it is reasonable to assume that discriminant analysis is worth a second look, and this topic will be considered later in more detail.

French made several assumptions regarding the limitations of discriminant analysis, which were reinforced by the experimental method he chose to illustrate the superiority of multiple regression.

French took four groups of candidates; beauticians, office workers, mechanics and carpenters, and gave them a battery of tests. Four validity coefficients were available to him in the shape of the subjects' vocational shop course grades. He used each of the validity coefficients as the dependent variable in a multiple regression for each job, and decided that the regression would select appropriate candidates based on his agreement with the various combinations of the tests indicated by each regression analysis as correlating most highly with the validity coefficient.

He then performed a discriminant analysis, with group membership as the grouping variable, and decided that the results, which indicate typical rather than ideal profiles for each job, would not necessarily produce the best candidates for all four jobs.

However, this was not necessarily to compare like with like. Discriminant analysis is, as French mentioned, commonly used to determine group membership, using a categorical rather than a linear variable to distinguish the groups. However, the technique can be employed in a method much closer to multiple regression than French describes, and does also allow for weighting of individual variables. This permits the construction of an equation similar to that employed in multiple regression for subsequent selection.

Indeed, the grouping variable, like the dependent variable in regression analysis, can as easily be success or failure, and is not limited to occupational indicators. Had French used an indication of vocational achievement as a grouping variable - high and low performers in shop course grades, maybe - it would have been easier to see which approach better predicted successful performance.

French was not alone in seeing multiple regression as the preferred method for establishing differential validity coefficients, although various methods were propounded for calculating multiple correlations with multiple criteria.

However, multiple regression does not allow for more than one dependent variable, whereas the usefulness of the differential approach is most apparent in its capacity to classify individuals according to several criteria, either for selection, where more than one criterion operates to cumulative effect, or in classification, where different combinations of criteria may obtain for different jobs. Nor does multiple regression allow for any relationship between the predictor variables, whose conceptual independence is assumed. Most importantly, multiple regression assumes a linear relationship between scores and criteria. Differential theory, on the other hand, argues for non-linear relationships, with, in addition, strong relationships between certain groups of predictor variables to influence their association with the criterion, which cannot be assessed by the multiple correlation coefficient of regression analysis.

Although there has been an undercurrent of interest in differential theories, especially in the field of vocational guidance, over the last thirty years, much of the work done in the 1940s and 50s had been ignored in favour of general cognitive theory, and the phrase "differential validity" (Hunter & Schmidt 1978,1979; Hattrup & Schmitt 1990) has been most often used in recent years to describe the concept that a test might be differentially valid for separate groups of people, with obvious implications for vocational guidance and equal opportunities research.

However, the original research on differential validity theory concentrated on the need to develop test batteries that differentiated between performance on various criteria, usually different jobs or job fields. Brogden's work, for example, directly linked measurement of classification efficiency to mean predicted performance and hence to utility. The term "differential validity" was used to describe the problem of validating a differential battery, or a battery designed to maximise individual differences between and within subjects for the purpose of classification and utility analysis. (Brogden 1955)

In the last five years, almost all the work relating to this form of differential validity theory has come from Joseph Zeidner, Cecil Johnson and Dora Scholarios and their colleagues at George Washington University, where Paul Horst did his original research. (Zeidner & Johnson 1990; Johnson, Zeidner & Scholarios 1990).

Building on the work of Horst and Brogden, they have been studying the possibility that weighted composite scores in a battery of tests would be differentially valid when measured against different criteria, and could thus be used for classification of Army recruits into maximally efficient job areas.

They have returned to the original ideas of differential validity theory, and have developed their own brand; differential assignment theory, (DAT), which stresses the difference between predicted performance measures across jobs, assuming that there are several factors that differentially predict performance in various jobs. The theory assumes a non-trivial multidimensionality of aptitudes within the joint predictor-criterion space which represent the multiple predictors of performance required by classification. These authors clearly distinguished predictive validity as the appropriate figure of merit for judging predictor value when predictors are to be selected for a battery to be used purely for selection, and differential validity, or the degree to which the battery can predict differences between criterion scores, as more appropriate for a classification battery.

Differential validity was determined by these authors by the use of multiple regression techniques. Full least squares composites, containing all the tests in the battery, were used as selection and assignment variables, and the criterion employed was mean predicted performance ratings, computed using a simulation procedure and then trialled on a back sample. In a series of studies they examined the utility of maximising differential validity, which they defined as the differences within and between the predicted performance scores of individuals for different jobs, and which they measured by means of full least squares estimates of performance for each job as composites of predicted performance.

Brogden (1955) showed that, when these were derived from the multiple regression equation of all predictors in a battery, the sum of the least squares estimates of performance was equal to the sum of the actual criterion scores. Thus, mean predicted performance is shown to be an increasing function of $(1-r)^{\frac{1}{2}}$ where r always represents the intercorrelations between the full least squares estimates of performance. Therefore, mean intercorrelations among job performance estimates can reach fairly high levels without much loss of classification efficiency.

This challenged the traditional view that test batteries should be so constructed that, at least ideally, each component test correlates positively and equally with the criterion; the "ideal measure" of the quality under discussion. Provided that each test intercorrelates with none of the others, such a battery of ten tests can give virtually perfect prediction even when $r < 0.3$.

Because the traditional battery assumes that each aspect of ability is separate from each of the others and bears no relationship to any of them, scores on each test can be simply added together and a summative measure of ability produced.

The basic assumption of such a battery is that each test correlates positively with the aspect of ability which it measures, and has no significant relationship, positive or negative, with any other aspect of ability.

However, it does seem that, in constructing a battery of tests to measure various aspects of ability, it is necessary to accept the fact of intercorrelations between tests. In practice, it is very unusual for such intercorrelations to fall below the level of the criterion correlations. This means that a battery of tests constructed along traditional lines would need to contain an enormous number of tests if it were to approach maximum predictive value.

The inevitable intercorrelations would greatly lower the efficiency of the traditional battery, because it could not be assumed that each test was measuring a particular ability and nothing else. Where there are such intercorrelations, even an infinite number of tests in a traditional battery would not allow perfect prediction to take place, but, in a differentially constructed battery, this need no longer be regarded as an insuperable difficulty.

This was a concept turned by Morrisby into a purer form of differential theory. He was to argue that negative correlations between tests could exist and be of predictive value, as the differences between certain tests would themselves correlate with particular vocational criteria.

In practice, we are all aware of high negative correlations between predictors and criteria - stereotypes such as the absent-minded professor suggest how often our implied theories of intelligence assume that practical competence and scientific or artistic genius have a negative relationship.

While not wishing to argue that incompetence in one area always predicts competence in another, it is possible that large differences between scores predict particular strengths in the higher score, and it is highly likely that an individual's "highest" score, whatever its absolute difference from the lower scores, is of greater predictive value by virtue of being relatively higher than those lower scores.

John Morrisby was to develop these ideas from the 1950s research on differentiating criteria, particularly from the work of Paul Horst and, indirectly, from that of William Mollenkopf.

William Mollenkopf devised equations for showing how to predict differences in performance on different criteria. He showed that the multiple regression weight to be applied to each test score to predict the criterion/difference score would be equal to the difference of the weights for predicting each criterion separately, and that the difference between the predicted scores in the two criteria is equal to the predicted differences, each test being assigned the appropriate multiple regression weight.

This led to the conclusion that the relationship between the multiple correlation between predicted and actual differences could be related to the two multiple correlations with the pairs of criteria in the following way: the square of the multiple correlation between predicted and actual criterion/difference scores would equal the sum of squares of the multiple correlations of the battery with each criterion, less the product of these correlations and the correlations between predicted scores, all divided by $2(1- r_{ab})$ where r_{ab} are the criterion inter-correlations.

The Mollenkopf formula for the validity of the prediction of a difference is as follows:

$$R_{d*d} = \frac{\sqrt{R^2_{a*a} R^2_{b*b} - 2R_{a*a} R_{b*b} r_{a*b}}}{\sqrt{2(1-rab)}}$$

R_{d*d}	=	validity of differential prediction (correlation between predicted and observed differences)
R_{a*a}	=	validity of battery for criterion a
R_{b*b}	=	validity of battery for criterion b
r_{a*b*}	=	correlation between predicted criterion scores
r_{ab}	=	correlation between observed criterion measures

(Mollenkopf 1950)

It is apparent from this equation that it is essential, for differential validity to be of use, that the correlations between predictions (r_{a*b*}) should be as low as possible. If two tests predict success on a measure at identical levels, so that scores for any one person would be the same on both, then r_{a*b*} would be 1.00, and, according to the Mollenkopf formula, R_{d*d} would be zero. Thus it must be an essential requirement for each test to have different validities for the different criteria, and this differential validity, despite Brogden's comments, is more likely to occur if the tests are not highly inter-correlated.

Paul Horst (1954, 1955) distinguished between two types of multiple prediction; multiple differential prediction and multiple absolute prediction. In both cases, he was referring to the use of a differential battery, in the sense that the end result would be a set of predictions of success in each of the criterion activities, based on equations employing least square regressions weightings for the different predictors.

By multiple absolute prediction, he meant the use of a battery of tests which would have the highest prediction efficiency for all the criterion variables, irrespective of how well it differentiated between them. The index of efficiency would therefore be the sum of the variances of the predicted criteria, irrespective of their covariances.

By multiple differential prediction, he meant the use of a battery which would predict differences between all possible pairs of criterion measures. The index of efficiency would be the average of the variances for the predicted difference scores for all possible pairs of criterion variables.

In the case of multiple absolute prediction, he was assuming a selection situation in which it would be necessary to assess candidates against a set of criteria which were identical for each candidate. In the case of multiple differential prediction, he assumed a situation where several candidates were being considered for several different kinds of work. (Horst 1954; Horst 1955).

Horst argued that the index of the differential predictive efficiency of a battery is a simple function of the average of the variances for the predicted difference scores for all possible pairs of criterion variables. The larger the average variance, the greater the differential prediction of the battery.

This he showed to be equivalent to the difference between the average variance of the predicted criterion measures, and the average of their covariances, assuming standard measures for both predictors and criteria, provided that the predicted criteria are "least squares" estimates. Assuming average covariances of predicted criterion measures were equal, the greatest differential prediction would come from a battery of tests with the highest average squared multiple correlation with all the criteria.

John Morrisby (1955) attributed much of the theory underlying his Differential Test Battery to the work of Horst, but in fact he took an alternative view of differential testing. Using the Horst and Mollenkopf equations, he developed a formula of his own to represent the "differential coefficient" of a pair of tests with a single criterion.

Using the same method, he derived equations to calculate summative, differential and multiple coefficients; that is, correlations of tests with a criterion; correlations of the difference between all possible pairs of tests with the criterion, and correlations of both types of coefficient taken together with the criterion.

The average of the differential coefficients was to be taken as the differential validity coefficient of the whole battery. Although he did not claim that his differential coefficient was likely to be greater than summative coefficients in most cases, he did argue that in some cases that might well occur, and he also claimed that greater predictive validity would result from using more than one type of coefficient.

This, he argued, was particularly likely if a battery were constructed along differential lines, with the intention of maximising differences rather than following the summative approach of constructing a battery of subtests derived through factor analysis; a process which depends on test intercorrelations, frequently arrived at adventitiously or for reasons which may have little to do with any true causal relationship between tests and criterion.

Apart from the Washington school of research, differential theory, using the phrase in the sense that scores on subtests can predict different levels of performance on more than one criterion measure, has not provoked much interest. Despite their strong arguments in favour of a classificatory approach, very little recent work has been done in examining the related concept that differences between test scores within a battery, rather than composites, weighted or not, might also be useful as classificatory variables when measured against several criteria, or that they might be useful in predicting performance on a single criterion.

This has attracted little attention, despite its implications for selection, development and vocational guidance.

A natural development of the differential approach is the possibility that profile similarities, irrespective of elevations and mean scores, play a part in performance. This is not to argue that people will necessarily confound prediction by always performing in the workplace at the same levels when their test scores differ widely in level but their patterns of scores are identical. Nor is it to deny that possibility. Achievement in tests and levels of actual performance are often poorly correlated because of test-specific difficulties; individuals may, for example, perform poorly in test situations because they are suspicious and nervous of speeded psychological tests, but in reality be highly competent in their field of work. This is most often the case when test candidates come from minority or disadvantaged groups, or groups which perceive themselves or are perceived as outside the target area for the test.

In such circumstances, it may be of far more value to examine the shape of the profile than the actual levels of scores, and to devise a means for assessing both the relationship of the profile to the criterion, and, if necessary, the relevance of mean scores and distances from higher-achieving, parallel profiles. This presupposes the need both for an index of profile similarity and a means of assessing the validity of a profiling method in terms of that index of similarity.

John Morrisby, using some small samples of school students, examined Morrisby scores in relation to French O level grades, and found in some groups of students a significant association between score differences (profile shape) and examination grade, and on other groups a significant association between score levels and grades. Morrisby suggested that the difference lay in methods of teaching and school culture.

The literature on profile analysis is somewhat sparse. In 1936, Zubin published a detailed analysis of patterns in questionnaire responses, and suggested the need to formulate a technique for determining their similarity. He devised a simple expression for the agreement of patterns, assuming only dichotomous scores, based on the number of dimensions for which scores are the same divided by the total number of items. This expression did not take into account expected frequency of agreements, but opened up possibilities for debate.

In the 1930s, Cyril Burt pioneered his Q-technique for this purpose. (Burt 1937). Burt's Q-technique was developed with the concept of "the idea of degree of similarity of personality pattern" in mind. In Q analysis, correlations between people rather than variables are factored, so that the resulting factors load on individuals who are then classified into groups. This means that the general factor that would normally emerge in R-technique (almost certainly one of actual size) will be missing, and so perfect correlations might be found between individuals of similar profile but very different score levels. Stephenson followed a similar procedure, on data gleaned from candidates' introspective or "solipsistic" responses, rather than behavioural data, but used the rank correlation coefficient instead of the product-moment formula, as pioneered by Burt.(Burt 1937).

Following this work, Cattell devised and described two possible indices; the "shape correlation coefficient", or index of pattern similarity, ignoring size, and the "coefficient of pattern similarity", which takes both shape and size into account. The shape coefficient, rs, he derives as: $rs = \frac{\sum x' y'}{n \sigma_{x'} \sigma_{y'}}$

in which n is the number of variables and x' and y' are the corresponding values of the two patterns in each variable. They are shown as primes because they have been first standardised with respect to the distribution within each separate category of the profile.

Cattell doubted the efficacy of a measure that ignores the effect of size of scores, so preferred the pattern similarity coefficient, r_P, which he derives as:

$$r_p = \frac{2k' - \Sigma d_2}{2k' + \Sigma d_2}$$

where k' is the median chi-square value on k degrees of freedom; i.e. the number of elements in the profile, and Σd^2 is the sum of the squared differences, in standard scores.

This also took into account expected frequency of agreement, which had been lacking in Zubin's work, but is open to the objection that it assumes orthogonality of factors, making it difficult to use with ability tests - indeed, he said himself that "It would be quite wrong to use it with any highly correlated variables" (Cattell 1969). This formula resulted in an index ranging from -1 to +1, similar to that of the correlation coefficient, which Cattell regarded as easily interpretable and generally advantageous. He warned that the technique had limits, and could only be used to deal with similarities within a single species, for mixed species would give multimodal distributions. "The coefficients will not indicate whether a cabbage is more like a king than a looking-glass".

However, he did devise a method for using r_p to deal with the situation when the two patterns are taken from different distributions, although of course with the same dimensions, by expressing Σd^2 in relation to different expected values.

Cattell later elaborated on the use of r_p within vocational guidance (Cattell 1969), and recommended its use within multiple discriminant function (Cattell, Coulter and

Tatsuoka 1966).

Methods of expressing pattern similarity by means of distances rather than correlation coefficients were suggested by Osgood and Suci (1952) and by Cronbach and Gleser (1946). This avoided the problem inherent in Q-technique and the Chi squared techniques of Burt and Cattell in that it assumed absolute comparability of the different scale units, and reified the size of the scale index into an actual concept of similarity that had no real existence.

In 1953, Cronbach & Gleser elaborated this point, warning that the attempt to devise a profile similarity index was based on a fallacy, since the reduction of the multiple information available to an absolute, single figure rendered that index meaningless. They also described Cattell's belief that similarity should be measured by an index comparable to a correlation, ranging from -1 to +1 as "neither necessary nor desirable". "We see no reason why the measure of separation should have a limit. Complete dissimilarity of persons is an undefinable concept."

Cronbach and Gleser described three concepts, elevation, shape and scatter. Elevation referred to the mean of all scores for an individual; scatter was the square root of the sum of squares of the individual's deviation scores about her own mean; that is, the standard deviation within the profile, multiplied by $\sqrt{k}$, where k is the number of scores. Shape is the residual information in the score set after equating profiles for both elevation and scatter.

Once the elevation has been established, by averaging the scores, it can be removed by finding the difference of each score from the mean. This was described by Thomson as "centring about persons" (Thomson 1959). Treating deviation scores only removes information about level of scores and examines profile pattern alone.

If the deviation profiles (the differences from the mean)are divided by each person's scatter, this will effectively remove scatter differences between individuals, and concentrate information on shape alone. In this case, all correlations between profiles have become essentially measures of distance.

Cronbach and Gleser point out that this method can only be relied on when scatter is large relative to the error dispersion for the individual. If profiles are flat, or nearly so, this method will be greatly influenced by random error, as differences between persons near the centre will be greatly magnified.

They recommend retaining the elevation in a new, weighting similarity measure, dw^2, which combines their original measure, D_2, with a weighting element.

$$D^2 = \Sigma(X_{ij} - X_{2j})^2$$

$$Dw^2 = D^2 - k(1-w)\delta^2 El$$

w, weight, can range from zero to 1.

D^2 = sum of squared differences on dimensions(k) between individuals

$\delta^2 El$ = difference in elevation between individuals

However, Horn (1961) describes Cronbach and Lee's various forms of D_2 as basically meaningless. Although the constructors indicated that the statistic should only be used in circumstances where the dimensions are independent in the population, and the scores be in standardised form, he argues that such restrictions have not in practice been observed, so there is no frame of reference within the reported investigations to assess the statistic. In any case, he argues that it is really only a special case of Mahalanobis' D^2. Instead, he considers the further use of Cattell's r_p, pointing out that in this case too it is often difficult to ensure independence of dimensions. He suggests that Fisher's test and Rao's test, as used in discriminant function analysis, might be utilised with r_p. Fisher's test examines the hypothesis that there are no differences in mean values on the K dimensions of Mahalanobis' D^2. Since this is a test of no difference, significance suggests real difference and r_p will be negative.

Nunnally (1962) also advocates the linear multiple discriminant function method(Tatsuoka & Tiedeman (1954). He prefers this method for discriminating the typical profiles of two or more groups, as this both provides the best (in a least squares sense) linear combination(s) for discriminating the groups, and offers a procedure for assigning individuals to one of the groups. He also advocates the use of factorisation of a Gramian matrix of raw score cross-products rather than of correlation coefficients to obtain clusters for profile analysis. (A Gramian matrix is by definition one whose elements consist of cross products). However, he believes that it is usually necessary to assess level, shape and dispersion by separate methods, as the attempt to devise a single index makes it difficult to establish the precise contribution of each of these important factors.

Tabachnik and Fidell take a different approach to profile analysis, advocating the use of multivariate analysis of variance, using modern computer programs which allow groups to be classified along more than one dimension. The technique allows adjacent difference scores, or segments, to be examined, and for a canonical correlation coefficient to be derived which can serve as an index of differential efficiency. The technique also permits for tests of parallelism, using Wilks' Lambda. One of the programs they evaluate, SPSS MANOVA, has been used in this study to evaluate profile similarities or "parallelism".

Wilks' lambda tests the hypothesis of parallelism when segments, or score differences, are used, evaluating the ratio of the determinant of the within groups cross-products matrix to the determinant of the sum of the within and between groups cross products matrices, recalling Nunnally's suggestion regarding the factorisation of the cross product matrix. If a sufficiently small level of significance is reached, the hypothesis of parallelism must be rejected. Hotelling's TRACE and Roy's greatest characteristic root can also be used as tests for parallelism, and can be found in the tables in Chapter 5 in the MANOVA studies. (Tabachnik & Fidell 1989).

Most of the debate surrounding profile analysis hinged on its efficacy rather than its methods. Methods that include elevation, or actual score differences, have always been assumed to predict performance more effectively that methods omitting score levels. "Measurement specialists have often expressed concern..over the fact that a moron and a genius may have closely correlated profiles, despite the absolute discrepancy between their scores" (Osgood & Suci, op. cit.) "Can we put a Mongolian imbecile, on the one hand, and a brilliant schizophrenic poet like Hölderlin, on the other, in any sense into the same type because both have a verbal ability one and a half standard deviations higher than their numerical ability?"(Cattell. R.B. op. cit.)

In fact, these and most examples offered in the debate obscure the issue by spuriously associating individuals from inappropriately different populations. More to the point would be to ask whether two individuals facing the same prospects of employment or promotion, whose verbal ability was in each case one and a half standard deviations above their numerical ability, but whose actual scores differed greatly, possibly as a result of different levels of test sophistication, could be "put..in any sense into the same type." Examination of occupational groups suggests that possibly they could. It will be seen that, in the data sets of apparently high achieving groups examined in

Chapter 5, very wide ranges of scores indeed may be found within the same occupational groups, and parallel profiles separated by wide ranges of ability are not necessarily distinguished by achievement.

However, in order to use the differential approach, a suitable test battery has to be found. Although Sternberg has been instrumental in recent years in advocating and testing first the componential and then the triarchic multidimensional views of intelligence, his experiments have all been with tests created by himself and his assistants for the purposes of those experiments, and he has presented very little external validation data for the tests used. Of all the test batteries available currently, only the Morrisby Differential Test Battery (now Morrisby Profile) seems to have been constructed along explicitly multidimensional lines, and for this reason it has been selected for further examination, as a testing ground for the differential theories on which it is based.

Chapter 3. Theory of the Morrisby Profile

John Morrisby's Differential Test Battery

When John Morrisby (1916-1976) constructed the Morrisby Differential Test Battery, which was a process begun in 1942 and completed with its publication in 1955, he did not feel the need explicitly to set out a theory of abilities to underpin his tests. He was not alone, in the 1940s and 1950s, in taking an empirical, atheoretical line in constructing tests to predict performance. As has already been indicated, most of the work in the 1940s and 1950s was concerned with principles and methods of test construction and validation techniques, rather than with the attempt to relate those tests to an established theory of human ability.

The prevailing position of many test constructors of the time is well illustrated by Douglas Vincent, Principal Test Constructor at N.I.I.P (National Institute of Industrial Psychology) between 1950 and 1970.

> "In your letter you speak of the "principles on which you based your tests". Are there any such principles? My own opinion is that, apart from a good knowledge of psychometrics, successful test production is based on experience, trial-and-error, and a good deal of hard work. Emphasis on trial-and-error." (Vincent 1980 pers. comm to C. Brotherton)

When John Morrisby first constructed his battery, he wrote a thesis describing his theories, particularly stressing the limitations of factor analysis, and presented it to Cyril Burt, who dismissed it out of hand. Morrisby then destroyed all his writing on the subject, except for the tests themselves, and a short monograph "The Theory of the Differential Test Battery", which is all that remains. He then agreed with NFER-Nelson that they should distribute his tests and conduct validation studies themselves concurrently, but after a difference of opinion regarding royalties he withdrew the tests from NFER-Nelson, and no record of the validation studies now exists.(pers.

comm Mary A. Morrisby 1992).

Morrisby then established his own company, Educational and Industrial Test Services Ltd., and distributed the tests himself to industrial clients attracted by his approach to selection problems. He showed no further interest in publishing his theories, in presenting data to the academic world or in joining in the debate on the nature of intellect, preferring to rely on the empirical success of his tests with lay selectors, who were content to claim they saw increases in efficiency and productivity as a result of using the battery.

What is known of his theories and sources has to be gleaned from the tests themselves, from his notes on their interpretation and from the "Theory of the DTB"(1955). In many instances in this work, the tests are explicitly related to theories Morrisby may never have held, but it is contended that the differential approach described here would be of use whether or not it was followed by the constructor of the differential tests used to illustrate it, and that Morrisby was a pioneer of differential testing theory rather than the founder of all its tenets. For this reason, although a full description is given in this section of what is known of Morrisby's views and the interpretation he advocated for the tests, this is done in order to set the tests themselves in context, rather than to argue ad hominem for the truth of the views expressed in the rest of the work.

What Morrisby was more concerned to establish, in describing the principles of test construction along which he worked, was, in the first place, the need for a causal rather than a purely factor-based link between test variables and criteria, and, secondly, the importance of creating a battery which produced results in line with his implicit theories of ability.

His perception of a "causally related" test was one which was linked functionally with a criterion, in that possession of the ability would directly affect performance on the criterion. He distinguished such causal tests from those which were linked correlationally but where the relationship was indirect or spurious.

This can be illustrated by the relationship of height and weight to the criterion task of high-jumping. Height is an attribute causally related to the criterion, as the task of high jumping requires the raising of the centre of gravity of a body over a bar. The tall person has a higher centre of gravity when upright than a shorter person, so in jumping over a bar will have a shorter distance to lift that gravitational centre. Weight is also causally related, though negatively, to the criterion, in that a large mass is harder to accelerate through the given distance than a smaller mass.

Height and weight are functionally independent of one another in relation to the criterion task. Height operates by virtue of the geometrical position of the centre of gravity of the body, whereas weight operates in terms of the acceleration of a mass along a vector. Nevertheless, height and weight are also positively related to one another; tall people tend to be heavier; shorter people tend to be lighter. Thus the positive relationship between the variables, which is irrelevant to success on the criterion task, will affect the real negative relationship between weight and the criterion task by reducing it. Similarly, it will distort the already positive relationship between height and the criterion and apparently increase it. The true causal relationship between the variables and the criterion task will therefore be masked by an irrelevant concomitant relationship between the variables.

The only authority to whom Morrisby refers in the Theory of the Differential Battery is Paul Horst, whose equation for dealing with suppressor variables was developed by Morrisby to link the predictive efficiency of a battery with the parameters of the component tests. (Horst 1941). Horst's work was published in 1941, the year before Morrisby began his work on the DTB, and eight years before the Mollenkopf development of the equations for differential purposes in 1950, although both these papers were published well before Horst's discussion of his differential index of validity in 1954.

Much of Morrisby's discussion of the theory of his battery is concerned with establishing his distrust of the current over-reliance on correlational and factor analytic techniques. He described three major flaws to which these techniques were subject: fallacious validity, suppressor variables and the confusion of functions and factors.

Fallacious validity

Fallacious validity can be said to occur when the correlation between a variable(e.g. height of children) and the criterion(success in a test) appear to be high, but this actually results from the true correlations between both the criterion and height with another variable, age. Thus tall children appear to do better on tests than shorter children, but what is actually the case is that older children, who do perform better than younger children, also tend to be taller.

The common problem of spurious correlations leading to fallacious validity claims is described by Morrisby with reference to an actual occurrence in the early days of the DTB.

"Observations have been made which indicate a high correlation coefficient between the scores obtained on the Differential Test Battery and the results of a certain first-year University exam. On the face of it this suggested that the Differential Test Battery would form a very sound basis for the selection of students for the course.

It became clear that a quite large proportion of the relationship between the Battery and the exam was accountable in terms of a pair of manual dexterity tests which were operating together in a way which had long been recognised as associated with handwriting skill.

These two aspects of the Differential Test Battery appeared to account for the large part of the relationship that existed between it and the criterion exam. In short, the "validity" of the DTB in this instance derived from the fact that it had picked up the artefacts of the exam. To use the tests blindly for this exam... would mean that students would be chosen mainly for their handwriting."

Suppressor variables

Morrisby also argued strongly that traditional correlational techniques were likely to ignore problems related to suppressor variables, or variables unrelated to the criterion but highly relevant to the candidate's ability to take the test in the particular form in which it is presented.

Suppressor variables were particularly likely to occur in test batteries derived from matrices of intercorrelations, and, as Wherry argued in the 1940s, it was practically impossible to eliminate them through computational means.

> "The most laborious solution is to compute all possible combinations of two, three or more tests up to the limit of testing time available. This approach would ensure the best solution, but is considered excessively time consuming if the number of possible predictions is at all large".(Wherry 1946)

In 1946 it would indeed have been an excessively laborious process, but the current availability and processing speed of statistical computing packages suggests the preference of Wherry and his colleagues for a non-computational solution is of historical interest only.

John Morrisby took an alternative approach to the elimination of suppressor variables. In the first place, he hoped that a battery which was not based on factors derived from intercorrelations would avoid the deliberate creation of such variables. Where they did arise, he employed a partitioning out technique which he described in his "Theory of the DTB":

> "(the Shapes test) measures a function which happens to emerge comparatively late in normal development. Those in whom the particular function is not developed are unable to perform the test and fall back on a alternative quasi-logical method of solution of the problems.
>
> It is necessary in the testing situation to enable those without the requisite "true" ability to continue with their alternative "false" solutions to the problems and a facility to enable them to do this has been built into the test itself. However this causes some interference with the performance of persons who do have the requisite "true" ability.
>
> A special "wrongs" key =- strictly speaking, a measure of the interfering suppressor variable - has been devised which frees the test subject's score from that part which is due to a "false" solution to the problems." (Morrisby 1955)

Factors and functions

Morrisby distinguished factors and functions in terms of their causal links with criterion tasks. "..primary variables or functions should be causally related to the criterion...Those variables which have basic or primary relevance to criteria are labelled <u>functions</u> in an extension of the sense that a criterion can be regarded as a function of one or more of these primary variables....at the present time, a function can only be defined in terms of its measurement...functions can only be recognised from the way they operate in relation to criteria." (Morrisby 1955)

Morrisby perceived functions as having direct meaning and a direct, causal relationship with the criterion. He recognised the difficulty in defining and quantifying these entities. Factors, however, were artefacts with little clear meaning, and such meaning as they did have might have no direct relationship with the criterion. Functions appeared to be on the side of the angels but similarly difficult to pin down. Factors, on the other hand, were easily identifiable and quantifiable, but the nature of their causal link with criteria could be questioned.

Morrisby was not alone in the 1950s in sounding a note of caution about the tendency to overstretch the tool of factor analysis. Guilford (1952) listed several common faults in the use of factor analysis, some of which were particularly stressed by Morrisby. In particular, Morrisby would have agreed with him that too many factors are often extracted for the number of experimental variables - Guilford recommended at least three variables for each factor - and too many experimental variables are factorially complex. Sometimes a common factor will fail to come out because it is substantially represented only in one experimental variable. The populations on which analyses are based are often heterogeneous, and, a particularly important point, correlation coefficients used in analysis are often spurious.

As Thurstone wrote, correlations used in factor analysis should be between variables that are linearly independent, so that no reason for covariation exists except that due to common factors. If a set of scores for factors were properly slanted so that each score measured only one factor, then the intercorrelations of those scores would indeed measure the intercorrelations of the factors, and factor analysis of the intercorrelations would give the second-order factors. However, if intercorrelations of factor scores are distorted by factorial complexity, analyses are difficult to interpret accurately.(Thurstone 1947).

Guilford also pointed to the number of times when specific and error variances actually contribute to intercorrelations, as when items are scored with weights for more than one trait variable. He suggests the possibility that two factors might be orthogonal or independent, but that the scoring keys for the two scores which were designed to measure them might contain a number of similarly weighted items. This would cause a positive correlation to appear between these unrelated variables. A negative correlation could also be brought about if the weights were in the opposite direction when items were scored for both factors. Thus positive and negative correlations among scores would actually be influenced by positive and negative correlations of weights.

In addition, he wrote, "At the present time we are not in a very good position to determine the extent of (factor intercorrelations)...there are..incidental disturbing conditions that give the appearance of intercorrelation of factors in an oblique solution in factor analysis."(Guilford 1952).

Although Morrisby agreed that factor analysis could be a useful tool for a test constructor, he was concerned by its misuse as an indication of test construct validity. He believed that it might not always isolate variables causally related to criteria, because of spurious correlations and interfering variables, and feared instead that factors artificially created might be regarded as reified by the technique.

Furthermore, Morrisby argues that, because of their basis in the correlation matrix, factor analytic models ultimately rest on the assumption that correlations with criteria are linear. If this were not the case, the use of curvilinear and nonorthogonal factors ("Functions") would give rise to a model of abilities which depended on patterning rather than scalarity.

Although it is not possible to support this approach incontrovertibly, there are some indications that it mirrors reality. If it is assumed that performance on a criterion is dependent upon the vector sum of the functions and that these functions are not all orthogonal, it follows that certain functions could have an inhibitory effect on performance on a specified criterion. However, generalisability and the inevitable appearance of intercorrelation (a super-dimension) will always mask the inhibiting effect.

Only if it were possible to partition out the communality and leave the vectorial functions for analysis, would it be possible to indicate whether abilities are truly cumulative or vectorially non-orthogonal.

> "The end result of a factor analysis is likely to have less real meaning than the original data upon which it is based; second, factor analysis is unlikely to reveal but is very likely to obscure any causal relationships that may exist between a test variable and a criterion. (This is only to be expected when it is remembered that the raw data upon which it is based, or at least a large part of it, are necessarily irrelevant to any criterion).
>
> Thirdly, the method of factor analysis leaves itself wide open to the development of false inferences through the effect of fallacious validities. If the variables being factorised are restricted to those which are causally related to the criterion and those with true validity, then rarely will there be enough data to enable the technical requirements of factor analysis to be fulfilled.
>
> If however enough variables are included to enable a factor analysis to be undertaken, then it is almost certain that many of these variables will have only an adventitious relationship with the criterion, or else will show fallacious validity. In such a situation important causal relationships are likely to be swamped in a mess of irrelevancies". (Morrisby 1955)

It is not known how Morrisby constructed the items of the original Differential Test Battery, but it is safe to assume that factor analysis was not the method employed. Although the battery, when factored with marker variables, does indeed come up with logical groupings (a "verbal" factor loads highly on verbal ability and language examinations, for instance; a numerical factor loads on Numerical ability, Maths GCSE and so on), Morrisby would undoubtedly argue that this, though convenient, does not reify the constructs of verbal and numerical ability; it only shows that correlations, which may be spurious, exist between these tests and examinations with similar labels.

Differential Validity

The larger part of Morrisby's discussion of his own work lays much more stress on methods of establishing differential validity coefficients, and here he did rely more heavily on correlational techniques. It is ironical that someone who wrote so incisively about the danger of spurious correlational validity should have developed a technique for establishing differential validity that depended heavily on those same test and criterion correlations, although it is true that his method sets out deliberately to counteract the problem of the suppressor variable.

Indeed, Morrisby based his equation for linking the predictive efficiency of a battery with the parameters of the component tests on a Horst equation developed specifically in relation to suppressor variables. His coefficient of multiple correlation is a function of two terms of the formula. The first is a function of the average magnitude of the criterion correlations, and the second is a function of their variance.

$$R^2 = \frac{n^2 \overline{X}_{r^o}}{1+c(n-1)} + \frac{n\sigma^2_{r^o}}{1-c}$$

where:

- n — is the number of component tests of the battery
- $\sigma_r o$ — is the standard deviation of the criterion correlation
- $\overline{X}_{r^o}$ — is the arithmetic mean of the criterion correlations
- c — is the value of the test intercorrelations, assuming these to be all equal.

If the correlation coefficient of the test with the criterion is as high as possible, no sub-test is inferior to any other and therefore the variance of the criterion correlations will approach zero. In these circumstances, the predictive efficiency of the battery, and therefore the multiple correlation, will increase as the number of tests and the size of the criterion correlations increase, and as the test intercorrelations decrease.

Once the test intercorrelations reach zero, the formula will become:

$$R^2 = \frac{nr_o^2}{1} = nr_o^2$$

This means that, with criterion correlations around a reasonable figure of .4, only six tests would be required for perfect prediction ($r_o = \frac{1}{\sqrt{n}}$)

However, the ideal case of zero intercorrelations is most unlikely ever to arise, for the reasons mentioned previously. In practice, intercorrelations are likely to equal or exceed criterion correlations. In the case of intercorrelations of .2 (a very conservative figure), criterion correlations of .4 would require an infinite number of tests to offer perfect prediction, following the Horst formula.

Morrisby then turned his attention to the construction of a battery which would allow notional "perfect prediction" on less stringent terms. He based this non-traditional, "differential" battery on the assumption that a criterion variable is causally determined by the interaction of correlated test variables which are still functionally independent.

The criterion correlations might be positive, negative or zero, provided they contributed causally to the criterion. They should ideally be arranged in pairs so that a test correlating positively with a criterion could be matched with another test correlating negatively to the same degree. This would mean that the arithmetic mean of the criterion correlations would be zero, although the variance would be positive.

Thus the first term of the Horst equation would vanish, and, assuming a mean of zero, the formula would be:

$$R^2 = \frac{n\sigma^2{}_r{}^o}{(1-c)}$$

This time the multiple correlation would increase with the number of tests, the size of the test intercorrelations and the variance of the distribution of criterion correlations.

In this case, assuming test intercorrelations of 0.5, and criterion correlations of .29, the number of tests required in the battery for perfect prediction would only be 6, in contrast to the infinite number required with similar or less likely conditions in the case of the traditionally constructed battery.

The may be compared with the Mollenkopf formula described in the preceding section.

Morrisby took the mathematics of multiple regression and added his own, vectorial, interpretation to derive a differential coefficient.

The estimated score on the criterion is usually derived from the sum of the standard score of the tests, each weighted by the appropriate beta coefficient. Morrisby pointed out that, if the second test assesses a different aspect of the criterion task from the first, which is one of the tenets of differential theory, the tests are not measuring in the same direction, and so must be regarded as summing only vectorially, not in a scalar fashion. Thus the correlation between the criterion and the sum (or average) of two tests would be derived from the square root of the sum of the two beta coefficients multiplied by their criterion correlations. This may be expressed as:

$$r_{o(1+2)} = \frac{r_{o1}\sigma 1 + r\sigma 2}{\sqrt{(\sigma^2{}_1 + \sigma^2{}_2 + 2R_{12}\sigma_1\sigma_2)}}$$

The correlation between the difference of the two tests may therefore be expressed as

$$r_{o(1+2)} = \frac{r_{o1}\sigma 1 - r\sigma 2}{\sqrt{(\sigma^2{}_1 + \sigma^2{}_2 - 2R_{12}\sigma_1\sigma_2)}}$$

If the tests distributions had unit variance, or were calibrated by some standardised scaling system so that they had the same variances, then the standardisation element could be removed from the above equations and they would become:

$$\text{Summation Coefficient} = \frac{r_{o1} + r_{o1}}{\sqrt{2(1+r_{12})}}$$

$$\text{Differential Coefficient} = \frac{r_{o1} - r_{o1}}{\sqrt{2(1-r_{12})}}$$

where:

r_{o1} = the correlation of Test 1 with the criterion
r_{o2} = the correlation of Test 2 with the criterion
r_{12} = the intercorrelation of the tests

Since these two coefficients are orthogonal, they should, between them, account for all of the variance that the two tests have in common with the criterion, and so their vector sum would be equal to the multiple correlation coefficient with the criterion.

$$\text{Multiple Coefficient} = R_{o.12} = \sqrt{(sum.coef.)^2 + (dif.coef.)^2}$$

The differential coefficient for the battery might then be expressed as the average of the differential coefficients for the paired tests of the battery, and the overall coefficient for the battery would be the average of the multiple coefficients. The phrase "differential coefficients", used in this sense, refers to the relationship between the criterion and the differences between paired subtests. "Summative coefficients" refers to the relationship between the criterion and the two combined test results, whereas "Multiple coefficient" is used to describe the relationship between the criterion and the combined summative and differential coefficients.

Morrisby never explicitly presented data showing the differential validity coefficient(s) for the whole of his test battery, although he did present sets of experimental data showing the performance of groups taking the battery whose scores on the various subtests and pairs of sub-tests could be examined in relation to the criterion of O level French pass grades. He relied on what he described as "clinical observation of real cases of real life criteria" to provide the basis "upon which forecasts of criterion performance of an astonishing accuracy can be made". (Morrisby 1955).

With relatively few subjects (N = 30) Morrisby had managed to obtain some differential coefficients of around 0.4 and 0.5., which would have reached levels of significance (p = .05) It is fair to add that all Morrisby's experiments were performed with relatively low numbers, which tends to inflate correlation coefficients. His group of students of French O level comprised 30 people, whereas there were over 600 in a replication of the study which will be described in chapter 5, which did not manage to obtain such high coefficients, although those it did obtain were significant at the .05 and .01 levels.

However, the essence of the differential position is illustrated in both the original study and the replication: firstly, it is possible to improve, if only slightly, on the summative position by taking the differential position into account, and, secondly, by so doing account is taken of patterns of achievement rather than pure levels. In other words, because it depends on averaging and summing scores, the summative approach advantages those who are generally good at tests, rather than being particularly able in certain areas. The differential approach described by Morrisby favours those with more strongly marked patterns of ability. Taking both into account allows selectors and employers to benefit from the abilities of both types of applicant, and avoids the prejudicial effect of over-valuing one type of candidate at the expense of the other.

The structure of human abilities

Although Morrisby, as has been said, devotes more time to discussion of his development of the differential coefficient than to the differential theories on which it is based, and does not refer explicitly to any named theory of abilities, he does make some oblique references to concepts of human ability structure which allow one to infer his probable views.

Although he did not deny the existence of "g", he regarded it as in some sense occupying a different dimension from other abilities, and related it more precisely to overall score levels, recalling Spearman's description of his single factor as "mental energy".

> "It is useful to regard the "level" of test scores as indicating some generalised kind of "ability", unassociated with any particular direction in which this ability is manifest: in short, a "scalar" quantity of ability. When the tests are of such a kind that the scores derive from certain mental processes which are usually labelled "intelligence", then we can call this scalar quality of ability the "intelligence" of the person; or, we may prefer the term "intellectual ability". In this case, the concept of "intelligence" is identical with that derived from the usual kind of intelligence test based upon the principles of summation.
>
> On the other hand it is useful to regard the "differential" between test scores as a specialised kind of "talent"... - an ability which is closely associated with a particular direction: that is, a "vector" quantity of ability. Apart from the Differential Test Battery, there seems not to have been any true recognition of the principle that abilities can have both scalar and vector components, and hence there can be no traditional,

technical equivalent of the conceptual meaning of "differential". (Some methods of factor analysis are based on the proposition that test scores are essentially vector quantities but the final referral back to the individual case is based upon the notion of the total ability in the person as a scalar quantity).

There is, however, popular recognition that "intelligence" and "talent" are not the same thing. It is often heard, for instance, that some person may not be very intelligent in most things but has "a flair for languages", or that another person may be "very intelligent" but hopeless with mechanical things". Such observations are usually ascribed to "imponderables", such as interest, motivation, or, as an eminent academic has tried to describe the business acumen of an illiterate, millionaire business man, "native cunning".

Such observations cannot be reconciled with the proposition that there is only one general ability which determines the performance of any and every task. Nor can the total performance of a task be adequately described by postulating a number of independent abilities, all positively related to the task and combined by simple summations.

However, these popular observations may be easily and usefully explained in terms of the "level" and the "differential" of test scores, as indicated above, and hence we may use them as the basis for our concepts." Morrisby J.R.(1955).

Despite this, Morrisby did not always regard high scores as either positive or desirable. His perception of the structure and balance of abilities meant that high scores in one area were frequently balanced by low scores in another, as many pairs of abilities occupied continua or orthogonal dimensions.

He wrote, "Mental tests are usually positively scaled in that a high score is taken to indicate a high degree of some quality in the person. There is no certainty, however, that the existence of a high degree of that quality, whatever it is, in the person will be to that person's advantage in relation to some criterion. The high degree of that quality can equally be taken to indicate the existence of a low degree of the opposite quality."(Morrisby 1955)

It is clear from the foregoing that Morrisby did not regard abilities as structured either hierarchically or horizontally. The stress he laid on differential relationships and the distinction he made between scalar and vectorial values, or sums and patterns, despite his fondness for assessing the intellectual ability of the individual as one scalar quality, mark him out as ahead of his time in perceiving the multidimensional nature of human ability structure. In his interpretation of the twelve tests of his battery, he argued further for a complex patterning of interrelationships within and between certain groups of tests, which in many ways anticipated recent models.

The Morrisby tests

The Morrisby tests cover five broad areas: abstract reasoning, specific information processing abilities, including verbal, numerical and perceptual, or diagrammatic, ability; practical abilities, including spatial and mechanical ability; manual dexterity, including manual speed and manual skill, and certain aspects of personality. These may be laid out schematically as follows:

Table 3.1

COLUMN 1 Test Dimension	COL 2 Label	COLUMN 3 Score	COLUMN 4 Test Groups
REASONING	CST	□□□□□□□□□□□□	PURE INTELLIGENCE
VERBAL	GATV	□□□□□□□	INFO.PROCESSING
NUMERICAL	GATN	□□□□□□□	"
PERCEPTUAL	GATP	□□□□□□□	"
SPATIAL	SHAPES	□□□□□□□□□	PRACTICAL ABILITY
MECHANICAL	MAT	□□□□□□□□□	"
AWARENESS	ST1	□□□□□□□□□□	PERSONALITY
PERSEVERATION	ST2	□□□□□□□□□□	"
INNER CONFIDENCE	ST3	□□□□□□□□□□	"
OUTER CONFIDENCE	ST4	□□□□□□□□□□	"
MANUAL SPEED	ST5	□□□□	MANUAL DEXTERITY
MANUAL SKILL	ST6	□□□□	"

The twelve tests are all given non-transparent labels, much as Cattell invented personality variables to avoid stereotypic responses to test results. The dimensions tested are listed in column 1, the actual labels or names of the tests are listed in column 2, and the names of the "blocks" or groups of tests in column 4. Column 3 attempts to show the groupings or blocks of tests visually, by allotting arbitrary, identical scores to each test in a block or group, and showing the "profile" made by the scores.

Compound Series

The first test, the CST, is a reasoning measure, involving a pattern completion task. It requires no "verbal" or "numerical" intelligence, so in that sense is a measure of "pure" intelligence. It is also as "culture-fair" as a pencil and paper test administered in English and requiring the usual test taking skills can expect to be, in that no previous literacy or numeracy skills are assumed. However, it is not of course suggested that any test of this kind can actually be described as entirely culture free or culture fair.

Indeed, there is some evidence (Jensen 1980) that minority groups actually perform less well on such "culture free" fluid intelligence tests than majority groups, and it is worth speculating as to whether, for such groups, the difficulty of this type of test task is actually considerably magnified. The challenge of the novel task in this test must be greatly diminished for a group accustomed to such tests and their assumptions; for a group fresh to the whole concept of the speeded paper and pencil "IQ" type of test, the novelty element must be greatly enhanced, and might account for group differences in a more satisfactory way than simply assuming majority intellectual superiority.

What the test is designed to do is to present problems which are likely to be unfamiliar, at least initially, to the test takers, and asks them to cope with novel difficulties at speed. The test items, initially very simple, increase in their challenge to the test takers, and vary somewhat in their nature, involving increasingly more complex combinations of size, colour and shape to arrive at the next moves in the sequence.

Morrisby wrote "The CST provides a measure of "pure intelligence". The term is taken to mean that the test avoids the use of materials or processes which are largely dependent upon acquired skill or knowledge....the mental function which confers the ..ability to absorb or integrate the totality of his experiences and observations, allowing for the development of an understanding of various phenomena..is measured. ...The CST is a scalar test; *the only one in the Battery,* (researcher's italics) and is therefore interpreted by the absolute level."

It is interesting to see the similarity of this concept to Sternberg's perception of the part played by "the experience of the individual", in his triarchic model. This is explicitly divided by Sternberg into the ability to cope with novelty and the capacity quickly to automatize responses to problems which initially require attention. Sternberg wrote in 1981 that one of the best ways to measure intelligence was to use tasks outside people's ordinary experience. He suggested that "intelligence is...ability to learn and think within new conceptual systems, which can then be brought to bear upon already existing knowledge structures".

It is also interesting to see Morrisby's use of a problem solving task as a measure of "pure" ("fluid") intelligence. Again, Sternberg, writing forty years after Morrisby constructed his battery, suggested that "fluid abilities are particularly well measured by reasoning items such as..series completions...fluid ability tests tend to stress ability to deal with novelty."

The CST is designed to measure fluid intelligence, and the measure includes the requirement both to cope with a novel task and to automatise it in a speeded test environment. This of course assumes that the novelty of the task can be held constant across the test population, a requirement rather easier to fulfil when Morrisby constructed the test than now with a much more diverse population in terms of culture and experience. In this context it is of some interest that the test is apparently currently being used by the Adams Institute in Israel for the assessment of non-literate Arab women. The format is no longer pencil and paper, but real beads are provided for stringing on real strings. It is a matter for debate whether the test remains one of fluid intelligence in this state, with novelty deliberately minimised by format.

Although there is no actual evidence for lower performance on this test in the UK by minority groups (see Chapter 4 for t-tests on minority/majority standardisation samples), anecdotal evidence from administrators would suggest that many minority and ethnic group members do have some difficulty owing to the absolute unfamiliarity of the test task. In addition, the need to "close" or select some answers while excluding others from multiple choice sets can disadvantage those faced with multiple choice questions for the first time. Such candidates may persist in allowing themselves what they perceive as a better chance of being right by offering more than the stipulated number of answers to each question. Practice tests are now available for those taking the Morrisby Profile, and it is recommended that they should be supplied, with all necessary assistance, to all adults taking the tests well before the test session.

Recent evidence suggests (Kurtz 1992) that practice items on similar tests do diminish the gap between minority and majority group average scores. If this is so, it would add weight to the argument that such tests do have a "coping with novelty" component in them; that this does form a part of intelligent functioning and is differentially distributed in the population, and that similar levels of novelty should apply for the tests to be fair if used with groups from different cultures with different thresholds of novelty.

Information processing: Verbal, numerical and perceptual abilities

The three tests that follow the CST; the verbal, numerical and perceptual tests, are intended to assess the capacity to deal with sequential and analogical reasoning in the three media of words, numbers and diagrams. Morrisby perceived these three methods of information processing as analogous to one another, although he recognised the greater practical importance in the educational world of his time of the verbal and numerical processing skills. Probably now in the 1990s more recognition is given to the claims of diagrams, graphics and flowcharts to be seen as potentially flexible and complex communication tools.

Morrisby described these three tests as "general abilities" (hence GATV, GATN, GATP for "General Ability Test - Verbal" and so on). He identified their functions as follows:

> "They relate to the ability of remembering verbal, numerical and perceptual symbols associated with conceptual knowledge. (They)..measure the base parameters of this part of the intellectual structure, which confers the ability to memorise, manipulate and utilise such conceptual knowledge. Clustering occurs in this three-dimensional vector field, so that various "talent" profiles can be identified....The profile points to the general occupational or learning area in which a person will operate most naturally. The stronger the profile, the more marked the talent and the better the likely performance in that area. A talent in a different direction from a stated interest or proposed occupational area, even though weak, may inhibit performance in that area." (Morrisby 1955)

The verbal and the perceptual tests are constructed along parallel lines: each consists of two parts; the first part presents six words/diagrams, four of which are associated in some way and two of which do not match the other four. The task is to isolate the two. In the second part of the test, an analogy is presented, and six alternative words or shapes are offered from which to construct a similar analogy -

e.g. Shoe is to Foot as Car GLOVE Window is to HAND Brick Smoke.

or ↑ is to ↓ as ⮝ ✸ ♠ is to ⮟ ♥ ☎

The numerical test also consists of two parts; again sequences and analogies are employed. In the sequencing test, the test taker is asked to fill in the two "missing" numbers in a sequence; in the second test, a 3 x 3 matrix of numbers is presented with the bottom right hand number omitted, and the pattern underlying the rest of the matrix has to be comprehended so as to establish the missing number. No knowledge of mathematics is required beyond the capacity to count, add, subtract, multiply and divide at a very simple level.

In analogies as described above, subjects are asked to encode (recognise the word, number or diagram of the presenting item). They then need to infer the relation between the words, numbers or diagrams (Shoe is worn on foot; up is the opposite of down, and so on). Then they need to map the relation between the words, diagrams or numbers so that it may later be transferred - wearing upon; converse direction etc.)

Next they must apply the mapping rule they have generated to the new concept - glove worn on hand. As there are distractors present, they must compare the solution they choose with alternative solutions that suggest themselves, and justify that choice, and finally they must operate the right response by marking the paper correctly.

According to Sternberg's componential theory of information processing, this precisely describes the inductive reasoning process. "Inductive information processing can be understood in terms of seven performance components that are common across inductive reasoning tasks: encoding, inference, mapping, application, comparison, justification and response." The use of analogies has a distinguished history in psychometric testing, and is still current in the 1990s.

When John Morrisby designed the Differential Test Battery, the three GATs each had a third part, which was in fact Part One of each test. In the verbal test, it took the form of a vocabulary test, of the "same/different" variety (i.e. hot/cold = different: amity/friendship = same).

In the numerical test, addition and multiplication sums were worked out with answers, and the test taker had to say whether the given answer to each sum was right or wrong. In the perceptual test, the test taker was presented with a sample diagram on one side of the page, and eight diagrams opposite it, which had to be marked as identical to or different from the sample diagram.

These parts were omitted in the 1991 version of the tests, and the rationale for this is briefly discussed in Chapter 4. However, when Morrisby designed the tests, in the 1940s and 50s, there was very little interest in the need to keep psychometric tests free from contamination by such factors as education, class and culture. Instead, it was perceived that tests measuring level of vocabulary and capacity to perform numerical calculations quickly and easily were likely to assess crystallised abilities, and such tests still formed part of the revised version of Thurstone and Thurstone's Primary Mental Abilities Test, seven years after Morrisby published the DTB (Thurstone and Thurstone 1962).

In Morrisby's view, the level of the scores on these three tests assessed that individual's underlying intellectual power, but it was the pattern of those abilities that was of particular interest in predicting performance. He perceived the tests as assessing and distinguishing between (broadly) four preferred approaches to information processing; a verbal approach, a numerical approach, a pictorial, object or diagram-centred approach, no preference, or any combination of the three.

The direction of preference, as has been described before, would give the profile added strength if it followed the direction of choice indicated by the criterion. For example, a moderately high, "flat" profile would not necessarily be as desirable as one in which the desired strengths were as high or higher, but the non-preferred score was at a lower level.

Although Morrisby did not argue that the abilities themselves were always paired and orthogonal, he did claim that to be the case in certain situations against certain criteria. For example, in a comparison of two applicant craft apprentices, one with verbal and mechanical ability at equally high levels, and one where the mechanical ability was higher than the verbal ability, the lack of differentiation between the scores of the first applicant would not be in her favour, and the differentiation and direction of the second candidate's scores would be of advantage to her.

A rather similar point was made by Burt, when defending the identification of group factors by clusters of positive and negative projections on bipolar axes. Thurstone rejected these bipolar axes because "negative abilities cannot exist". Burt responded that classification must proceed by logical dichotomy and antithesis. Negative projections, he affirmed, do not imply a person has less than zero of a quality. They only record a relative contrast between two abstract qualities of thought. In other words, more of one quality usually goes with less of another. (Burt 1939)

Practical abilities: Shapes Test and Mechanical Ability Test

The practical section of the DTB - reminiscent of Vernon's k:m factor - consists of the Shapes (spatial ability test) and MAT (mechanical ability test). Morrisby describes these tests as measuring:

"those functions within the intellectual structure which underlie practical performance, and also general practicality in the ordinary activities of everyday living....The tests are bi-polar in that they are mutually opposed in the type of practicality measured...there is no clustering; the abilities can fall at any point...the bias between the two tests indicates a continuum from the person who has a preference for dealing with whole systems (relatively high Shapes score) to the person who has a preference for dealing with discrete components in a serialistic manner".

The Shapes test consists of groups of three shapes, each of which is identical in form but not in rotation or orientation. The candidate is asked to identify the only one of the three which is viewed as if from the reverse side; the other two may well be rotated but not reversed. The Mechanical Ability test consists of mechanical, practical problems similar to those in the Bennett Mechanical Comprehension test, in which a diagram of a practical problem is presented and multiple choice solutions are given, only one of which is correct (Which of these cars is most likely to skid going round the corner?)

Most people taking the Shapes test are confronting a spatial ability test for the first time, and perceive the type of problem as entirely novel. The MAT, on the other hand, does utilise experience, and is probably of all the tests in the battery the one most subject to the practice factor.

Taken together, the two tests are used to show, from the mean of the two standardised scores, how practical an individual is in general, and how easily he or she can cope with practical problems. Taken separately, the tests are used in the first instance to assess the candidate's possession of separate abilities of spatial and mechanical reasoning.

Taken in conjunction, as differential tests, they have another function, as a continuum ranging from strategic, holistic, long-term planning to tactical, serialist, short-term planning. In the spatial test, there is a need to take in the whole picture without being distracted by detail - a capacity in some degree negatively related to field independence, but carrying positive implications of its own - whereas, in the mechanical test, candidates need to work sequentially and serially through detailed elements of a diagram to arrive at a more immediate solution. A higher score on the spatial test is taken as indicating, in comparison with the mechanical test, whether individual candidates find it easier to "see the wood for the trees"; to take a long term view or see the whole system, whereas, if the mechanical score predominates, they are seen as preferring to take a serialist, more tactical and short term approach.

If the Shapes score is very much greater than the mechanical score, that is taken to indicate that the person is inclined not to work in a methodical manner, and to neglect some components of a situation. Similarly, if the mechanical bias is strongly marked, such an individual may lose sight of the overall objective in carefully attending to immediate problems and details. If the scores are similar, that suggests the capacity to keep both the overall aim and the constituent parts in perspective, although, as with the "flat GAT" profile, in a field strongly requiring one particular kind of approach, they may not fare quite as well as an individual with the appropriate bias, as might be expected from an individual whose scores are at the centre of a continuum.

Block Score Differentials

As well as arguing for the differential predictive power within the GAT (general abilities) and practical (Shapes/Mechanical) blocks, Morrisby argued that the relative arrangement of the three "blocks" consisting of the Compound Series, the general abilities, and the practical blocks, also had predictive power.

He suggested that the relative mean levels of these three blocks could be used to indicate the way in which a test candidate preferred to learn and solve problems. If the CST predominated, for example, its relative level when measured against the other two blocks could be seen as a measure of a person's potential for understanding and integrating their immediate experience into their total "bank" of experience.

Those with a marked strength in this area might be able to bring apparently unrelated facts and pieces of information to provide fresh insights and novel solutions to problems and situations. However, they may be somewhat "overpowered" intellectually, without specific talents through which their potential for understanding and integration can be used.

If the general abilities block predominated, the bias would be in the direction of using conceptual knowledge; working in a theoretical world of accepted concepts where knowledge and information are more important than either practical or novel solutions. The actual distribution of the abilities would indicate whether the area was more likely to be in the arts, commerce, science, or so on, but the approach would be influenced by the predominance of the general abilities over both CST and Practical scores.

If the Practical (Shapes/Mechanical) block predominated, the candidate would be likely to operate at the level of direct action, rather than at an intellectual or theoretical level. They are likely to work by identifying problems and implementing solutions, probably preferring a trial and error approach to determining the principles on which to act, and they would be unlikely to enjoy spending much time on preparatory planning and analysis. Their thinking style identifies options on which to act, rather than establishing underlying principles or speculating about the relative values of theoretical approaches.

If, however, the practical and the CST block predominate together over the general abilities, a rather different thinking style would emerge, in which the individual would use their understanding to grasp the essentials of the problem, and their practical ability to produce a relevant solution.

It is a characteristic of this profile that, although often possessed by very high performers in the industrial and commercial world, it does not indicate academic success, as the general abilities which make for the acquisition and reproduction of knowledge are relatively low. In Sternberg's terms, although without implying shared definitions of those terms, the metacomponents and performance components outweigh the knowledge-acquisition components.

Sternberg described Vernon's "practical/mechanical" abilities as equivalent to Cattell's fluid abilities; in the sense that such abilities are tested best by the demand to respond to novel problems, presumably in an appropriately practical context. This sounds very like the combination of integrative understanding and practical ability described above, and indicates again the usefulness of this approach in covering aspects of human development not normally measured by traditional test batteries.

It can be seen, in examining John Morrisby's approach to his six ability tests, that his concept of test interpretation was far from the assumption that all that is required is to identify the best predictors from a battery and arrive at a composite score. This complex use of both scalar levels and differentials to indicate not only the level of functioning, but also the direction and manner of that functioning, allowed a subtler, more flexible use of psychometrics, while retaining objectivity in scoring and interpretation. Provided that the approach could be shown to work, Morrisby was clearly ahead of his time in his perception of what a test battery might be asked to do.

Sternberg wrote: "Aspects..dealt with most inadequately by present tests are, I believe, a) adaptation to, selection of and shaping of real-world environments; b) dealing with novel kinds of tasks and situations; and c) metacomponential planning and decision making." (Sternberg 1985) Sternberg had, at this time, never heard of John Morrisby or of his tests (pers. comm. Sternberg 1993).

Speeded Tests - personality and manual dexterity

Besides the six ability tests discussed above, the Morrisby Profile consists of six further speeded tests, which are administered at the end of the battery and together take no more than thirty-five minutes, including instructions. These tests fall into two groups, four personality measures and two measures of manual dexterity. Although the personality measures follow the ability tests sequentially in the profile, the dexterity measures are more akin to traditional tests and will be discussed next.

Manual Dexterity tests

Speed test 5 - a measure of manual speed - involves placing three marks in a square as many times as possible within the space of forty-five seconds. Speed test 6, a measure of manual skill, involves drawing lines between progressively narrower parallel lines using the preferred and non-preferred hand consecutively. 45 seconds are allowed for each "hand". The arithmetic mean of the two tests shows the individual's usual motor speed.

To obtain a measure of the dexterity of the individual, ST6 is subtracted from ST5. For the purposes of interpretation, the two scores are presented as if they were two parts of a bi-polar continuum, one end of which is characterised by the excess of motor speed over skill (SP>SK) and the other end of which is characterised by the excess of motor skill over speed (SK>SP). Thus, an individual profile may be fast or slow(mean), skilful or unskilful (differential), and characterised as more speedy than skilful or the reverse (direction).

This approach is similar to that taken with the ability scores, with the exception of the use of ST5 as a speed suppressor on ST6, which is particular only to those tests. An individual may be able or less able (mean), with more or less strong biases (differential) in favour of one or more abilities (direction). This can be applied to the general abilities, to the practical abilities, and to the three "blocks"; in each case the mean, the differentials and the direction form part of the profile and influence its interpretation.

Objective personality tests

The first four speed tests are fundamentally different from the other tests of the profile, in that they give information about a person's "characteristic mode" of intellectual behaviour, as distinct from their intellectual talents and abilities. The tests are objectively scored, speeded measures of behaviour rather than personality questionnaires or projective tests, and as such are not subject to faking or to subjective modes of interpretation. Because they are presented as speeded tests, and speed of test-taking forms little part in the evaluation of personality traits, the interpretation procedure is based on differentials and direction rather than on absolute levels.

Descriptions of the separate tests follow, but interpretation depends on their inter-relations, and cannot as effectively be made on levels of individual scores.

Speed Test 1.

This test involves rapidly marking pairs of numbers, names (and, in the revised MP, shapes) to determine if they are the same or different. The intellectual component of the task, given very basic literacy levels, is minimal, and differences in performance will relate to the speed and accuracy of the individual. Indeed, the test task follows a format which is common to many such speed and accuracy clerical tests.

Speed Test 2.

In this test, the candidate is asked to carry out a simple manual task repetitively. The task is then changed but performed repetitively again. The candidate then has to alternate between the two tasks, thus breaking mental set quickly and frequently. Differences in score are due to the speed with which the candidate is able to break mental set, and the speed taken on the original simple tasks is used to calculate the ease with which this is done. The task follows a standard form of "alternation tests", similar to the signature task in Cattell's Objective/Analytic Battery, but less subject to error owing to idiosyncratic signature patterns.

Speed Test 3.

In this test, the candidate is asked to add two letters to a given pair of letters to construct a four letter word. The task is similar to word fluency tests such as those of Thurstone, but is further affected by the incomplete nature of the instructions given to complete the task. Candidates respond differently when given incomplete instructions, and it is this difference in response, coupled with word fluency which is a drive of internal, emotional origin, that indicates the candidate's internal self-confidence.

Speed Test 4

There are two parts to this test. In the first part, the subject is asked to list things which fit into a highly ambiguous category - e.g. to name metal objects with straight sides. Those who lack the ideational fluency to produce the ideas or who lack the confidence to interpret the ambivalence of the cue and "risk" a wrong answer achieve lower scores.

The second part of the test is a graphical version of the first. Candidates are given a series of meaningless lines and are asked to make them into meaningful drawings and label them. This item type is similar to those used by Cattell to show ideational fluency, which is a fluency of intellectual, rather than emotional, origin, related to outward confidence and initiative.

The four speed tests are interpreted with reference to three dimensions. The first dimension is assessed by examining the differential between Speed Test 1 and Speed Test 2. The candidate is classified as Flexible, Inflexible or Tenacious depending on whether Speed Test 1 is greater than, less than or equal to Speed Test 2. The scale is thus determined by the difference between Conceptual Speed and Perseveration.

When Conceptual Speed dominates, a flexible mind is indicated, in the sense of someone quick to pick up and react to new ideas. When Perseveration dominates, the tendency is towards inflexibility, in the sense of the preference to stick with a particular train of thought. When the two results are balanced it indicates a tenacious nature, in the sense of being reasonably quick to pick up ideas but tending to stay with a course of action until it proves necessary to change.

The second dimension is assessed by examining the differential between Speed Test 3 and Speed Test 4. The candidate is classified as Personally Confident, Outwardly Confident or of Matched Confidences according to whether Speed Test 3 is greater than, less than or equal to Speed Test 4.

This compares the internal fluency or self-assuredness of ST3 with the more public confidence of ST4. The continuum ranges from those who are inwardly sure of themselves, work to the full scope of their ability, but cannot take a front line position, and are unhappy when asked to initiate or innovate, to those who are ready to show a high level of confidence to others but are inwardly uncertain, finding it difficult to invest their abilities fully in one task, preferring to get on with the next project or to initiate one.

The type of confidence assessed by the third dimension may predominate or be subordinate in the profile. This is assessed by the third dimension, in which the larger of Speed Tests 3 and 4 (or the mean, if the confidences are matched), is compared with Speed Test 1. The candidate is classified as Non-confident, Confident or Even depending on whether Speed Test 1 is larger than, smaller than or equal to Speed Test 3/4.

A concomitant of conceptual speed is an awareness of the surroundings and the environment, so the person with higher levels of this awareness than of confidence is likely to be generally sensitive to the outside world, and less likely to show their confidence. If the converse is true, they will be more likely to show their confidence - of whichever sort - than to respond to their surroundings. If their inwardly confident nature dominates, they will be more likely to insist on getting things right rather than on taking the lead; if the outward confidence dominates, their behaviour will be characterised by fluency of ideas and public confidence. In practice there are many different arrangements possible and the continuum includes behaviour from highly reticent to assertive and over-confident.

Objective tests of personality have not been widely used in the UK, (Fox & Brotherton, 1991) but Cattell describes "T data", or data which comes from measurements derived from tests whose purpose is hidden from their subjects, as the third main source of quantitative data in the field of personality. (Cattell 1973). This type of measurement was used in his Objective-Analytic Tests (1955).

Such tests have advantages over questionnaires and projective tests, in that they are impervious to guessing, are not subject to response sets, and, in some cases, may be useful in cross-cultural situations, where questionnaires are particularly subject to cultural bias (Kline1977). In view of the lack of face validity of such measures, it is particularly important to establish that they are in fact valid tests. Kline 1993) recommends the use of factor analysis with marker variables in the field of personality, ability and motivation. It is a little difficult to imagine, however, which criteria he would regard as acceptable as marker variables for personality, since none of those available would be able to substantiate the claims made for objective testing as a superior method to questionnaire data, in its ability to over-ride faking, inaccurate self-perceptions and cultural bias. Nevertheless, some of Kline's recommendations were followed in the course of this study, and are described in chapter 4.

Although there are 27 possible modal types, there are clearly many variations possible of each, depending on the actual differentials within each category. As Morrisby wrote "The user must view these categorisations as initial starting points, using his understanding of the dimensions involved to identify the salient features of the person's profile. They should not be used as discrete compartments into which profiles are dropped and a resulting interpretation automatically provided." (Morrisby 1955).

When Sternberg first presented his triarchic theory he barely touched on the part personality might play in the structure of human intelligence. However, he modified this approach considerably in his later work, and introduced a whole alphabet of personality factors, mainly negative, that might militate against the proper exercise of a person's abilities. (Sternberg & Suben 1986). He described them as impediments which interfere with the development of intelligence, interacting with the elements of ability structure so as to maximise or minimise their effect.

Some of the aspects (or failures) of personality he describes are echoed in the dimensions of the modal profile outlined here, particularly lack of motivation (ST3) lack of perseveration (ST2); inability to translate thought into action (similar to the clinical interpretation given to the circumstances when a "block differential" occurs where the CST predominates over the other abilities); failure to initiate (ST4) and too little/too much self confidence (ST3/4).)

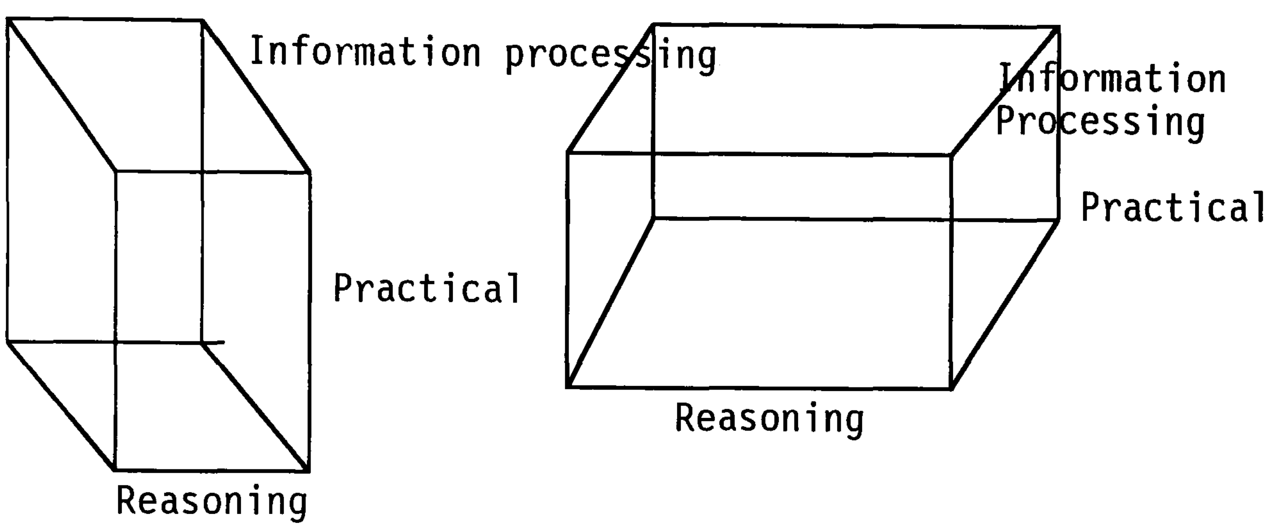
The approach to personality testing outlined here makes no attempt to offer a full personality profile of the individual; nor is it closely related to a "Big Five" model, although there is a clear link between extroversion and Speed Test 4. In empirical testing programmes, using the full profile to measure occupational or educational performance, it is clear that these tests do have some predictive validity, as may be seen from the data analysed in the following chapter, particularly that relating to success in sales. Motivation, flexibility and confidence probably do have links with cognitive achievement, as Sternberg suggested, and the unfamiliar and challenging format of psychometric tests may well appeal to some candidates because of their personal make-up more than they do to others, which is justifiable if that reflects fairly the fact that certain tasks and work styles will be more or less congenial to them in reality. On the other hand, it is equally possible that success in psychometric tests engenders a temporary, situational, confidence. However, it has not been the task of this study to erect a new theory of personality within which to set the speeded Morrisby tests; rather it is the contention of this work that they assess various behavioural facets of performance which link personality with abilities.

The Morrisby Profile depends on a conception of intelligence which owes something to Vernon and Guilford and to a degree anticipates both Carroll and Sternberg, but cannot wholly be categorised by reference to alternative theories. It presupposes a capacity to understand complex matters which is not unlike Cattell's fluid intelligence, and which is perceived as likely to be innate. This capacity may be possessed in any amount, and exists as a dimension irrespective of the quantity of other, "learned" or crystallised intelligences which Morrisby sees as the information processing abilities: verbal, numerical and perceptual or diagrammatic, somewhat akin to Vernon's v.ed, and forming a three-dimensional dimension of its own. A third, "two dimensional" dimension of ability is formed by the practical intelligences, measured by spatial and mechanical ability. There is also a fourth dimension which is occupied by personality factors which cannot be wholly divorced from intelligence or ability.

In addition, a fifth dimension of work style and planning exists, which is related to previous dimensions but exists independently of them. The planning style affected by the pattern of the Shapes/Mechanical tests, and the problem solving style influenced by the pattern of the CS, General Abilities and Practical tests, is a strong influence on the effectiveness with which the individual uses the profile she has. This may contextualise the structure of intellect possessed by any individual, partly because various planning styles and problem-solving styles may be perceived as more or less appropriate for different work-related situations, and partly because a less effective problem solving style may vitiate a high scoring profile, and an effective one may maximise a low scoring profile.

This multidimensional cuboid may take almost any appearance, with any one of the dimensions larger or smaller than any one of the others. This concept allows due weight to be given to the value of patterns of ability, and also permits the overall size of the "cube" to be influenced by the whole without undue importance being artificially attached to the "g" dimension. It is possible, for example, for a pair of "cubes" of identical volume to be formed either of a very small amount of "innate understanding" and a very large amount of practicality, or vice versa. Sheer quantity does not indicate the quality of ability.

Fig. 3.1 Possible "cuboids":



More practicality — More reasoning

Less reasoning — Less practicality

The Morrisby Profile (MP) allows for the relationship between aspects of ability, both positive and negative, and this is what gives it its differential nature. Many of the tests intercorrelate, either positively or negatively, because most aspects of ability bear a positive or negative relationship to one another.

This is not a disadvantage, as it would be in a traditionally conceived battery, because instead of contributing equally towards a final ability score, the separate tests and subgroups of tests allow a profile of the candidate to be built up from the differences between scores on pairs or subgroups of the test.

Unlike a traditional test battery, the separate tests of the MP do not each measure some abstraction called ability. Instead, each measures some specific aspect of ability, and none of them measures each and every ability.

This allows the battery greater flexibility. Not only can single subtests measure separate dimensions directly, but the very low correlations of some tests with some abilities which they were not specifically designed to measure allow differentials to be established between the scores on the ability measure - that is, the ability they were designed to measure - and scores on the "non-ability" measure. When interpreting test scores, the differences - the differentials - between two test scores are as important as their average level. For example, higher scores on a numerical than on a verbal test would, in the absence of contaminating factors, predict greater success in a commercial field than would another individual's scores which stood at a higher level in absolute terms but which reversed the preference.

The Morrisby Differential Test Battery (Morrisby 1955) assumes that the difference between an individual's scores on subtests is predictive of occupational success, although not of any absolute level of intellectual ability. Thus it could be argued that the functions measured by the Morrisby tests are those required for success in various occupationally linked activities, rather than being intellectual abilities. However, it is difficult in practice to see how intelligence or "absolute ability" can be considered, divorced from ordinary human measures of success in occupationally related terms. Without wishing to add to all the definitions of intelligence quoted at the beginning of this work, it could be argued that a quality which is unrelated to performance in daily life, and therefore the work-place, deserves the label less than those qualities which facilitate success.

Although many lay test users would support the two propositions that strengths are better predictors of performance than actual score levels, and that different types of intelligence are differentially appropriate, there is little validational evidence for either contention. Much of Chapter 5, therefore, will be taken up with establishing the effectiveness of this alternative approach, both by examining score patterns made up of individual subtests, and by examining patterns made up of larger "blocks", to assess the theory of "block differentials" or problem solving approaches.

However, before examining the effectiveness of differential approaches, the battery itself must come under closer scrutiny.

Chapter 4 - Standardising and validating the Morrisby Profile

When John Morrisby died in 1976, he had established the empirical usefulness of his test battery to his own and his clients' satisfaction, but, presumably as the result of his experiences with Burt, had not been particularly interested in producing or publishing traditional validation data himself, or in standardising the tests again once the original 1955 norms had become out of date.

This meant that, in order for this work usefully to be progressed, it was necessary as part of this study for the author, who was employed at the time as the only full-time psychologist of the organisation, to update, standardise and validate the Differential Test Battery so as to ensure that it could appropriately be used as a selection and development tool. In order to establish the usefulness of the differential approach, a comparison had to be made with the traditional approach, and therefore it was important to ensure that the battery had been standardised and validated along accepted lines so as to comply with current requirements.

Although the conception of the old Differential Test Battery remained unaltered, some minor rewriting was also necessary, partly because of changes in vocabulary, educational patterns and cultural assumptions since the test was written, and partly in order to ensure, as far as possible, that the battery did not disadvantage candidates from minority groups. There was also considerable care taken to ensure that, so far as possible, the tests were assessing aptitude, or potential, rather than educational attainment.

It should be acknowledged here that the present author's involvement with the Mechanical Ability Test was limited to establishing its reliability and examining individual point biserial correlations to determine the fairness of individual items in relation to gender and ethnicity. She was not responsible for rewriting and redrawing items for the Mechanical Ability Test, which would have been well outside her sphere of competence, and this was done instead by Malcolm Morrisby, son of the original author of the tests.

Accordingly, during 1990/91, an initial item analysis was performed on the tests of the DTB which resulted in a number of items being replaced and a new item order created. The revised version was piloted on groups of volunteer students in schools and colleges. (N = 678)

Item analyses and Rasch analyses were run on the results of these pilots, establishing the separate reliability and difficulty levels of the individual items. In view of the results the items were re-ordered and a number were dropped, on the basis that point biserial correlations did not reach a cutoff point of 0.3.

A final pilot established the reliability levels of replacement items, and the new tests were then standardised on a stratified sample of 3016 students from fifth forms across the United Kingdom, covering the full range of ability, approximately equal numbers of both gender groups, and 9% of whom classified themselves as belonging to one of the ethnic minority groups. The sample included over 200 who regarded English as their second language. The standardisation was done with the assistance of twenty-four of the separate Careers Services in England, Scotland and Wales, who suggested suitable schools for the purpose. Over forty schools took part in the procedure. Explanatory notes were sent to schools, parents and pupils before each testing session, and translations of the parents' notes into Bengali, Gujurati and Hindi were available. Pupils were asked to volunteer for the sessions, and to complete an interest inventory in advance of the session to assist in the preparation of appropriate computerised narrative feedback. The present author visited each school to administer the new form of the tests, and also briefed careers officers and teachers to explain the written feedback which was sent to each person who took part in the standardisation.

A further 1,000 candidates, representing sixthformers, graduates, managers and technicians, were then used to complete the initial norming process. General population norms, as well as separate norms for 14 and 15 year olds, graduates and managers, were prepared. These norms were computed by Malcolm Morrisby, and not by the present author.

The changes from the DTB to the Morrisby Profile could be characterised under two headings: presentation and content.

Presentation

A simpler and more economical method of presentation was devised, replacing test booklets which contained both questions and answers with reusable books, multiple choice format questions and single answer sheets. A small study of equivalence was performed on the only test which had not previously been presented in multiple choice format, the Numerical test, and no significant difference was found between the groups. (T-values between .04 and .30)

Table 4:1.

Subjects:

28 Year 11 students
15 female; 13 male

Group 1: Multiple choice presentation
Group 2: Original presentation

Materials:
Morrisby Profile.

Method:

The students were asked to sit the Morrisby Profile, and expected to receive vocational guidance as a result. They were administered the tests together at one sitting, but in reality they had been split at random into two groups of 14, and each group received different versions of the Numerical Ability tests. Group 1 received the "old" version, in which students were asked to respond to open numerical questions by writing the answers directly on to question sheets.

Example: *Complete the sequence 2 4 6 ? ? 12. Replace the question marks with the correct numbers.*

Group 2 received the "new" version, in which multiple choice responses were to be written onto separate answer sheets:

Example: *Complete the sequence 2 4 6 ? ? 12. The correct answers should be selected from the boxes below, choosing the first response from Box 1 and the second response from Box 2.*

Box 1: 6 8 10 11 12 Box 2: 8 9 10 12 100

T-tests were then performed on each part of the Numerical Ability Test, and on the test as a whole, to determine whether or not the groups differed significantly in their scores. It was assumed that any significant difference would result from the different manner of presentation.

Results:

Table 4:2

T-test for: Numerical (Part 1)

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	13	34.77	10.54	2.92
Group 2	15	34.93	10.05	2.60

t-value	-.04
2-tail Prob.	0.97

Table 4:3

T-test for: Numerical (Part 2)

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	13	28.77	8.93	2.48
Group 2	15	29.33	9.03	2.33

t-value	-.17
2-tail Prob	0.87

Table 4:4

T-test for: Numerical (Whole test)

	Number of Cases	Mean	Standard Deviation	Standard Error
Group 1	13	8.92	2.43	.67
Group 2	15	9.20	2.46	.63
t-value	-.30			
2-tail Prob.	0.77			

On the basis of these tests it could be seen that there was no significant difference between the two methods of presentation ($p < 0.77$), and it was therefore presumed unlikely that candidates would be greatly disadvantaged by the move to separate answer sheets and multiple choice. However, all the tests of the new Morrisby Profile were standardised on a large and representative reference group to ensure norms were meaningful and fair.

Content:

Compound Series Test.

On the basis of the Rasch analyses referred to, the items of the CST were reordered to conform more closely to difficulty levels established on item analysis. No other significant changes were made.

General Abilities

The Morrisby general abilities (Verbal, Numerical and Perceptual ability) are designed to assess potential in these areas rather than attainment and learned skills. It was necessary therefore to consider how closely the level of vocabulary and amount of arithmetic or geometric knowledge might obscure a true measure of potential.

For this reason, sub tests requiring candidates to compare words of increasing levels of difficulty, or to perform addition and multiplication tasks at speed, were dropped from the test, and items selected which depended for their discrimination effect upon the ability to manipulate verbal concept or numbers, rather than on vocabulary or arithmetic skills.

During the item analyses run on these tests, all verbal items which appeared to discriminate between members of the ethnic minorities were dropped, as were those items which tended to discriminate against those speaking English as a second language on the grounds of language rather than actual difficulty.

Spatial ability - The Shapes Test

Apart from the re-ordering of items according to a Rasch analysis, this test was not altered.

Mechanical ability

Three principles operated in rewriting this test; the wish to ensure an acceptably simple level of language in the written questions, so that candidates low in verbal ability were not disadvantaged in taking this test; the wish to ensure that the item content covered mechanical principles but did not unduly stress learned mechanical knowledge, and the wish to present the pictures and concepts of the test in an acceptable way to gender and minority groups.

The test was rewritten to incorporate as much material as possible which, while still assessing mechanical concepts such as leverage, the use of action and so on, was presented in terms likely to be familiar to those without much formal mechanical experience. This test was also rewritten with due regard for a balanced presentation of male/female black/white figures in the accompanying pictures. Item analyses were run by the author using gender groups, ethnic groups and those speaking English as a second language, and items which discriminated against members of those groups on the grounds of group membership rather than ability, as shown in item analyses of the whole population, were dropped or, in the case of "language" items, rewritten in simpler terms.

Equivalence of DTB and MP

MP was intended to be a more appropriate measure for the 1990s than DTB, in that it is relatively attainment-free, was written with a due regard for equal opportunities, and, in its Vocational Guidance Service form, is closely linked with the needs of a modern careers service. However, its conceptual equivalence with the old DTB was established to ensure its validity in occupational terms.

In an interim study of 32 adults between the ages of 19 and 54, who took MP between 12 months and 2 years after taking DTB, the following correlations were observed:

Subjects:
N = 32
Age range 18 to 54
18 female: 14 male

Materials
Morrisby Differential Test Battery
Morrisby Profile

Method:

32 adults volunteered to take Morrisby Profile who had already taken the Morrisby Differential Test Battery within the past 2 years. The correlation between the two versions of each test were then calculated, using SPSSPC.

Table 4.5
Correlation between DTB and MP tests

Test	r
CST	.92
Verbal	.82
Numerical	.81
Perceptual	.93
Shapes	.90
Mechanical	.86
Speed Test 1	.76
Speed Test 2	.75
Speed Test 3	.85
Speed Test 4	.62
Speed Test 5	.77
Speed Test 6	.79

Although a larger study is necessary before the equivalence can be assured, it would appear that the underlying concepts of the DTB are retained in MP.

The reliability of the Morrisby Profile

The reliability of the Morrisby Profile was determined by the author using the test-retest method and Cronbach's alpha, to establish internal consistency.

Subjects:
N = 341
Year 11 and Year 12 students in secondary education
169 female: 172 male.

Materials:
Morrisby Profile

Method:

The Morrisby Profile was administered to the students, and the reliability coefficient for each test was calculated, using the Reliability module of SPSSPC. The results may be seen in Tables 4.6 and 4.7 below. Using the test-retest method, the Morrisby Profile was administered a second time to the group, after an interval of six weeks, and the correlations between first and second testing for each test were calculated, as well as an overall figure for the whole battery (0.951).. These results may be seen in Table 4.8

Table 4.6

Internal Consistency:

TEST	No. Items	r	Mean	S.D.	SEm σ	SEM R.S.
Compound Series	60	.93	38.68	8.33	0.26	2.17
Verbal	60	.91	35.26	6.98	0.30	2.12
Numerical	64	.96	37.72	7.61	0.21	1.61
Perceptual	60	.89	35.83	10.29	0.34	3.45
Shapes	60	.95	34.34	10.22	0.23	2.31
Mechanical	40	.81	24.01	5.87	0.44	2.59

N = 341

The Verbal, Numerical and Perceptual tests consist of two parts each. Separate statistics for these subtests are given overleaf:-

Table 4.7

TEST	No. Items	r	Mean	Std. Dev.	SEM σ	SEM r
Verbal 1	30	.85	16.39	3.43	0.38	1.31
Verbal 2	30	.81	18.87	4.31	0.44	1.88
Numerical 1	32	.94	19.45	5.96	0.24	1.45
Numerical 2	32	.88	18.27	5.38	0.34	1.83
Perceptual 1	30	.82	17.62	4.20	0.42	1.78
Perceptual 2	30	.79	18.21	4.39	0.46	2.02

N = 341

Table 4.8

Test-retest reliability

Test	"r"	SEm σ	SEM R.S.
Compound Series	0.90	.32	2.73
Verbal	0.85	.39	1.96
Numerical	0.82	.42	3.13
Perceptual	0.81	.43	2.51
Shapes	0.92	.28	3.99
Mechanical	0.83	.41	2.16
Speed Test 1	0.79	.46	6.87
Speed Test 2	0.72	.53	4.30
Speed Test 3	0.81	.43	2.12
Speed Test 4	0.81	.44	3.26
Speed Test 5	0.82	.42	2.73
Speed Test 6	0.78	.47	4.25

Test-retest reliability over all 12 tests

Sample: as above.

r = .95

The validity of the Morrisby Profile

Construct validity data is presented for the ability tests of the Morrisby Profile by showing the correlations of the tests with established ability tests. Concurrent validity data for the ability tests is presented by showing correlations with state examinations. Concurrent validity is also presented in four multiple regression studies, using examination performance and three different measures of performance in the workplace as the criteria. It is considered that multiple regression is preferable to separate correlations in seeking to establish the validity of a whole battery rather than that of its separate components.

The objective personality tests are factor analysed with dimensions of the OPQ as marker variables (Kline 1993) and their validity established by performing multiple regressions with success in sales as the dependent variable.

Validity of Morrisby Profile ability tests

Correlations with ability tests: 1

Subjects:
103 careers guidance officers
Age range: 22 to 58
59 female, 43 male.

Materials:

Morrisby Profile: the first six ability tests and Speed Test 1, which measures speed and accuracy
Ravens Standard Progressive Matrices
Alice Heim 5
Alice Heim 6 (AG)
In the results section, the tests are indicated as follows:

AH51 - Alice Heim 5, Part 1: Verbal and Numerical
AH52 - Alice Heim 5, Part 2: Diagrammatic
AH5T - Alice Heim 5 - Total: Verbal, Numerical and Diagrammatic
AH61 - Alice Heim 6 - Part 1: Verbal
AH62 - Alice Heim 6 - Part 2: Numerical and Diagrammatic
AH6T - Alice Heim 6 - Total: Verbal, Numerical and Diagrammatic
AHt - Alice Heim Total (AH tot) = AH5 Total OR AH6 Total(whichever was taken).

Method:

The subjects for this study were all graduate careers advisers working in one of the local authority careers services. All took the Morrisby Profile as part of their training to use psychometric tests with young people. A number of them subsequently took some or all of a selection of ability tests made available to them during their training in the use of psychometric tests. It should be noted that they were under no pressure of selection, nor were they in need of guidance, so motivation may not have been high with many of the group, which would be likely to reduce the size of the correlation coefficients.

The results were then correlated with the separate tests of the Morrisby Profile, and the correlations are shown below:

Table 4.9

Morrisby	Raven's	AH5 1	AH5 2	AH5t	AH6 1	AH6 2	AH6T	AHT
N	103	30	30	30	54	54	54	84
CS	.56**	.17	.41#	.33	.49**	.57**	.59**	.45**
Verbal	.47**	.58**	.22	.49**	.54**	.30*	.47**	.42**
Numerical	.57**	.59**	.52**	.67**	.51**	.49**	.56**	.56**
Perceptual	.65**	.48**	.49**	.56**	.50**	.50**	.56**	.53**
Shapes	.60**	.59*	.56**	.68**	.48**	.48**	.55**	.53**
Mechanical	.53**	.30	.50**	.43*	.30*	.40**	.39**	.40**
Speed & Acc	.35*	.12	.10	.14	.21	.12	.18	.19

** Pr < .001   * Pr < .01   # Pr < .05

It should be noted that the correlations between the CST and Raven's Progressive Matrices, which is regarded as a measure of fluid intelligence, reach high levels (0.56 $p < .001$), although it is interesting to see the even higher correlations between RPM and the Morrisby Perceptual test. Ravens has a very high visual component, with a higher diagrammatic and spatial element than the CST, as well as being less attractively set out, and these factors are possibly reflected in these correlations.

The relationship between the Verbal and Numerical tests, and the Alice Heim tests assessing these elements, follows a predictable pattern. The Verbal test correlates with the Verbal element of the AH5 and AH6 (0.58, 0.54) and the Numerical 0.59, 0.49). The Perceptual correlates with the Diagrammatic elements (0.49, 0.50) and in all cases $p < .001$.

Correlations with ability tests: 2

Subjects:

54 Careers Guidance Officers

Ages: ranging from 23 to 54

(32 female; 22 male)

Materials

Morrisby Profile

Bennett's Mechanical Comprehension Test.

Method

The subjects took the Morrisby Profile and, two days subsequently, were given the Bennett's Mechanical Comprehension Test. The results were correlated and are given below. Although it was assumed that the strongest correlation would be with the Morrisby Mechanical Ability Test, it was expected that significant associations would appear with other cognitive tests, partly because of generalised test-taking skills, and largely as Bennett's test requires spatial and, conceivably, verbal skills.

Table 4.10

Correlations between Morrisby Profile and Bennett's Mechanical Comprehension Test:

Morrisby	**Bennett**
N	*54*
CS	.51**
Verbal	.51**
Numerical	.59**
Perceptual	.43**
Shapes	.67**
Mechanical	.82**
Speed and Accuracy	.11
ST2	.00
ST3	.21
ST4	.16
ST5	.05
ST6	.19

** Pr < .001   * Pr < .01   # Pr < .05

The expected correlations are those between Bennett's test and the Mechanical Ability test (0.82 $p < .001$). However, those scoring highly on mechanical tasks often show corresponding strengths on Spatial tasks, and this can also be seen in the table ($r = 0.67$, $p < .001$). The lower (0.43) correlation, although still highly significant, with Perceptual (diagrammatic) ability may be surprising, but Bennett's test utilises highly representational pictures rather than diagrams. It is also possible that the gender imbalance in the group affected this particular association, as women tend to underachieve on mechanical tests, but men and women perform equally well on tests of perceptual ability as may be seen in the data relating to Fairness described later in this chapter.

Concurrent validity data: correlations with state examinations

Concurrent validity data is also presented of correlations of the separate ability tests with a number of GCSE examinations taken by 658 students at secondary school.

Subjects

534 Year 12 students (ages 16-17)

286 male

248 female

Materials:

Morrisby Profile

GCSE results, grades A to G. Each grade was "scored" (8 points for A, 7 for B, 6 for C and so on). A "composite" GCSE result was also produced for each student by totalling her/his scores, thus reflecting both the number and quality of GCSE passes.

Method:

Morrisby Profile measures were used in a predictive validity study with 638 school students to predict GCSE results. Each ability test was correlated with individual GCSE results, and also with a composite score ("GCSE"), derived from each subject's total number of GCSE passes, with higher scores allotted to higher grades. The results of this study are set out below. It should be noted that the numbers varied unavoidably between subject groups, with a very large number - over 500 - taking the commoner subjects, and very few taking Computer Studies or Economics, and that few conclusions can therefore reliably be drawn from the smaller groups, despite the apparently high correlation coefficients.

Table 4.11
Correlations with GCSE examination results

	GCSE	Eng.Lng.	Maths	Econ.	Bus.St.	French	Physics
N	*534*	*520*	*518*	*18*	*18*	*397*	*188*
CS	.28**	.22**	.45**	.07	.35	.20**	.32**
Verbal	.41**	.40**	.46**	.50*	.36	.42**	.29**
Numerical	.40**	.31**	.55**	.43	.65*	.35**	.35**
Perceptual	.35**	.27**	.51**	.47#	.32	.31**	.40**
Shapes	.28**	.14**	.41**	.63*	.15	.22**	.25**
Mechanical	.23**	.03	.36**	.70**	.14	.11#	.25**
Speed & Acc	.23**	.27**	.27**	.43	.23	.21**	.29**

	Chem	Biol	Sci.1	Sci.2	Cmp.St.	Home Ec.	Music
N	*196*	*227*	*73*	*186*	*37*	*27*	*38*
CS	.32**	.34**	.29*	.35**	.30	.38#	.28
Verbal	.24**	.33**	.47**	.46**	.56**	.32	.17*
Numerical	.37**	.30**	.41**	.39**	.33	.43#	.42#
Perceptual	.33**	.38**	.22#	.44**	.23	.30	.35
Shapes	.33**	.30**	.28#	.44**	.06	.42#	.19
Mechanical	.33**	.33**	.39**	.44**	.23	.30	-.03
Speed & Acc.	.12	.18*	.15*	.32**	-.09	.34	.35#

** Pr < .001   * Pr < .01   # Pr < .05

As can be seen in the table, correlations between Verbal and English (0.4); Numerical and Maths (0.55), Spatial/Mechanical/Perceptual with double science (0.44 in all three cases) Perceptual with Physics (0.4); French with verbal (0.42) and most of the other coefficients follow expected patterns.

Although these tables only show the relationships between the separate tests and other accepted measures of supposedly related constructs, they are reproduced here both to establish the traditional "validity" of the battery, and to suggest that, without prejudice to the foregoing, there are strong correlations between most tests and most academic criteria. The tables above show a number of high correlation coefficients between subtests which are not especially designed to measure their particular constructs, but presumably a relationship still exists between the capacity to pass an examination - almost any written examination - and the capacity to take and do well on a paper and pencil psychometric test. It is difficult otherwise to account for the correlations between Bennett's Mechanical Comprehension test, for example, and the Morrisby Verbal Test (0.51 $p < .001$) or the relationship between Perceptual Ability and French (0.31; $p < .001$)

Although further discussion of validity more appropriate to the whole battery rather than separate tests may be found later in Chapter 5, it is useful to show here four further demonstrations of the validity of the battery, using the tool of multiple regression

Concurrent validity: multiple regression.

- The tool of multiple regression is less commonly used in test manuals to validate tests, even when the same test manuals recommend the tests should be used as composites. Nevertheless, it is probably more helpful to see whether the battery as a whole can predict performance on a criterion, than to attempt in the face of reality to associate single tests with performance. High performance in examinations may well be associated with a number of factors, each of which may conceivably be assessed by a separate test. For example, scientific achievement may be linked to numerical and perceptual reasoning, spatial and mechanical ability, some element of fluid intelligence, motivation and, especially in the case of those who have not traditionally perceived themselves as "scientists", self-confidence.

Multiple regression: Study 1.

Accordingly, using the group of students described above, (N = 534) and taking as the first criterion the number of GCSE's passed at Grades A to C, the following multiple correlations were obtained with the battery. Figures are given for the multiple squared correlation (R) with all twelve tests.

Table 4.12

Multiple R	.45
R Square	.21
Adjusted R Square	.19
Standard Error	25.63

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	106548.97	8879.08
Residual	626	411107.17	656.72

F = 13.52 Signif F = .00005

Beta coefficients and T values:

VARIABLE	BETA COEFFICIENT	T VALUE	SIGNIF.
ST6	-.01	-.40	.6976
VERBAL	.38	8.62	.00001
ST2	-.01	-.301	.7639
ST1	-.11	-2.22	.0267
CST	.02	.45	.6541
ST4	-.03	-.52	.6007
SHAPES	.07	1.33	.1832
NUMERICAL	.09	1.87	.0625
ST5	.17	3.43	.0006
MECHANICAL	-.02	-.31	.7535
PERCEPTUAL	-.03	-.47	.6354
ST3	.01	.154	.8775
CONSTANT		48.42	.0000

It may be seen from this table that the multiple correlation coefficient of 0.45 shows that the battery can explain about 20% of the variance in number of GCSEs passed. This figure may be compared with the correlations of GCSE with the ability tests alone (0.23 to 0.41) which were given in table 4.11. The higher figure here is due to the inclusion of the speeded personality tests, which increase the predictive power of the battery. In assessing the value of this figure, it is important to consider that the tests are merely tests of aptitude and potential, whereas GCSE also assesses teaching standards, knowledge, opportunities and motivational factors. It may be seen from the table of Beta coefficients and T values that Verbal carries a significant amount ($p < .00001$) of the weighting, as does the measure of manual speed ST5 ($p < .0006$). In the light of comments relating to general test taking ability, this supports the suggestion that academic prowess is not always related to intellectual power.

Nevertheless, in traditional terms, in the light of this experiment, the multiple correlation can be regarded as the index of efficiency of the battery (0.45, $p < .00005$).

Multiple regression: Study 2

A second multiple regression was carried out with a small sample of 24 managers, for whom appraisal grades relating to their promotion prospects were available.

Subjects:

24 managers

Age range 23 to 42

20 male; 4 female

Materials:

Morrisby Profile

Appraisal grades

Method:

Senior managers rated the candidates on a scale of 1 to 5, ranging from 1 (Unsatisfactory) through to 5 (Outstanding). This rating was used as the criterion and the six ability tests and four personality measures of the Morrisby Profile were used as the dependent variables in a multiple regression, to establish whether the whole battery of tests could predict performance.

Table 4.13

Multiple R	.55
R Square	.30
Adjusted R Square	-.24
Standard Error	1.11

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	6.95	.70
Residual	13	16.05	1.24

F = .57 Signif F = .82

In this experiment, the multiple correlation,0.55, although it is actually higher than the coefficient in the previous experiment, cannot be taken as the index of efficiency of the battery. The numbers are very low in this second dataset and the F value is not significant. The criterion available was also not very satisfactory, as most of the managers' appraisals fell around the mean. Appraisers are often reluctant to rate individuals at extreme levels, (Woodruff 1990) which can reduce the usefulness of this sort of exercise, although it has been included here for the sake of completeness.

Several studies have found that factors unrelated to the structured assessment procedure affect supervisor ratings, such as the nature of the organisation, personal characteristics of the rater and the position within that organisation of the individual under review. Halo effect, errors of central tendency (particularly noticeable in this study), range restriction, "similar to me" syndrome, contrast, recency and primacy are all likely to occur. According to Landy and Farr (1980),individual differences in raters cause most of the errors. Mitchell & Wood (1980) and James and White (1983) claimed that supervisors are liable to attributional biases. Lane and Herriot (1990) found that in all performance areas, except the capacity to delegate and control, self ratings were superior to supervisor ratings.

Multiple regression: Study 3.

A third dataset was examined, consisting of a group of insurance salespeople, for whom sales figures and promotion levels based on sales and length of service were available.

Subjects:

28 insurance salespeople

3 female, 25 male

Age range: 23 to 28

Materials:

Morrisby Profile

Sales figures, length of service and promotion grades.

The six grades, in ascending order, are trainee, consultant, senior consultant, sales executive, branch executive and unit manager.

Method:

A multiple regression were performed, with grade as the dependent variable, using the full Morrisby Profile.

Results:

Dependent Variable. Grade

Independent Variables: 10 tests of Morrisby Profile (ability and personality measures, but excluding the two dexterity measures)

Table 4.14

Multiple R	.80558
R Square	.64896
Adjusted R Square	.42956
Standard Error	.96771

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	27.69963	2.76996
Residual	16	14.98350	.93647

F = 2.95788 Signif F = .0260

It may be seen that the profile predicts sales success ($R = 0.81$, $p = .03$) even with a small sample. In view of the difficulty in reaching significance with any tests using numbers as low as 28, this can be interpreted as indicating the strength of the measures, particularly with reference to the preceding study. However, this particular dataset will be examined later in the light of the importance of personality measures in predicting sales success.

Multiple regression: Study 4.

A fourth dataset was examined, consisting of 135 applicants for engineering and technical posts, 54 of whom were short listed, eighteen of whom were selected and then considered for further promotion, and six of whom were selected to head their teams.

Subjects:

135 applicants for engineering and technical posts:

Age: 19 to 42

Male: 123: Female: 12

Materials:

Morrisby Profile

Engineering and process test results

Team qualities assessments

Promotion decisions

Selection decisions

Method:

All applicants took the Morrisby Profile, and were selected or rejected on the basis of an application form, previous record and interview. Multiple regression and discriminant function analysis were used, with selection and "Standard" as the Dependent/discriminating variables. "Standard" was a composite variable, made up of the engineering and process test results and the team qualities assessment. Only those subjects assessed for promotion (N = 18) could be used in the regression analysis and discriminant function analysis which used "Standard" as the dependent/discriminating variable.

The criteria for selection were the application form, previous record and an interview. In addition, all the applicants were given the Morrisby Profile, although the results were not used by the selectors, who wished to evaluate the tests against their present system. Those considered for further promotion were given engineering tests and process tests, as well as being assessed for various team qualities. The final decision relating to promotion was made by a panel after interview and test assessment.

Two variables were used, "Selection" and "Standard". In "Selection," a value of 0 meant an individual was not selected; 1 meant he was short listed (all those short listed were male), 2 meant he was appointed and 3 meant he was appointed as team leader. Two multiple regressions were then performed on the tests of the Morrisby Profile, with selection and "Standard" as the dependent variables. "Standard" was a combined score, made up of the engineering tests, the process tests and the team qualities assessment.

When all twelve of the tests were used, with selection as the dependent variable, (N = 135) a multiple R of 0.37 was obtained. When only the six ability tests were used, a multiple correlation of 0.29 was obtained.

When all twelve of the tests were used with "Standard" as the dependent variable (N

= 18), a multiple R of 0.84 was obtained. When only the six ability tests were used, a multiple correlation of 0.64 was obtained. Although the figures for the ability tests and the full profile sound very high, it should be remembered that the numbers were low and significance levels reflect this. The full figures may be found in Appendix B but a summary follows:

Table 4.15

Summary of multiple regression tables showing the association between the 6 ability tests of the profile, and the full 12 tests, to predict the selection of engineering technicians, and the promotion of those selected.

Test	DV	N	Multiple R	Sig of F
12 tests	Selection	135	0.37	.11
6 tests	Selection	135	0.29	.07
12 tests	Internal promotion	18	0.8	.50
6 tests	Internal promotion	18	0.65	.34

Discriminant Function Analysis

Using selection to group the applicants in a discriminant function analysis, the following classification results were obtained:

Table 4.16

Discriminant Function analysis: Classification Results:

Discriminating variable: "Standard" (performance on engineering & assessment tests)

Actual Group	No. of Cases	Predicted Group Membership High	Low
Group 1 (High performers)	9	9 100.0%	0 .0%
Group 2 (Low performers)	9	1 11.1%	8 88.9%

Percent of "grouped" cases correctly classified: 94.44%

The full classification figures have been included in Table 4.16 to show the high degree of accuracy with which in fact the men were classified into groups. The criterion was a particularly strong one, in that the tests used (engineering and process tests) were closely related to the work the men were doing; the same assessors were used on all the men for all tests, and the personality assessment measures, although somewhat subjective, were again consistently applied by the same assessors. It is unfortunate that the study contained so few subjects, but it has been included as an example of the effectiveness of the battery as a composite measure, without, at this point, examining it as a measure of profile patterns, which will be done in the next chapter.

It should be noted that, to avoid attenuated results, it has been necessary to use only those groups of test subjects for whom test scores and selection or performance criteria were available, but who had not been selected, promoted or appraised either partly or wholly on the basis of those test scores. Such samples are particularly difficult to find, which may explain why all such datasets using performance criteria in this study are unavoidably small.

Multiple Regression: Study 5

Vocational Guidance candidates (N = 13000)

M = 7468

F = 5532

A much larger sample of mixed candidates for vocational guidance, ages ranging from 15 to 58, mean age 26.3, was used in the final study in this section. The grouping variable used was taken from their score on an interest inventory, which classified individuals according to their strongest interest taken from People, Data and Things (Practical /Technical/ Scientific). There were some misgivings as to the usefulness of the criterion, as so many candidates for vocational guidance complete such interest inventories virtually at random, and in any case expressions of interest in careers may not entirely match the ability to fulfil those aspirations. Another factor, partly controlled for in this study, was the well documented stereotypic effect of gender on career aspirations (Hammond and Dingley 1989), which could well have encouraged some girls to avoid the scientific and technical careers, despite their aptitudes, and some boys to avoid the caring professions.

Subjects

13000 Vocational Guidance candidates

7468 males

5532 females

Materials

Morrisby Profile results

MVQ interest scores

Method:

13000 candidates for vocational guidance, who had also completed the MVQ (Morrisby Vocational Questionnaire) supplied their Morrisby Profile and MVQ scores. The majority of the subjects were 16-18 year old school students. 7468 males and 5532 females took part in the study. Three multiple regressions were performed on both the male and female subjects separately, to control for gender bias, using People, Data and Things as the dependent variables.

Results

In all six treatments, the battery was able significantly to predict interests from the ten variables of the battery. With each of the criteria it performed best with males.

When males were the subjects, it performed best in predicting interests in Things (R = 0.42, p = .00005) It significantly predicted interests in Data (R = 0.34, p = .00005) and , less well but still significantly, interest in people (R = 0.28, p = .00005).

With females, it performed best at predicting interest in people (0.27, p = .00005), then in things (0.26, p = .00005) and least well, although still significantly, in data (R = .24, p = .00005).

Multiple Regression 5A.

Dependent variable	People
Independent variables	Morrisby ability and personality measures (10 variables) excluding the manual dexterity measures
Subjects:	7468 Males

Table 4.17

Multiple R	.28
R Square	.08
Adjusted R Square	.09
Standard Error	3.42

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	7535.39	745.84
Residual	7448	86946.14	11.67

F = 63.89 Signif F = .00005

Multiple Regression 5b

Dependent variable:	People
Independent variables:	Morrisby ability and personality measures (10 variables) excluding the manual dexterity measures
Subjects:	5532 Females

Table 4.18

Multiple R	.27
R Square	.07
Adjusted R Square	.07
Standard Error	3.21

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	4263.52	426.35
Residual	5506	56636.46	10.29

F = 41.45 Signif F = .00005

Multiple Regression 5C.

Dependent variable: Data

Independent variables: Morrisby ability and personality measures (10 variables) excluding the manual dexterity measures

Subjects: 7468 Males

Table 4.19

Multiple R	.34
R Square	.12
Adjusted R Square	.11
Standard Error	3.01

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	8742.79	874.28
Residual	7448	67554.43	9.07

F = 96.39 Signif F = .00005

Multiple Regression 5D.

Dependent variable: Data

Independent variables: Morrisby ability and personality measures (10 variables) excluding the manual dexterity measures

Subjects: 5532 Females

Table 4.20

Multiple R	.24
R Square	.06
Adjusted R Square	.06
Standard Error	2.96

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	2946.77	294.68
Residual	5506	48211.57	8.76

F = 33.65 Signif F = .00005

Multiple Regression 5E.

Dependent variable: Things (Practical/Technical/Scientific)

Independent variables: Morrisby ability and personality measures (10 variables) excluding the manual dexterity measures

Subjects: 7468 Males

Table 4.21

Multiple R	.42
R Square	.17
Adjusted R Square	.17
Standard Error	3.69

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	21118.22	1759.85
Residual	7446	101645.56	13.65

F = 128.92 Signif F = .00005

Multiple Regression 5F.

Dependent variable: Things (Practical/Technical/Scientific)

Independent variables: Morrisby ability and personality measures (10 variables) excluding the manual dexterity measures

Subjects: 5532 Females

Table 4.22

Multiple R	.26
R Square	.07
Adjusted R Square	.06
Standard Error	3.12

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	3714.70	309.56
Residual	5504	53469.02	9.72

F = 31.87 Signif F = .00005

Table 4.23
Table of beta weights and T-values for the above dataset:

Variable	Beta	T	Sig T
ST6	.01304	.92	.36005
ST2	-.04675	-2.63	.00860
SHAPES	.09039	4.62	.00005
ST1	-.04215	-2.39	.01690
ST4	-.01917	-.98	.32790
ST3	-.04049	-2.08	.03770
ST5	-5.555E-03	-.26	.79290
CST	.02593	1.01	.31420
GAT-V	-.34080	-12.02	.00005
NAT	.28283	10.04	.00005
GAT-N	-.10800	-3.78	.00020
GAT-P	.22778	6.85	.00005
(Constant)		34.27	.00005

It is interesting to note that, in the group of female subjects, interest in practical scientific/technical subjects is most strongly associated with Spatial ability ($t = 4.62$; $p = .00005$), Mechanical ability ($t = 10.04$; $p = .00005$) and Perceptual ability ($t = 6.85$, $p = .00005$). It is also negatively associated with Verbal ability ($t = -12.02$, $p = .00005$)

Validity of Morrisby Profile - objective personality measures.

The construct validity of objective personality measures, which do not resemble the more familiar questionnaire-based dimensions of the 16PF, OPQ and similar measures, is not easy to establish, in view of the dissimilarity of most available criteria. However, in this study two methods have been used.

First, taking dimensions from the OPQ as marker variables (Kline 1993), the objective measures have been factor analysed to examine high and low loadings on factors where the marker variable can be seen to load high. Secondly, to establish concurrent validity, the group of insurance salespeople referred to earlier have been used, employing the criterion of sales success to see if it can be predicted from the objective measures alone.

Although there are some objections to the use of self report measures to "validate" those based on non-self report, Kline's approach has been taken on the assumption that, even if inter-correlations are low, some meaningful associations will be apparent. Accordingly, three factor analyses were performed, one using the Morrisby Speed Tests with the OPQ (N = 34); one with the Myers Briggs Type Indicator (N = 32) and one using the Speed Tests with the Morrisby Compound Series Test (N = 3016), which allowed a measurement of reasoning ability to be included in the assessment. The results may be found below, and they would appear to indicate that the four measures are assessing constructs that are also produced by more conventional measures. There is, of course, no means of knowing by this method whether they are more objectively valid than data extracted by questionnaire, or whether they extract other types of information not easily extracted by the more traditional approach..

The Morrisby dimensions are given in the analysis as follows:

ST1: awareness, responsiveness, sensitivity, can mean, if very high in relation to the other scores, shyness, lack of confidence and over sensitivity.

ST2: perseveration: tenacity, resistance to change, task completer

ST3: inner certainty and conviction, commitment, inner drive, can be stubbornness.

ST4: outward confidence, initiative, decisiveness, innovativeness and ideational fluency.

In the analysis with the OPQ, the four tests were entered as single dimensions: in the MBTI study, they were entered as "strengths" on continua, as follows:
ITF (Inflexible/Tenacious/Flexible); PMO (Inner conviction/outward decisiveness/ideational fluency)NEC(Non-confident/even/confident - relating to the primary confidence score)CLMH Confidence low/medium/high(relating to the secondary confidence score).

The full explanation of these continua may be found in chapter 3, in the section on the objective personality tests.

Factor analysis: 1. OPQ and speed tests.

In view of the sheer number of OPQ dimensions, only those considered to relate to the dimensions of the Morrisby modal profile were used as marker variables. Each was put into a factor analysis along with the objective personality measures, and an OBLIMIN rotation was performed. In most cases, as might be expected from so few variables, only two factors emerged. Structure matrices produced are listed below.

Table 4.24
OPQ 2 Controlling

	FACTOR 1	FACTOR 2
ST1	.31437	-.79331
ST2	.78031	.02932
ST3	.06445	-.68564
ST4	.79688	-.17576
OPQ2	.35746	.56959

In Factor 1, this is associated with both ST4 (leadership) and ST2 (perseveration; desire to complete). However, it loads most heavily on Factor 2, and is therefore most strongly associated with insensitivity and lack of inner conviction.

Table 4.25
OPQ5 Affiliative

	FACTOR 1	FACTOR 2
ST1	.75380	.35768
ST2	.03516	.80781
ST3	.75334	-.02734
ST4	.17323	.81785
OPQ5	.79001	.06131

This is associated with ST1 which is designed to measure awareness and responsiveness, and with ST3 (inner confidence; one-to-one assuredness).

Table 4.26
OPQ6 Socially confident

	FACTOR 1	FACTOR 2
ST1	.19809	.85157
ST2	.66899	.12639
ST3	.03182	.70015
ST4	.85589	.17287
OPQ6	.63882	-.50906

This appears to be associated with ST4, a measure of outward confidence and leadership, and low ST3 (usually perceived as internal, but low outward, confidence)

Table 4.27
OPQ 7Modest

	FACTOR 1	FACTOR 2	FACTOR 3
ST1	.79898	-.20206	.24486
ST2	.08619	.00924	.94846
ST3	.83254	.18664	-.03463
ST4	.29363	-.72513	.57618
OPQ7	.12838	.89894	.14920

Modesty seems associated with commitment (ST3), but not outward confidence, and shyness (ST1). The combination of ST1 and ST3, if ST1 is greater, is usually seen as showing lack of confidence. Modesty seems to load most strongly on Factor 2, where it is associated with lack of confidence(ST3).

Table 4.28
OPQ 14 Traditional

	FACTOR 1	FACTOR 2
ST1	.12098	.88611
ST2	.56770	.20701
ST3	.08650	.69601
ST4	.87773	.24123
OPQ14	-.77002	.26550

A low score on this measure seems to be associated with ST4, which shows ideational fluency and innovativeness. It is described as showing those who would "introduce change wherever possible and ..prefer to be in an area..pioneering new methods."

Table 4.29
OPQ 15 Change orientated

	FACTOR 1	FACTOR 2
ST1	.28647	.78997
ST2	.67818	.09094
ST3	-.09649	.83068
ST4	.78671	.27159
OPQ15	.71461	-.12437

This factor again is associated with ST4, and both measure willingness to change and accept new methods

Table 4.30
OPQ 17 Innovative

	FACTOR 1	FACTOR 2
ST1	.14095	.86699
ST2	.68325	.12699
ST3	.01580	.71282
ST4	.81941	.24688
OPQ17	.71072	-.36634

As might be expected, both ST4 and OPQ 17 load on this factor. Innovativeness in its various forms seems measured by many of the OPQ dimensions.

Table 4.31
OPQ 19 Detail conscious

	FACTOR 1	FACTOR 2
ST1	.18480	.82907
ST2	.75654	.08676
ST3	-.05534	.76926
ST4	.72873	.32386
OPQ19	.64418	-.20214

Detail consciousness is associated with perseveration.

Table 4.32

OPQ 23 Toughmindedness

	FACTOR 1	FACTOR 2
ST1	.25253	-.79254
ST2	.82445	-.03396
ST3	-.15596	-.83493
ST4	.75979	-.29198
OPQ23	.21668	.04355

This is not a particularly high loading for OPQ23, but it is interesting that the tough-minded factor associates ST2 and ST4, as this combination of outer confidence and tenacity might well be described as tough-minded.

Table 4.33

OPQ 24 Emotional control

	FACTOR 1	FACTOR 2	FACTOR 3
ST1	.79880	-.11244	.29473
ST2	.04867	.21885	.88527
ST3	.83141	.15581	-.07284
ST4	.29658	-.54988	.71665
OPQ24	.12456	.91504	.13187

This appeared to load most heavily on the second factor, and to be most strongly associated with a negative score on ST4, which measures, when positive, outward, but not internal. confidence.

Table 4.34

OPQ 25 Optimistic

	FACTOR 1	FACTOR 2
ST1	.31047	.76182
ST2	.65878	.12088
ST3	-.09906	.82348
ST4	.79892	.24438
OPQ25	.63466	-.30487

Optimism is associated with high outward confidence - a classic extroversion measure.

Table 4.35

OPQ 28 Competitive

	FACTOR 1	FACTOR 2
ST1	.35956	.76955
ST2	.78908	-.02729
ST3	.03879	.85235
ST4	.78816	.22917
OPQ28	.41626	.17451

Competitiveness is associated with decisiveness, confidence and completing (ST4 and ST2). The secondary factor appears to load on ST3, which is concerned with being self-driven and committed, but not interested in outward forms of competition.

Table 4.36

OPQ 29 Achieving

	FACTOR 1	FACTOR 2
ST1	.21807	.79259
ST2	.78270	-.00019
ST3	.11552	.83669
ST4	.74864	.26066
OPQ29	.73550	.20857

Achieving is associated with decisiveness (ST4) and completing (ST2)

Table 4.37

OPQ 30 Decisive

	FACTOR 1	FACTOR 2
ST1	.32768	.74348
ST2	.72842	-.07051
ST3	-.04094	.82624
ST4	.84520	.28503
OPQ30	.45778	.38712

This factor loads most heavily on ST4 which would appear to be associated with OPQ30. Since both are intended to measure decisiveness (inter alia) this was a satisfactory finding.

Factor analysis 2: MBTI and Speed Tests

In a factor analysis of all four Morrisby personality dimensions and all four MBTI dimensions, three factors emerged after OBLIMIN rotation. The first factor appeared to load on virtually all the MBTI components; the second factor loads on high confidence (NEC is a non-confident /even/confident dimension) and a secondary high confidence (CLMH) dimension, and the third factor loads on flexibility and "thinking" rather than "feeling".

Table 4.38

	FACTOR 1	FACTOR 2	FACTOR 3
STR1ITF	-.06620	.27154	.90913
STR2PMO	.14234	-.50611	.10485
STR3NEC	.18809	-.86204	-.22480
STR4CLMH	-.12927	-.87351	-.19851
MBS1EI	-.69322	.43775	-.06397
MBS2SN	.75655	.13725	-.30229
MBS3TF	.65778	-.20564	.52039
MBS4JP	.82315	-.07531	.11264

Separate factor analyses were then performed for each of the MBTI marker variables.

When Extroversion/Introversion was used as the marker variable, two factors emerged. The first appeared to have little to do with the marker variable, and be a general low confidence and inflexibility factor. The second factor loads most highly on extroversion and outward confidence, which seems to give some credibility to the outward confidence measure.

Table 4.39

	FACTOR 1	FACTOR 2
STR1ITF	-.84799	.32655
STR2PMO	-.02545	.76307
STR3NEC	.75535	.33891
STR4CLMH	.75752	.25952
MBS1EI	-.06555	-.69766

When Sensing/Intuitive was used as the marker variable, again, the first factor appears to have little to do with the MBTI measure, and to load on confidence and, this time, inflexibility. The second factor loads heavily on sensing and inflexibility, or feeling and flexibility, which appears to match the interpretation of both measures.

Table 4.40

	FACTOR 1	FACTOR 2
STR1ITF	-.50656	-.56025
STR2PMO	.43374	-.02597
STR3NEC	.85670	.06406
STR4CLMH	.92290	-.15650
MBS2SN	-.18984	.91284

When Thinking/Feeling is used as the marker, Feeling appears to load with Flexibility. The first factor is largely confidence again, with some loading on Feeling.

Table 4.41

	FACTOR 1	FACTOR 2
STR1ITF	-.43470	.67773
STR2PMO	.45671	.07011
STR3NEC	.88993	.01963
STR4CLMH	.87989	-.09107
MBS3TF	.25115	.87737

The Judging/Perceiving factor

Table 4.42

	FACTOR 1	FACTOR 2
STR1ITF	-.64855	.03627
STR2PMO	.30488	.70922
STR3NEC	.85416	.22545
STR4CLMH	.89414	.10184
MBS4JP	-.10425	.80966

Perceiving seems to be associated with outward confidence ("ideational fluency") intuition and initiative.

Although little can be gleaned definitively from a study with such small numbers (N = 32) this does give prima facie evidence for the existence of the confidence dimensions and probably supports the Flexibility dimension.

Factor Analysis 3: CST and Speed Tests

Table 4.43

	FACTOR 1	FACTOR 2
ST1	.73162	.38079
CST	.57360	.19677
ST3	.41356	.35440
ST2	-.21778	-.08396
ST6	.16218	.70053
ST5	.41809	.60252
ST4	.29095	.46136

The first factor appears to load on awareness, speed and inner conviction or commitment, and some flexibility.. Intellectual functioning appears more strongly associated with this factor. The second factor seems to load on outward confidence, manual speed and deftness associated with writing quickly - it has to be remembered that these are speeded tests.

Table 4.44

	FACTOR 1	FACTOR 2
FACTOR 1	1.00000	
FACTOR 2	.42340	1.00000

Concurrent validity of Morrisby objective personality measures

Subjects:
28 insurance salespeople.
Age range 23-38
3 female, 25 male.

Materials:
Morrisby Profile
Sales figures for 1991/2
Promotion levels for group
Length of service for group.

Method:

28 sales staff, selling insurance and pensions for a large insurance group, provided records of their year's sales figures, their length of service and their level of promotion, which is entirely based on past sales successes. A variable "Grade" was computed from the product of their year's sales figures and their current promotion level, divided by their length of service. ((Sales * Rank)/Time) "Grade" and "Sales" (the year's sales figures) were used as the criteria.

The four personality measures of the Morrisby Profile were standardised and, in addition, three variables were calculated, "CONF1, CON4 and CONF3" . These were, respectively, the mean of the sum of Speed Tests 3 and 4, representing overall confidence; Speed test 4 minus Speed Test 1, representing the differential between outward confidence and awareness, and Speed Test 3 minus Speed Test 4, representing the differential between internal confidence, or commitment, and Speed Test 1.

The intention of the research was to establish, first, the capacity of the personality measures to predict sales success from weighted composites, using multiple regression and discriminant function analysis, and, secondly, the ability of the differentiated approach to predict the same success, by using differentiated scores in the same way.

Results:

Multiple regressions were performed with the standardised measures and the differentiated measures, using Sales and Grade as the dependent variables.

The prefix "Z" indicates a standardised variables: CONF, CONF4 CONF3 are the differentiated variables described above.

Using multiple regression, with "GRADE" as the dependent variable, and the four standardised speed tests as the independent variables, the following results were obtained:

Table 4.45

Multiple R	.58
R Square	.33
Adjusted R Square	.21
Standard Error	.71

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	4	5.56	1.39
Residual	22	11.11	.51

F = 2.75 Signif F = .05

Using the "CONF1 CONF3 CONF4" differentiated scores, similar results were obtained.

Table 4.46

Grade Dependent variable
CONF1 CONF3 CONF4 Independent variables

Multiple R	.53
R Square	.28
Adjusted R Square	.18
Standard Error	.72

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	3	4.65	1.55
Residual	23	12.02	.52

F = 2.93 Signif F = .05

This procedure was repeated with "SALES" as the dependent variable:

Table 4.47
SALES - dependent variable
ZST1 ZST2 ZST3 ZST4 independent variables

Multiple R	.60
R Square	.36
Adjusted R Square	.24
Standard Error	1.34

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	4	21.74	5.44
Residual	22	39.22	1.78

F = 3.05 Signif F = .04

Table 4.48
SALES - dependent variable
CONF CONF3 CONF4 - independent variables

Multiple R	.59
R Square	.35
Adjusted R Square	.26
Standard Error	1.32

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	3	21.09	7.03
Residual	23	39.88	1.73

F = 4.05 Signif F = .02

From these results, it may be seen that both the overall sales figures of the year, and the group's promotion rate, could be effectively and significantly predicted from the objective personality measures. There was little to choose between the performance of the standardised and differentiated variables, given the slight difference in number of variables, in the multiple regression studies. With GRADE, and standardised variables, R = 0.58, p = .05; with differentiated variables, R = .53, p = .05. With SALES, and standardised variables, R = .60, p = .04; with differentiated variables, R = .59, p = .02. The difference in number of variables would account both for the slightly higher coefficients and the lower significance level.

It would seem, therefore, that the objective personality measures effectively predict sales success.

Fairness

Any battery which assigns traditional importance to levels of scores is likely to disadvantage groups of individuals whose performance on traditional tests does not properly reflect their true ability. This will be discussed in greater detail in the following chapters, when the differential approach and its relevance to minority groups and non-traditional methods of assessment will be examined. However, this chapter is concerned with establishing the usefulness of the Morrisby Profile in traditional terms, and so it is necessary to examine the degree to which it may disadvantage minority groups, when used in this way.

Special Needs

Special needs groups in schools will undoubtedly under-perform, relative to the majority of school students. This may not necessarily be discriminatory if real differences in ability are being highlighted, but it is important that such students should feel they are being fairly assessed and that comments on their performance should be couched in realistic but constructive terms. The Morrisby Profile allows for considerable stress to be laid on the performance of individuals in terms of their own relative strengths and weaknesses, and it was hoped that this could be examined by inviting a group of special needs students to comment on their experience with the battery, rather than by examining their actual performance, which would certainly have been below that of the average student.

It is recommended that Morrisby Profile should be used with those whose reading age is above 9.5 years. The tests were standardised on the full mainstream ability range, including some designated as special needs pupils within the main stream.

However, an additional small group of special needs students (all described as "readers" by their teachers) in a special needs school were tested using Morrisby Profile, and reports were sent to them after the test session, based on a differential scoring system which described their test scores relative to their individual mean score. All were asked to complete a short multiple choice questionnaire describing their reaction to the tests and the reports. Seventeen returned the questionnaires and the answers are summarised below. The full text of the questionnaire can be found in Appendix F .

All found the instructions easy or mostly easy to understand, and 94% found they could do something on every one of the tests. One candidate found one test too hard. Given the nature of the group, this was a particularly satisfactory finding.

Most (81%) of the students approved of the presentation of the report, and all but one (94%) found the report helpful. The majority of the group (81%) found both the test and diagrams in the report easy and the remainder "found the text OK, but the diagrams difficult". This is again encouraging, as computerised reports are more difficult to present in a simple and fluent style.

All agreed with some or all of what was said about their ability and all but one agreed with all or some of what was said about the sort of person they were. This was particularly interesting in view of the differential scoring technique employed, which described their strengths and weaknesses in terms of their own average scores, rather than relating them too strongly to the population mean. 81% said they had learned something new about themselves which was worth thinking about.

88% found the job suggestions "very helpful" or "quite useful". 12% found the suggestions "unhelpful".

Only 1 of the students talked the report over with the careers teacher, and 2 with the careers officer. 80% felt they needed someone to talk over the report with them.

The whole group felt that others should have the chance to do the tests, and that people should have the opportunity to receive such reports when they went for careers advice to a careers adviser.

This is not put forward as validation data, as it is clearly a subjective response to a new situation for which no clear criteria existed. Groups may show high levels of courtesy and compliance when faced with questionnaires of this nature, but these responses do suggest that the multidimensional approach of the battery would appear to offer a positive experience to a group which often finds psychometric testing depressing and unconstructive.

Gender and ethnicity

Testing mixed gender and ethnic groups is always a sensitive issue. In some fields, it is well established that groups do perform differently at occupational tasks. Men and women do not perform, on average, at the same levels on visual-spatial or mechanical tasks (Maccoby & Jacklin 1974), so it is reasonable to expect that a test would show up these differences. Similarly, it is likely that a group of native English speakers and a group speaking English as a second or third language would perform, on average, differently on verbal tests of competence in written or spoken English.

It would not be discriminatory for a test to highlight these differences in ability, provided that decisions unrelated to competence in English language use were not made on the basis of these scores.

The Morrisby Profile is designed to present test results as a profile of sub-test scores, so differences between separate tests are used constructively to highlight strengths and weaknesses, and probable occupational preferences, rather than contributing to a higher or lower total ability score. The stress is laid on an individual's personal strengths, not on a level of ability artificially computed by averaging scores on widely differing measures. Thus it would not be helpful to use a separate norming system to reduce real differences in ability, as this would only result in a distortion of an individual's profile of strengths.

In practice, few of the Morrisby tests do distinguish between either the gender groups or the ethnic groups, although it should be said that neither the majority nor the ethnic group contained anyone whose first language was not English. These individuals were excluded from this part of the study, as it was impossible to quantify independently their degree of English knowledge and familiarity. Accordingly, although it is very likely that the tests would disadvantage non-English speakers, the degree of that disadvantage is not clear.

Gender Differences

Subjects:

N = 625

All subjects were in year 11 of secondary school

300 were females and 325 males.

For both groups, the average number of GCSE subjects to be taken at the end of the year was 5.

Materials:

Morrisby Profile

Method:

The students were administered Morrisby Profile, and T-tests were performed on each of the Morrisby tests, to examine whether any significant differences were to be found between the mean scores of the male and female groups. Differences were not regarded as significant unless they reached the .05 level (marked with an *). The .01 level is marked **, and .001 with ***.

N.B. Positive mean difference scores are in favour of females, and negative scores in favour of males.

Table 4.49

Test	Difference of Means	SEM	Significance of Difference
CS	1.31	2.75	0.079
Verbal	0.36	2.91	0.541
Numerical	0.36	3.08	0.638
Perceptual	0.95	4.02	0.141
Shapes	-6.08	2.42	0.001***
Mechanical	-3.55	2.60	0.001***
ST-1	8.26	9.53	0.001***
ST-2	0.37	4.97	0.682
ST-3	0.30	2.24	0.549
ST-4	0.11	4.22	0.864
ST-5	2.75	3.49	0.001***
ST-6	-3.33	9.72	0.014**

*** Pr < .001 ** Pr < .01 * Pr < .05

Although some of these differences do reach significant levels as indicated by the final column, it should be noted that most of the mean differences are less than the range allowed for by the standard error of measurement. However, the Spatial and Mechanical tests, and Speed Test 1 (Speed and Accuracy) do differentiate between males and females, to the disadvantage of the latter in the case of the first two mentioned, and to the disadvantage of males in the case of Speed Test 1.

Ethnic Differences

Subjects:

1016 students in secondary education, age range 15-16

939 belonged to the majority group

61 belonged to an ethnic minority group

Materials:

Morrisby Profile

A brief ethnic monitoring questionnaire (Which of these best describes your family group - Afro-Caribbean, Arab, Asian, Bangladeshi, Chinese/Japanese, Indian, Latin-American, Pakistani, white, other, prefer not to say.)

Method:

1016 fifthformers were tested, of which 61 belonged to an ethnic minority group and 939 to the majority group (16 preferred not to say). For both groups, the average number of GCSE subjects to be taken at the end of the year was 5. All those tested regarded English as their first language. This ensured that ethnicity, not language as such, was being examined.

T-tests were performed on each of the Morrisby tests, to examine whether any significant differences were to be found between the mean scores of the ethnic and non-ethnic groups.

N.B. Positive mean difference scores are in favour of the minority group, and negative scores in favour of the majority group.

Table 4.50

Test	Difference of Means	SEm	Significance of Difference
CS	-1.75	2.75	0.106
Verbal	-1.51	2.91	0.158
Numerical	-0.94	3.08	0.475
Perceptual	-0.56	4.02	0.633
Shapes	-4.45	2.42	0.111
Mechanical	-0.77	2.60	0.313
ST-1	0.38	9.53	0.889
ST-2	2.57	4.97	0.103
ST-3	-0.59	2.24	0.420
ST-4	-2.41	4.22	0.018*
ST-5	3.11	3.49	0.005*
ST-6	-3.32	9.72	0.228

* P = .05 ** P = .01

On Speed Test 4, examining initiative and outward confidence, the ethnic group performs at a slightly lower level. This may be explicable in social rather than psychometric terms. On Speed Test 5, a measure of manual speed, the ethnic group performs at a significantly higher level. However, standard error on both these tests would suggest the differences are not of much practical significance.

Testing members of different religious communities

143 fifthformers were tested at two schools in the South of England. The catchment area was very similar, but one of the schools had a 90% Roman Catholic intake, and the other, which was close by, contained very few pupils who did not describe themselves as C of E. Although the pupils were not asked to identify themselves by religious belief, it has been assumed for the purposes of this study that one group was Catholic and one was not.

T-tests were used to establish the level of discrimination between the means of the groups on the separate tests, and a discriminant analysis was performed on the whole profile, excluding the manual dexterity measures, to examine the differences between the groups in the shape of the whole profile.

Table 4.51

School 1 (Non-RC) 62 pupils: 35 male: 27 female (i.e. higher proportion of males)
School 2 (RC) 83 pupils: 41 male: 42 female

Test	Difference of Means	Significance of Difference
CS	-0.79	0.557
Verbal	-0.22	0.861
Numerical	-2.02	0.153
Perceptual	-0.67	0.709
Shapes	2.85	0.337
Mechanical	1.96	0.025*
ST-1	-2.65	0.500
ST-2	1.8	0.265
ST-3	-0.24	0.794
ST-4	-0.45	0.734
ST-5	5.34	0.0005***
ST-6	1.30	0.592

*** Pr < .001 ** Pr < .01 * Pr < .05

Positive mean difference scores are in favour of the Non-RC school; negative scores in favour of the RC school.

Apart from the results on the Mechanical Ability test and the Manual Speed test (ST5), which both showed a higher mean score for the Non-RC school, there appear to be no significant differences between the two schools on any of the ability or personality tests. The higher Mechanical score is probably accounted for by the fact that there was a larger proportion of boys to girls in the non-RC school, and male mean scores on mechanical ability tasks and tests have been shown by most researchers to be appreciably higher than female mean scores.

Results relating to the whole profile (ability and personality measures)

Canonical correlation coefficient (eta) = .2942

Significance = .253

It can be seen that the percentage correctly classified did not reach a significant level, and it can be assumed that the tests do not produce a "characteristic" profile for either group.

A further study of religious differences was carried out in N. Ireland, although the small size of the sample and the unequal numbers would suggest that this can give at best only confirmatory experimental evidence.

36 volunteer students took the MP, and were asked to describe their religious group. One student preferred not to be identified by religious grouping, so was dropped from the study. Independent sample T-tests carried out on the individual tests showed no

significant differences between the groups except on the Numerical test (GATN) where the difference favoured the Catholic group ($p<0.15$) (SEdiff 2.18).

Table 4.52
N = 35
Catholic = 27
Protestant = 9

Test	Mean difference	Significance of Difference
CST	2.73	.245
GATV	2.88	.205
GATN	5.83	.017*
GATP	2.01	.382
SHAPES	-2.38	.688
MAT	-0.43	.824
ST1	10.05	.316
ST2	-1.16	.671
ST3	2.33	.168
ST4	-0.99	.719
ST5	-4.54	.092
ST6	0.03	.053

Positive mean differences are in favour of the Catholic group; negative mean difference in favour of the Protestant group.

* $P < .05$

External assessment of the Morrisby Profile

Rewriting and standardising the Morrisby Profile; assessing its reliability and establishing its validity and fairness took place between 1991 and 1992. Norm tables from the standardisation were computed for a general population group, for various age bands, for graduates and managers and for technical groups.

Once the battery had been rewritten and assessed, it was submitted for review to the BPS Review of Tests used in Vocational Guidance, and it received a favourable report. The assessing body recommended the battery for use in both Employment training and Youth Training, and gave it acceptable or above acceptable ratings in all areas. The criteria for this favourable assessment appeared to be the numbers in the validity studies and the sizes of the correlation coefficients with the criteria.

However, this assessment did not take into account either the battery's ability to provide useful information for selection, development or guidance decisions on the basis of the combinations of test scores, or its capacity to operate as a truly differential battery. These were apparently not regarded by the reviewers as relevant criteria in the context of vocational guidance, and instead the batteries assessed in the Review were regarded either as composites or as collections of separate tests, each of which required separate correlational validation.

It is the contention of this study that such an assessment does not really address the requirement that a battery should, as a whole, predict occupational membership and performance, not by summing scores and correlating the sum with a criterion, but by measuring both levels and profile patterns of scores against relevant criteria.

Morrisby's detailed clinical interpretation of his tests; the profiling, work styling, planning, problem solving and so on, still relied on a large element of faith and very little evidence beyond anecdote and questionnaire "client satisfaction" data, which could not really establish the efficacy of the differential approach. An index of differential efficiency was needed, and this will be discussed in Chapter 5.

Chapter 5. Validating the differential use of the Morrisby Profile

As has been seen in the previous chapter, it is a relatively simple matter to establish an index for validating the Morrisby Profile, or any other differential battery, as a traditional measure, by performing multiple regressions on the subtests using relevant criteria as dependent variables. Alternatively, correlations with separate criteria for the individual subtests can be used to show their construct and/or predictive validity. However, this only shows its efficacy as a traditional battery; it neither proves nor disproves its efficacy as a differential battery; that is to say, its ability to predict performance or group differences from candidate specific inter-test relationships.

It is less easy to provide validation evidence for the predictive power of patterns of scores than for single tests or composites, largely because simple correlational methods are no longer appropriate, but also because quantifiable criteria against which the groupings or patterns may be measured tend to be rather poor. It is true that correlational methods could still be used for purely confirmatory studies, but in exploratory studies the number of potential pairings precludes the use of such methods.

There are two aspects to the validation of the differential approach; showing the effectiveness of the differential treatment of score groups, and providing an index of differential validity against which such a battery may be measured. Four questions therefore have had to be addressed in this study:

1. Can the differential battery predict group differences, using profile patterns and levels of scores?

2. Can a purely differential approach, examining score differences, predict group

membership or performance?

3. By what standards are we to assess the performance of a battery which takes the differential approach?

4. If differentials are seen to add usefully to the process of selection, development or guidance, is it necessary for them also to outperform score levels in predicting performance?

The first two of these questions will, it is hoped, be addressed by this chapter, and further considered, with the rest of the questions, in the discussion which follows in the final chapter.

Can the differential battery predict group differences?

The prediction of group differences in the traditional way, using multiple regression coefficients was addressed in the preceding chapter, in the examination of regression coefficients using various criteria of performance. However, it is appropriate at this point to examine alternative ways of classifying individuals into groups, and to take a further look at the implications of the purely differential approach.

Discriminant analysis has already been referred to as the method employed by French (1955) in order to assess the ability of a test battery to find typical profiles of scores for an occupational group. It is particularly appropriate in this context, as it is designed for situations where the dependent variable is categorical rather than linear. As a validating technique for the Morrisby Profile, which is frequently used in vocational guidance and job development, it is certainly useful, as it is able to show the ability of the battery to distinguish between occupational groups with a high degree of accuracy, and to generate typical profiles of group members.

There are clearly limitations to this approach, in that generation of typical profiles risks cloning majority membership of the group. In managerial selection, for example, white heterosexual middle class males probably make up the majority of managers, and may well generate a typical profile which differs from those in other groups without necessarily being an ideal managerial profile. Shared characteristics among a group do not necessarily distinguish the highest performers. Nevertheless, the balance of probabilities suggests that where two or more occupational groups are compared, the distinguishing factors are more likely than not to be those which make success in one group more probable than in another, and close examination of the weighting coefficients usually shows this to be the case.

For example, a comparison of accountants and engineers shows spatial ability and numerical ability to be particularly important factors in determining group membership, in the expected directions. It seems counter-intuitive and perverse, although logically possible, to claim that the possession of numerical ability characterises the majority of accountants but that its lack would single out exceptionally high flyers.

In any case, it is not suggested that typical profiles, once generated, should take on ideal status and be recommended as templates for future selection or guidance decisions. Using discriminant analysis in this way allows a differential battery to be examined for concurrent validity, rather than using it to create rigid moulds.

Although predicting occupational group membership has its limitations, the groups used as classifiers in discriminant analysis need not just be occupational divisions; grouping variables may be success in management, course completion, promotion rating and so on. This allows more flexibility to the technique, which bears more relation to multiple regression, in that it should be possible to use the battery to devise useful equations for future selection and promotion, based on weighted coefficients similar to the beta weights of regression. In clinical psychology this technique has been used to distinguish patient groups; e.g. schizophrenic/non-schizophrenics (Newmark 1980).

Discriminant function is also a useful technique to employ when determining the differences, if any, between minority and ethnic groups and the majority; t-tests, which are commonly used to indicate test bias, only operate on single tests, whereas discriminant analysis can use membership of ethnic group as the dependent variable, and so establish whether a battery as a whole substantially discriminates between the groups even if the individual tests do not reach any level of significance in distinguishing their performance.

In discriminant analysis, discriminant scores are computed for each case in order to predict its group. These scores are obtained by finding linear combinations of independent variables, multiplying each variable by a constant and then summing the products. However, it does not take account of any non-linear relationships between dependent and independent variables which are characteristic of differential or profiling theory, and therefore is not presented as an answer to the problem of finding a truly differential index of efficiency, but as a halfway house between wholly traditional and wholly differential methods. While profiling is not wholly unrelated to scalar scoring, it may better be regarded as an approximation of a multiple regression or discriminant function analysis which includes negative beta weights and in which the magnitude of the beta weights varies in a curvilinear manner.

Four sets of subjects were used to examine the validity of the battery using discriminant analysis. Sets A and B were distinguished by their occupations : Set A consisted of technicians, managers and careers guidance specialists, and Set B consisted of engineers and teachers. For both these sets, occupation was the grouping variable. Set C consisted of sixthformers who had taken GCSE the previous year, and grades in Maths GCSE were used as grouping variables. Set D was a large collection (13000) of candidates for vocational guidance whose scores on an interest inventory were available as criteria.

Discriminant function analysis A

Set A. Guidance officers, Technicians and Managers. (N = 332)

Subjects:

N = 107

19 guidance officers

44 technicians

44 managers.

Age range 22 - 54

M = 52

F = 65

Materials:

Morrisby Profile

Occupational labels

Method

Three volunteer occupational groups were administered the Morrisby Profile, on the understanding that they would receive printed reports describing the results. A discriminant analysis was performed on the group, using occupation as the discriminating variable, with the standardised scores of the Morrisby Profile as the predictor variables.

Results

Group membership was predicted at a significant level (canonical correlation coefficient 0.67; $p = .00005$)

Table 5.1

N = 107

Actual Group	No of Cases	Predicted Group Membership 1	2	3
Group 1 Guidance	19	10	2	7
		52.6%	10.5%	36.8%
Group 2 Technicians	44	0	37	7
		0	84.1%	15.9%
Group 3 Managers	44	4	8	32
		9.1%	18.2%	72.7%

73.83% of grouped cases correctly classified.

Table 5.2

Function	Canonical Correlation	Wilks' Lambda	Chisquare	DF	Sig
1	.67	.43	82.187	24	.00005
2	.45	.80	22.49	11	.0209

The battery predicts group membership accurately nearly 75% of the time ($p < .00005$), which, given that there are three groups, is very much better than chance, which would be 0.33% of the time.

Table 5.3

Correlations between rotated canonical discriminant functions and discriminating variables:

	FUNCTION 1	FUNCTION 2
CST	-.20	.17
VERBAL	-.09	.55
NUMERICAL	-.02	.40
PERCEPTUAL	.05	-.01
SHAPES	.56	.05
MECHANICAL	.58	.28
ST1	-.23	-.08
ST2	.20	-.08
ST3	-.13	.41
ST4	-.04	.47
ST5	-.11	.19
ST6	.06	.34

The number of negative weightings, although none is particularly large, shows the importance of score differences, although using this technique the differential nature of the battery cannot be fully explored. What matters in differential testing is the fact that an important strength, such as, for example, numerical ability for an accountant, is greater than an unimportant strength, such as mechanical ability. A linear approach would over-emphasise the actual size of the score differences, so that mechanical ability would be seen, ipso facto, as militating against accounting success, whereas in fact it is only in those rare cases that mechanical ability dominates the profile that it could have such an effect. This is likely to be so rare that the beta weight would reflect the usual intercorrelation between numerical and mechanical ability and so come out at zero. Usually the intercorrelation between numerical and mechanical ability will ensure that mechanical ability correlates positively with success.

It would appear that the first function loads most heavily on spatial and mechanical ability, whereas the second function is most concerned with verbal ability, confidence and commitment. This second function appears to distinguish the Guidance Officers, whereas the first function strongly differentiates them from the technicians. The managers somewhat muddy the water, in that they would naturally share characteristics of both groups, without necessarily having particular areas of expertise to distinguish them.

However, it is evident from these results that the battery can distinguish occupational groups on the basis of the relative levels of the relevant subtests.

Discriminant function analysis B

Set B

Subjects:

(N = 449)

344 engineers

105 teachers

Materials:

Morrisby Profile

Occupational group labels

Method

The method employed was the same as in the previous study. Using the standardised scores of the Morrisby Profile as predictor variables, as above, and occupation as the grouping variable, the following figures were obtained:

Results

Again, it was possible to distinguish significantly between the two occupational groups, on the basis of the Morrisby scores (canonical correlation coefficient 0 .54; p = .00001) The groups were correctly classified 89% of the time, which is very considerably greater than a chance classification. However, gender data was not available for these subjects, and it may well be that the likely preponderance of males in the engineering groups assisted in classifying the individuals along gender rather than wholly occupational lines. However, it is unlikely that this would have altered the results significantly, as the variables most likely to distinguish the gender groups (MAT and Spatial ability) are also those likely to be of particular importance in differentiating between the engineers and teachers in reality.

Table 5.4

Canonical Discriminant Functions

Fcn	Pct of Variance	Cum Pct	Canonical Corr		After Fcn	Wilks' Lambda	Chisquare	DF	Sig
				:	0	.54	270.34	12	.00001
1*	100.00	100.00	.68	:					

Table 5.5
Classification Results -

Actual Group		No. of Cases	Predicted Group Membership 1	2
Group Engineers	1	344	330 95.9%	14 4.1%
Group Teachers	2	105	35 33.3%	70 66.7%

Percent of "grouped" cases correctly classified: 89.09%

It can therefore be seen that, with more clearly differentiated samples, good discrimination can be obtained. In the above example, 50% might be expected to be classified correctly by chance. In fact, almost 90% were correctly classified, and a canonical correlation of 0.68 ($p < .0005$) was obtained.

It appears therefore that, with the three occupational groups already described, which were deliberately chosen so as not to be too easily classified by this method, as the mean ability levels are not very different for this all graduate/managerial sample, the battery of tests is able to distinguish groups at a level very significantly greater than chance, even given the probable gender imbalance.

Used in this way, discriminant analysis is still vulnerable to charges that it only isolates typical features of an occupational group, rather than distinguishing between higher and lower performing groups (French op. cit).

It was therefore decided to perform a further discriminant analysis using more performance-linked grouping variables.

Discriminant function analysis C

Set C: Sixth form students (N = 517)

Subjects:

517 Year 12 students

M = 296

F = 221

Materials:

Morrisby Profile

Maths GCSE results

Method:

Using sixth form students, a discriminant function analysis was performed, using Maths GCSE grades as the discriminating variable, and the standardised scores of the Morrisby Profile as the predictor variables. The discriminating variable was recoded into 4 values: Grade A (Group 4) Grade B (Group 3) Grade CDE (Group 2) and Grade FG or below (Group 1).

Results:

The Morrisby Profile scores correctly classified the various students into their various performance groups 64% of the time, which was at a level significantly better than chance (canonical correlation coefficient 0.62; p = .00005)

Table 5.6

N = 517
Grouping variable = Maths GCSE

Canonical Discriminant Functions

Fcn	Cum Pct	Canonical Corr		After Fcn	Wilks' Lambda	Chisquare	DF	Sig
				0	.60	257.25	36	.00005
1*	95.20	.62	:	1	.97	15.48	22	.8407
2*	98.28	.14	:	2	.99	5.57	10	.8498
3*	100.00	.11	:					

Table 5.7

Classification Results -

Actual Group		No. of Cases	Predicted Group Membership 1	2	3	4
Group FGFU	1	5	2 40.0%	3 60.0%	0 .0%	0 .0%
Group CDE	2	249	1 .4%	203 81.5%	2 .8%	43 17.3%
Group B	3	107	0 .0%	65 60.7%	2 1.9%	40 37.4%
Group A	4	156	0 .0%	39 25.0%	1 .6%	116 74.4%

Percent of "grouped" cases correctly classified: 62.48%

In view of the fact that there were four groups, and the size of each group was included in the analysis using the SPSS (PRIORS SIZE) command, chance performance would be put at 25%, so a classification rate of 63% is satisfactory. The canonical correlation of 0.62 is also satisfactory. Wilks' lambda is significant ($p<.00005$) which suggests that the group means on the discriminant function differ .

Discriminant function analysis D

Set D: Vocational Guidance candidates (N = 13000)

M = 7468

F = 5532

A much larger sample of mixed candidates for vocational guidance, ages ranging from 15 to 58, mean age 26.3, was used in the final study in this section. The grouping variable used was taken from their score on an interest inventory, which classified individuals according to their strongest interest taken from People, Data and Things (Practical /Technical/ Scientific). There were some misgivings as to the usefulness of the criterion, as so many candidates for vocational guidance complete such interest inventories virtually at random, and in any case expressions of interest in careers may not entirely match the ability to fulfil those aspirations. Another factor, partly controlled for in this study, was the well documented stereotypic effect of gender on career aspirations (Hammond and Dingley 1989), which could well have encouraged some girls to avoid the scientific and technical careers, despite their aptitudes, and some boys to avoid the caring professions.

Subjects

13000 Vocational Guidance candidates

7468 males

5532 females

Materials

Morrisby Profile results

MVQ interest scores

Method:

13000 candidates for vocational guidance, who had also completed the MVQ (Morrisby Vocational Questionnaire) supplied their Morrisby Profile and MVQ scores. The majority of the subjects were 16-18 year old school students. 7468 males and 5532 females took part in the study. Discriminant analyses were performed on both the male and female subjects separately, to control for gender bias, using Interest in Things (Practical/ Technical/Scientific) as the discriminating variable.

Results

In all cases, the battery was able significantly to predict interest in Things from the ten variables of the battery. In all cases it produced significant canonical correlation coefficients. With the whole dataset, the canonical coefficient was 0.4 ($p = .00005$); with males it was 0.38 ($p = .00005$) and with females it was 0.2 ($p = .00005$).

In a discriminant function analysis performed with the whole group, the battery was able accurately to classify 62% of the group according to the level of their interest in Things.

With the males only, it classified 56% correctly, and with the females, 75%. However, this result was almost certainly affected by the very large number of females expressing a low level of interest in Things, and, in reality, the battery performed better with the male subjects.

Discriminant function analysis D1

Grouping variable - Things (Practical/Technical/Scientific)
Subjects - Whole dataset (13000)

Table 5.8

Fcn	Pct of Variance	Cum Pct	Canonical Corr	:	After Fcn	Wilks' Lambda	Chisquare	DF	Sig
				:	0	.84	2293.50	24	.00005
1	96.73	96.73	.40	:	1	.99	81.38	11	.00005
2	3.27	100.00	.08	:					

Table 5.9

Classification Results -

Actual Group		No. of Cases	Predicted Group Membership 1	2	3
Group	1	7070	5910 83.6%	1150 16.3%	10 .1%
Group	2	4890	2809 57.4%	2045 41.8%	36 .7%
Group	3	1016	320 31.5%	661 65.1%	35 3.4%

Percent of "grouped" cases correctly classified: 61.58%

Discriminant function analysis D2

Grouping variable - Things (Practical/Technical/Scientific)
Subjects - 7468 Males

Table 5.10

Canonical Discriminant Functions

Fcn	Pct of Variance	Cum Pct	Canonical Corr	:	After Fcn	Wilks' Lambda	Chisquare	DF	Sig
				:	0	.85	1203.63	24	.00005
1*	95.67	95.67	.38	:	1	.99	55.98	11	.00005
2*	4.33	100.00	.09	:					

Table 5.11

Classification Results -

Actual Group		No. of Cases	Predicted Group Membership 1	2	3
Group	1	2949	1521 51.6%	1420 48.2%	8 .3%
Group	2	3563	906 25.4%	2628 73.8%	29 .8%
Group	3	947	83 8.8%	832 87.9%	32 3.4%

Percent of "grouped" cases correctly classified: 56.05%

Discriminant function analysis D3

Grouping variable - Things (Practical/Technical/Scientific)
Subjects - 5532 Females

Table 5.12

Canonical Discriminant Functions

Fcn	Pct of Variance	Cum Pct	Canonical Corr	:	After Fcn	Wilks' Lambda	Chisquare	DF	Sig
				:	0	.96	240.91	24	.00005
1*	94.95	94.95	.20	:	1	.99	12.39	11	.33500
2*	5.05	100.00	.05	:					

Table 5.13

Classification Results -

Actual Group	No. of Cases	Predicted Group Membership 1	2	3
Group 1	4121	4103 99.6%	13 .3%	5 .1%
Group 2	1327	1315 99.1%	12 .9%	0 .0%
Group 3	69	67 97.1%	0 .0%	2 2.9%

Percent of "grouped" cases correctly classified: 74.62%

Results:

These tables all show the ability of the battery to classify individuals according to group membership, whether that membership be "typical" occupational grouping, academic prowess, vocational interests or performance related work ratings. However, at this point in the study, the classification is still based on discriminant scores culled from group means, rather than any differences between the scores. Paired score differences as predictors will be examined fully in the next section.

The differential approach is based on the predictive power of groupings and patterns of scores and score differences, rather than on that of a single, or single composite, score. This is the concept underpinning the multidimensional, rather than the unitary, structure of abilities. It assumes that there are several aspects of intelligence, and that they operate both separately and in various relationships to predict performance differentially on different criteria. For example, in certain fields it is preferable to have a preponderance of practical intelligence; in others, information processing ("knowledge-acquisition") components are more useful and, within those components, verbal and numerical abilities should be the greater.

Sternberg suggested that there are several possible relationships between an individual's various intellectual abilities, some of which might produce identical success rates in very different ways. "Tentatively, one might venture that grammar-intensive and learning from context methods may work better for those with higher general intelligence but only average abilities specialised for language learning; mimicry and memorisation methods may work better for those with higher specialised language-learning abilities but only average general intelligence." (Sternberg 1985)

Because of the difficulty of establishing the boundaries between "higher" and "average" in Sternberg's terms, and of the lack of knowledge of the various teaching methods employed with the individual students, the central proposition here cannot be adequately tested. In any case, this seems to describe a curvilinear rather than a linear approach in line with truly differential assessment which will be directly addressed in the next section.

However, the idea that some forms of intelligence are more valuable than others in predicting particular types of performance is well in line with the theories of differential intelligence. (Brogden 1951; Johnson, Zeidner & Scholarios 1990).

On the basis of these discriminant analysis studies, as well as on the multiple regression studies described in the previous chapter, the evidence suggests, therefore, that the Morrisby Profile can be used to distinguish occupational groups and high/low academic performers with a high degree of accuracy. It would appear that they can also be used to distinguish between high and low performers in the workplace, although the datasets used were too small for this evidence to be conclusive.

There is some evidence that they are significantly associated with vocational interests and personality attributes, as well as with ability, which would make them a useful tool in guidance as well as in selection and development.

However, showing that the battery of tests can differentiate between groups using a linear technique does not make a case for non-linear differentiation. This will be addressed in the next section, in an attempt to answer the second of the four questions posed earlier.

Can a purely differential approach, using differences between scores, predict group differences?

There are several ways in which score differences can be used to predict performance. The two most commonly examined are direct differences between paired criterion correlations, the approach taken by Brogden and Horst, and paired score differences which are then correlated with a single criterion, which was the direction taken by John Morrisby. In addition, contiguous differences may be analysed by means of multivariate analysis of variance so as to examine the relationship of profiles of scores to a criterion. A further method, using deviation scores, which examines profile patterns without any reference to levels of scores, will also be described in this section.

Paul Horst (1954) assumed the purpose of a differential battery to be the classification of candidates so as to be able to predict their relative chances of success in a wide field of activities. In other words, the intention should be to predict a person's good performance on one criterion and poorer performance on another. His index of differential efficiency was taken to be the average of the variances of the predicted difference scores for "all possible pairs of criterion variables".

This approach stresses the multidimensional nature of abilities, accepting that different combinations of scores will be more or less appropriate for different jobs. However, it does not take into account the actual predictive power of the differences between scores, or predictor variables, on single or multiple criteria; nor does it allow for any examination of the pattern of scores irrespective of levels of achievement.

If the tests are acting as a differential battery in Horst's sense, they should be able to distinguish between those performing better on one criterion and less well on another, and vice versa. The "differences" here would be the differences between the criterion variables; not the differences between the test scores. However, there is a further sense in which the battery may claim to be a differential battery, and that is in the sense in which actual differences between the scores may be predictive of performance, either irrespective of level or in addition to it.

The Morrisby Differential Coefficient

This question was introduced in the third chapter, in the examination of the differential coefficient derived by John Morrisby from the Horst equation. The Horst equation theoretically can be used to establish a single index of differential efficiency from a full battery of tests, but the matrices of inter-correlations and co-variances required to establish the index would be so unwieldy, and the opportunity for imperfect criteria to contaminate the final figure so large as to make this a tool suitable for simulated data only.

The Morrisby differential and summative coefficients, already described, are designed to examine only pairs of variables against a single criterion.

Summation Coefficient $= \frac{r_{o1} + r_{o2}}{\sqrt{2(1 + r_{12})}}$

Differential Coefficient $= \frac{r_{o1} - r_{o2}}{\sqrt{2(1 - r_{12})}}$

Multiple Coefficient $= R_{o.12} = \sqrt{(sum.coef.)^2 + (dif.coef.)^2}$

where:

r_{o1} = the correlation of Test 1 with the criterion
r_{o2} = the correlation of Test 2 with the criterion
r_{12} = the intercorrelation of the tests

Morrisby used these coefficients to establish correlations between score differences and the criterion of GCE French, and found differential coefficients as high as -0.56 with some pairs of tests. In addition, he suggested, although he did not illustrate, the use of the average of the summative, differential and multiple coefficients to establish the average summative, differential and multiple coefficients for the whole battery.

Two studies have been undertaken using the technique of Morrisby's differential and summative coefficients. The first study used GCSE students, with GCSE results for English language, Physics, Mathematics and French as the criteria. The second study used insurance salespeople, with success in sales as the criterion.

Morrisby Coefficients - Study A1

Subjects:

648 Year 12 students (16-17 year olds)

M = 335

F = 313

Materials:

Morrisby Profile

GCSE results for the cohort in English Language, Mathematics, French and Physics.

Method:

In an attempt to replicate John Morrisby's conclusions regarding the efficacy of the differential coefficient, 648 Year 12 students (16/17 year olds) were administered Morrisby Profile in 1992. Using the traditional technique of multiple regression, the results of the first six ability tests were then correlated first with their GCSE results, which were available for this group, and then with their English language separate result. The GCSE criterion was established by adding points for subject passes on a scale of 1 to 9, 1 representing the lowest grade and 9 the highest, the whole forming a rough indication of both quality and quantity of examination passes. The English language criterion used the grades given for the subject (1 to 9). All correlations with this criterion were therefore in reality negative, as 1 is the highest grade, but the signs have been reversed below for convenience of interpretation. It should be noted that this approach assumes that number as well as grade of GCSE passes is prima facie evidence of quality of performance. Clearly, if the sample contains many students from schools which encourage poor students to aim at a large number of low grade GCSE passes, whereas good students are encouraged to concentrate their efforts on achieving a few high grades, the premises of this study will be invalid.

It was to be expected that a summative coefficient would be more relevant to a criterion made up of a composite examination score, such as GCSE, but it was hoped to show that, with a more specific, differentiating criterion such as English language, the differential approach would be shown to have some value, although the subject is too knowledge based and affected by such aspects as teaching style and examination taking proficiency to be regarded as an ideal criterion, and probably the coefficients given below underestimate the reality.

Results:

With GCSE as the criterion, the summative coefficient performed better in all cases, as expected, than the differential, although the multiple coefficient, using both coefficients together, often proved of significantly greater value than either. However, using English as the criterion, the differential coefficient outperformed the summative in the case of five test pairings, highlighted in the table, despite the limitations of this choice of criterion.

The differential coefficient increased markedly when both test correlations with the criterion differed sharply, and the intercorrelations were fairly low. Summative coefficients were highest when both criterion correlations were high and test intercorrelations were low.

The highest summative coefficient, 0.43, occurred with GCSE as the criterion, between GATV and GATN. This pair has almost the highest criterion correlations, and the intercorrelation is a little lower than that between the highest pair (GATV and GATP).

The highest differential coefficient (GATV and MAT), with English as the coefficient, as might be expected from the Mollenkopf research, occurred when the intercorrelation was low (0.25) and one test (GATV) correlated with the criterion much better(0.4) than the other (0.03). It may be seen that, with GCSE as the criterion, the same tests performed best in the differential condition, and under very similar circumstances.

Table 5.14

Test intercorrelations

N = 648

	CS	GATV	GATN	GATP	SH	MAT
CS	1.0					
GATV	.51***	1.0				
GATN	.60***	.55***	1.0			
GATP	.70***	.58***	.68***	1.0		
SH	.47***	.34***	.42***	.53***	1.0	
MAT	.30***	.25***	.32***	.35***	.55***	1.0
GCSE	.28***	.44***	.21**	.28***	.20**	.15
Eng	.22**	.40***	.31***	.27***	.14	.03

* = p < .05 ** = p < .01 *** = p < .001

Table 5.15

Criterion GCSE

TEST PAIRS	SUMMATIVE 'R'	DIFFERENTIAL 'R'	MULTIPLE 'R'
CS + GATV	0.42***	0.16*	0.45***
CS + GATN	0.33***	0.03	0.33***
CS + GATP	0.34***	0.04	0.34***
CS + SH	0.28***	0.08	0.29***
CS + MAT	0.27***	0.11	0.29***
GATV + GATN	0.43***	0.14	0.45***
GATV + GATP	0.40***	0.17*	0.44***
GATV + SH	0.39***	0.21**	0.44***
GATV + MAT	0.37***	0.24**	0.44***
GATN + GATP	0.32***	0.03	0.32***
GATN + SH	0.30***	0.10	0.32***
GATN + MAT	0.28***	0.14	0.31***
GATP + SH	0.27***	0.08	0.50***
GATP + MAT	0.26***	0.11	0.28***
SH + MAT	0.20**	0.05	0.21**

Average summative coefficient 0.35***
Average differential coefficient 0.11
Average multiple coefficient 0.39

* = p < .05 ** = p < .01 *** = p < .001

Table 5.16

Criterion ENGLISH LANGUAGE

TEST PAIRS	SUMMATIVE 'R'	DIFFERENTIAL 'R'	MULTIPLE 'R'
CS + GATV	0.36***	0.18*	0.40***
CS + GATN	0.30***	0.10	0.32***
CS + GATP	0.27***	0.07	0.28***
CS + SH	0.21**	0.08	0.22**
CS + MAT	**0.16***	**0.16***	**0.23****
GATV + GATN	0.40***	0.09	0.41***
GATV + GATP	0.38***	0.14	0.40***
GATV + SH	0.33***	0.23**	0.45***
GATV + MAT	**0.27*****	**0.30*****	**0.40*****
GATN + GATP	0.32***	0.05	0.32***
GATN + SH	0.27***	0.16*	0.31***
GATN + MAT	**0.21****	**0.24****	**0.32*****
GATP + SH	0.24**	0.13	0.27***
GATP + MAT	**0.18***	**0.21****	**0.28*****
SH + MAT	**0.09**	**0.12**	**0.15**

*=p<.05 **=p<.01 ***=p<001

Average summative coefficient 0.31

Average differential coefficient 0.15

Average multiple coefficient 0.32

Morrisby Coefficients - Study A2

As has already been discussed in Chapter 3, the Morrisby Profile includes the assumption that there are broadly three components of intellectual functioning; abstract reasoning, or the capacity to comprehend new complex ideas; information processing, using words, numbers or visual representations, and practical intelligence, using spatial and mechanical abilities. Although these three "blocks" of intelligence; reasoning, information processing and practicality, may themselves be subdivided, an individual's strengths in terms of her "blocks" may give useful information as to her preferred methods of solving intellectual problems.

In view of the logic of the theory of abilities discussed in this section, the process was then repeated with the three "block" mean scores for abstract reasoning, information processing and practical ability.

Subjects:

M = 335

F = 313

Materials:

Morrisby Profile

GCSE results for the cohort in English Language, Mathematics, French and Physics.

Method

Differential, summative and multiple correlations were computed for the three "blocks" as predictor variables, using first GCSE and then English Language as criteria. The intercorrelations may be found in Table 5.17, and the results in tables 5.18 and 5.19

Results

Table 5.17 Block intercorrelation

	Reasoning	Info Pro	Practical
Reasoning	1.00		
Info. Pro	.71	1.00	
Practical	.28***	.20**	1.00
GCSE	.28***	.40***	.20**
Eng.Lang.	.22**	.39***	.11

* = p < .05 ** = p < .01 *** = p < .001

Table 5.18

Criterion GCSE

	Summative	Differential	Multiple
Reasoning/Info Proc.	.38***	.16*	.40***
Reasoning/Practical	.30***	.07	.31***
Practical/Info Proc.	.37***	.24**	.45***

* = p < .05 ** = p < .01 *** = p < .001

Average Summative Coefficient: 0.35***
Average Differential Coefficient: 0.16*
Average Multiple Coefficient: 0.39***

Table 5.19

Criterion ENGLISH LANGUAGE

	Summative	Differential	Multiple
Reasoning/Info Proc.	.34***	.20**	.39***
Reasoning/Practical	.19*	.11	.22**
Practical/Info Proc.	.28***	.30***	.41***

Average Summative Coefficient: 0.27***
Average Differential Coefficient: 0.20**

Average Multiple Coefficient: 0.34***

* = p < .05 ** = p < .01 *** = p < .001

As would be expected, the best differential coefficient occurs when intercorrelations are relatively low and the difference between the criterion correlations is highest (Information Processing and Practical). The best summative coefficient occurs when intercorrelations are not too high (but not the lowest figure) and test/criterion correlations are both higher.

The logic of Morrisby's version of the differential position can be seen more clearly in these scores; practical ability is in itself unlikely to predict high GCSE scores, but it is quite possible that there is a true negative relationship between practical ability and GCSE attainment, which is being masked by the intercorrelation of practical ability and information processing, as the latter attribute clearly does play a part in GCSE success.

Morrisby Coefficients - Study B

Subjects:

28 insurance sales people

Materials:

Morrisby Profile

Sales figures for previous year

Method:

Summative, differential and multiple coefficients were calculated for all possible pairs of variables in the Morrisby Profile, excluding the manual dexterity measures. The Morrisby equations were used, as described earlier. The criterion used was SALES; which represented the sales figures from the previous year for each salesperson. The summative and differential and multiple coefficients were compared so as to establish which set of coefficients performed better overall in predicting sales figures.

Results

The multiple coefficients, as expected, outperformed both the summative and the differential coefficients. However, in contrast to the results of the previous study, the differential coefficients, overall, were higher than the summative coefficients. 26 of the differential coefficients, out of a possible 45, outperformed the summative coefficients, and 19 of the summative coefficients outperformed the summative coefficients.

The best performing summative pairs were Speed Tests 1 and 4 (0.52), Speed Test 1 and 3 (0.51), and Speed Test 1 with the Numerical Ability test (0.51). This suggests, reasonably enough, that high levels of personality scores and Numerical ability are valuable aptitudes in sales. The best performing differential pairs were Mechanical and Speed Test 4 (0.59), Mechanical and Speed Test 1 (0.53) and Mechanical and Numerical (0.53). This suggests that it is important for the personality measures and numerical ability to be higher in the profile than practical ability. Using the multiple coefficient allows both level and differences to be taken into account, and gives much higher coefficients - the best performing pairs are Mechanical and Speed Test 4 (0.59), Speed Tests 3 & 4 (0.54), Mechanical and Speed Test 1 (0.54) and Mechanical and Numerical (0.53).

Summative Correlation Coefficients

(The figures in bold type are those where the summative coefficient is greater than the differential coefficient).

Table 5.20

	CST	GATV	GATN	GATP	SH	MAT	ST1	ST2	ST3	ST4
CST										
GATV	.01									
GATN	.25	.30								
GATPL	.05	**.13**	**.33**							
SPAT.	**-.07**	.02	.27	.06						
MAT	**.32**	-.19	.06	-.13	-.22					
ST1	.28	**.36**	**.51***	**.38**	.29	.09				
ST2	**-.18**	-.09	.18	-.04	**-.19**	**-.42**	.20			
ST3	.13	**.24**	**.44**	**.30**	.12	-.08	**.51***	.05		
ST4	.29	**.34**	**.51***	**.37**	.29	.08	.52*	.23	**.52***	

* p < .05

Average summative coefficient: 0.238

(18 of the summative coefficients are greater than the differential coefficients)

Differential Correlation Coefficients

(The figures in bold type are those where the differential coefficient is greater than the summative coefficient).

Table 5.21

	CST	GATV	GATN	GATP	SH	MAT	ST1	ST2	ST3	ST4
CST										
GATV	**.12**									
GATN	**.37**	**.32**								
GATP	**.20**	.06	.29							
SPAT.	.01	**.10**	**.34**	**.15**						
MAT	.19	**.30**	**.53***	**.43**	.35					
ST1	**.39**	.31	-.04	.27	**.38**	.53				
ST2	.09	**.17**	**.41**	**-.23**	.09	.10	**.46***			
ST3	**.21**	.11	.15	.06	**.23**	**.41**	.16	**.25**		
ST4	**.36**	.32	-.03	.27	**.36**	**.59****	.01	**.40**	.15	

*** p < .05 ** p < .01**
Average differential coefficient: 0.251
Root mean square (which allows signs to be disregarded):
(24 of the differential coefficients are larger than the summative coefficients)

Table 5.22

Multiple Correlation Coefficients

(All the multiple correlation coefficients are greater than either the summative or the differential coefficients).

	CST	GATV	GATN	GATP	SH	MAT	ST1	ST2	ST3	ST4
CST										
GATV	.12									
GATN	.45*	.44								
GATP	.20	.14	.44							
SPAT.	.07	.10	.43	.16						
MAT	.37	.36	.53*	.45*	.42					
ST1	.48*	.47	.52*	.47*	.48*	.54*				
ST2	.20	.19	.45*	.24	.21	.43	.50			
ST3	.25	.26	.46*	.31	.26	.41	.53*	.25		
ST4	.46*	.46*	.51*	.46*	.46*	.59**	.52*	.46*	.54*	

* p = .05 ** p = .01

Average multiple coefficient = 0.36

The technique devised by John Morrisby clearly can effectively predict performance, particularly if both summative and differential data is utilised in the form of the multiple coefficient. However, its disadvantages are, first, that it is cumbersome to compute, even with present day statistical techniques, and, secondly, that it only deals with paired differences, rather than allowing the differences between other groupings of scores to be taken into account. It is also a substantially linear method which is necessarily dependent on correlations which, as has been seen, may be spurious.

Ipsative measurement using deviation scores

The technique of using deviation scores, described by Cattell (1944) as "ipsative measurement", may be applied to ability tests and be used with fewer limitations than in its more common application to personality measures. In this technique, level or elevation of scores is first partially and then entirely removed by the process of standardising the candidate's scores, mapping them in relation to the candidate's mean, which presents the scores as difference scores, and then restandardising them to take out all element of individual level of score. This eliminates both elevation and scatter, leaving only the shape of the profile, and is analogous to the process described by Thomson (1950) as "centring around persons."

This method allows new scores to be calculated on each test of the battery for each individual, showing each score as a function of the whole profile, centring around the candidate's own mean to provide a measure of her strengths and weaknesses in her own terms. As with all ipsative measures, strictly speaking this does not allow for direct comparisons to be made between candidates on the basis of the scores, as one individual's score of one standard deviation above the mean (his own mean) may bear no relation to another individual's apparently identical score, when the original means were quite different. However, despite this, the possibility of predicting performance on the basis of ipsative scores alone was examined in this study. The rationale for this has already been touched on, and relates to the spurious nature of linear correlations between the test scores and criterion performance. If size of test scores is affected by factors other than ability, and if pattern of scores is some indication of the direction and strength of ability, it might be possible to base predictions at least partially on ipsative scores.

Although no work has been done to the present author's knowledge on ipsative scaling of ability measures, there has been considerable debate about the feasibility and desirability of deriving validity coefficients relating to external criteria from ipsative data from personality questionnaires. Johnson, Wood & Blinkhorn (1989) made a strong case against the use of any of the standard statistical techniques with ipsative data. "Correlations..between ipsative scales are uninterpretable because the scales are mathematically interdependent...therefore any method which relies on the analysis of correlation matrices is also inadmissible. ...because of their mathematical properties, scale means are also uninterpretable." (Johnson et al. op.cit.)

However, in the same article, Johnson et al. agree that "it is certainly permissible to compare individuals in terms of score profiles or patterns, because then the absence of a common metric does not matter..obviously this is only non-trivial when the number of variables is three or more."

Hicks(1970) listed seven possibilities for reducing the effects of ipsativity.

1. Respondents only partially order item alternatives.
2. Scales have differing numbers of items
3. Not all alternatives ranked by respondents are scored
4. Scales are scored differently for respondents with different characteristics
5. Scored alternatives are differently weighted
6. One or more of the scales from the ipsative predictor set is deleted when data are analysed
7. The test contains normative sections

With the exception of (4), all these are common practice in the case of ability tests originally normative which have been rendered ipsative by the process of centring around the mean, not by using forced choice paired comparisons. It is true that, as with paired comparisons in the ipsatised version of ability tests the sum of the covariances obtained between a criterion and a set of ipsative scores will equal zero. However, it is not the case, as with the forced choice questionnaire, that high scores on one scale will force low scores on any other scale. This is only the case for the final or twelfth scale, by which time all the degrees of freedom have been used up, so that after scoring eleven of the twelve tests of the battery the twelfth score can be predicted.

In view of this, it would seem that the only corrective action that need be taken in applying traditional statistical techniques to this type of "ipsatised" data would be to subtract another degree of freedom before establishing significance levels. There seems little reason to apply the total ban advocated by Johnson et al, or by Kline (1993) in ignoring all validities calculated with such data.

A further objection that may be made to those who would claim a great divide between normative and ipsative measures is the illusory nature of the difference. The assumption that normative measures rely on a common metric may somewhat overstate the case. Individuals vary widely in their ability to perform well on psychometric tests. For example, Table 5.22 below shows the range of scores on the Morrisby Profile Numerical ability test for 156 school students who had all received the highest grade in their GCSE Maths examination.

Subjects:

156 Year 12 school students

Mean Age 16yrs 15 months (range 15-19)

M = 50

F = 106

Materials

Morrisby Profile - Numerical Ability scores (raw and "ipsatised")

GCSE Mathematics results.

Method.

517 Year 12 students took Morrisby Profile Numerical Ability Test and supplied their GCSE results for analysis. All those who obtained Grade A in Mathematics (N = 156) were included in this study.

The range and variance of scores in the Numerical Ability test, using both raw and "ipsatised" scores as described above, was calculated for the group. The intention was to examine the range of aptitude test results in groups achieving a similar level of performance, and to see if "ipsatised" scoring showed a narrower range, or explained any more of the variance than the raw scores.

Results.

Table 5.22

Descriptive statistics: using the raw score for the Morrisby Numerical Ability test

Mean	40.244	S.E. Mean	.509
Std Dev	6.362	Variance	40.469
Kurtosis	-.403	S.E. Kurt	.386
Skewness	-.015	S.E. Skew	.194
Range	32.000	Minimum	24.00
Maximum	56.00	Sum	6278.000

Table 5.23

Descriptive statistics: using the "Ipsatised" difference score for the Morrisby Numerical test

Mean	.104	S.E. Mean	.053
Std Dev	.656	Variance	.431
Kurtosis	-.513	S.E. Kurt	.386
Skewness	.081	S.E. Skew	.194
Range	3.103	Minimum	-1.41
Maximum	1.69	Sum	16.269

The Morrisby Numerical Ability measure has high construct and predictive validity - it correlates with GCSE Maths itself 0.55 ($p < .0001$)($N = 518$ - see chapter 3). Nevertheless, the range of value obtained by the highest performers in the raw score condition is very wide. The possible range of score obtainable is 0 - 64, with a mean of 40 and a standard deviation of 6.4 for this high achieving group. The actual range of scores obtained by the group, as may be seen in the table, is 24 to 56, which is 5.03 standard deviations; or ± 2½ standard deviations from the mean.

As may be seen from the tables, there is a considerable amount of unexplained variance (6.36 standard deviations) in these results. This may well reflect problems in the linear scale model used, and suggests a need for a different approach. If differential prediction can be seen to account for some of this unexplained variance, it must be seen as a useful tool in prediction.

In the ipsatised condition (Table 5.23), it may be seen that the range (3.103) is 4.72 standard deviations, or ± 2.36 standard deviations, which is appreciably less than the range for the raw scores (± 2.51 standard deviations). The index of skew is also positive, (0.81), showing that more of the candidates are scoring highly, as would be expected for a measure of high performers, whereas that for the raw score condition is negative, showing that slightly more are scoring low. (-0.15) The ipsatised condition also accounts for more of the variance, as the unexplained variance is only 14% of the range, whereas, in the raw score condition, it was actually greater than the max/min range, and 71% of the maximum score.

If difference scores can reduce unexplained variance, it would seem they are worth consideration as part of the predictive process.

If normative scores are subject to such a degree of subjective bias in the forms of nervousness, test-taking familiarity, speed of responses and so on, it is difficult to avoid the conclusion that high and low scores on such measures are not necessarily as precisely mapped onto normative conceptions as one would wish. The top percentile bands may indeed predict "better" criterion performance than the lower bands - but this is not necessarily so. If it is not so, then normative tests are as vulnerable as ipsative tests to the charge that conclusions based on statistics drawn from suspect means and standard deviations must lack methodological rigour. The spurious hierarchy of achievement mapped by normative tests encourages their constructors to make claims for accuracy that responsible constructors of ipsative tests explicitly forego, stressing the need for careful interpretation of scores, in the light of their relationship to the client or candidate's own standards.

Saville and Willson (1991) do not directly address the methodological arguments of Johnson et al in their response to the 1989 paper; instead they offer empirical evidence that in fact ipsative measures perform adequately in the field. This argument may carry more weight if it is accepted that the reason ipsative and normative measures in fact correlate with one another reasonably well, when traditional, correlational or means-based statistical measures are used, in that normative measures are less truly linear than they appear, whereas the ipsative nature of deviation scores should not obscure the real force of performance rated against a candidate's own mean.

In view of the special nature of normative, "ipsatised" measures and the doubt as to the force of arguments raised against the use of statistical techniques with such data, studies using both multiple regression and discriminant analysis with "ipsatised" and scaled scores will be presented and compared. The separate contribution of deviation scores and scale scores to group variance will also be assessed, using partial eta squared, and this will be described in more detail later in this chapter. Meanwhile, in order to assess the various contributions of normative (absolute levels) and deviation scores to predictive validity, and despite the comments of Johnson et al., the empirical approach advocated by Saville and Willson will be followed in a provisional bid to examine the evidence.

In order to examine whether or not the Morrisby Profile could act as a differential battery, either in Horst's sense or in the "purer" sense; that is, whether the tests could distinguish between those performing better on one criterion and less well on another, or whether they could distinguish high performers on a criterion by the use of score differences rather than score levels, a number of studies were set up, using multiple regression or discriminant function, depending on whether the dependent variables were linear or categorical in nature, and comparing the presentation of variables as raw scores or as deviation, "ipsatised" scores.

The role of discriminant analysis in predicting group differences from a set of predictor variables has already been discussed. The technique is not unlike that employed by multiple regression. A linear combination of the independent variables, weighted by coefficients derived from correlations between the values of the discriminant function and the variables, serves as the basis for assigning cases to groups. The information contained in the multiple independent variables is summarised in a single index, the discriminant score, which is obtained for each case by multiplying the unstandardised coefficients by the values of the variables, summing these products and adding the constant. When the mean of the discriminant scores is the same for all groups, there can be no discrimination. In most cases, therefore, the technique is used to distinguish on the basis of high or low scores on the relevant variables.

However, discriminant analysis can be also "forced" to some extent to ignore actual levels of scores by the technique of finding the distance of each score from the mean of all the subtests, and then finding the z-score of the result. As has already been described, this has the effect of standardising each candidate's score around his or her own mean, and then comparing the scores with other candidates only to the degree by which each exceeds or is exceeded by that mean - in other words, removing levels of scores and replacing them with patterns of scores.

Using this technique with discriminant function analysis is intended merely as an example of how the technique can be expected to work. It does not radically alter the concept of discriminant function, and could indeed be taken equally well with multiple regression - as will be shown later - and with several other statistical methods. However, the technique does attenuate the dataset by removing that aspect - levels - which is normally considered most important.

These deviation, "ipsatised" profiles can be used alone or with conventional scores as the predictor variables in order to compute group membership, and the resulting weights may then be used with a back sample to compute new scores and examine the effect of the criterion grouping variable on the back sample. The method has the disadvantage of reducing variability, as it assumes the standard deviations appropriate to the normal curve, which may not always be the case, but it does allow examination of the individual's direction of abilities without being distracted by the actual size of the score.

In order to examine the usefulness of this approach, the sample of managers, technicians and guidance officers described previously was utilised once more.

Deviation score study: DFA 1

Subjects

N = 107

19 guidance officers

44 technicians

44 managers.

Age range 22 - 54

M = 52: F = 65

Materials:

Morrisby Profile

Occupational labels

Method

The three occupational groups were administered the Morrisby Profile, and all twelve of their test scores were standardised as z-scores. The first six ability scores were summed and their mean calculated, and the mean of the second six speeded tests was also taken. The appropriate mean was then subtracted from each test score, so as to produce a score difference from the mean in standardised scores. These differences were then restandardised, so that all elements of elevation and scatter were removed from the scores, which now only represented a difference in the candidate's own terms from his or her own mean.

Results

Table 5.24

N = 329

Canonical Discriminant Functions

Fcn	Pct of Variance	Cum Pct	Canonical Corr		After Fcn	Wilks' Lambda	Chisquare	DF	Sig
:					0	.53	204.02	20	.00005
1*	92.18	92.18	.66	:	1	.94	20.37	9	.02
2*	7.82	100.00	.25	:					

Table 5.25

Classification Results -

Actual Group		No. of Cases	Predicted Group Membership 1	2	3
Group Guidance	1	74	42 56.8%	5 6.8%	27 36.5%
Group Technicians	2	132	2 1.5%	98 74.2%	32 24.2%
Group Managers	3	123	20 16.3%	43 35.0%	60 48.8%

Percent of "grouped" cases correctly classified: 60.79%

This may be compared with table 5.26, showing the same groups classified by the raw scores:

Table 5.26

Fcn	Pct of Variance	Cum Pct	Canonical Corr		After Fcn	Wilks' Lambda	Chisquare	DF	Sig
:					0	.49	230.64	24	.0000
1*	86.91	86.91	.67	:	1	.89	37.60	11	.0001
2*	13.09	100.00	.33	:					

Table 5.27

Classification Results -

Actual Group	No. of Cases	Predicted Group Membership 1	2	3
Group 1 Guidance	74	52	6	16
		70.3%	8.1%	21.6%
Group 2 Technicians	132	6	104	22
		4.5%	78.8%	16.7%
Group 3 Managers	123	26	37	60
		21.1%	30.1%	48.8%

Percent of "grouped" cases correctly classified: 65.65%

As may be seen from the above tables, (Tables 5,26 and 5.27), in the raw score condition, classification is correct 66% of the time (canonical correlation 0.67, p < .00005). In the "ipsatised" condition, it is correct 61% of the time, (canonical correlation 0.66, p <.00005). The differences do not appear greatly significant, and it is of particular interest that these three groups can be differentiated in this way, using a method owing nothing to score levels. Chance classification would be 0.33%, as there are three groups in the analysis.

Deviation score study: DFA 2

A further analysis was performed with the engineers and teachers previously described (N = 463).

Subjects:

344 engineers

205 teachers

Criterion:

Occupational membership

Method: as above

Results

In the following table, the "ipsative" approach was examined, as the score differences were restandardised to eliminate elevation. Again, the "ipsatised" approach (Tables 5.28 and 5.29) classified groups correctly at virtually as high a rate as the raw scores shown in table 5.30 and 5.31 for comparison.

Table 5.28 - "ipsatised"

Canonical Discriminant Functions

Fcn	Pct of Variance	Cum Pct	Canonical Corr		After Fcn	Wilks' Lambda	Chisquare	DF	Sig
				:	0	.59	243.11	10	.00005
1*	100.00	100.00	.64	:					

Table 5.29 - "ipsatised"

Classification Results - in the "ipsatised" condition

Actual Group		No. of Cases	Predicted Group Membership 1	2	
Group	Engineers		354	339	15
			95.8%	4.2%	
Group	Teachers		109	40	69
			36.7%	63.3%	

Percent of "grouped" cases correctly classified: 88.12%

Table 5.30 - raw scores

Canonical Discriminant Functions

Fcn	Pct of Variance	Cum Pct	Canonical Corr		After Fcn	Wilks' Lambda	Chisquare	DF	Sig
				:	0	.54	270.34	12	.00005
1*	100.00	100.00	.68	:					

Allowing for one less degree of freedom does not alter the significance level

Table 5.31 - raw scores

Classification Results - using the actual score levels

Actual Group		No. of Cases	Predicted Group Membership 1	2
Group	1	354	324 90.5%	30 9.5%
Group	2	109	23 21.1%	86 78.9%

Percent of "grouped" cases correctly classified: 88.9%

As may be seen from a comparison of these tables, in the raw score condition, prediction is 88.9 % correct (canonical correlation 0.68, p<.00005). In the "ipsatised" condition, prediction is 88.12% correct,(canonical correlation 0.64, p<.00005) The difference would appear to be so slight as to be insignificant. Allowing one less degree of freedom, as described earlier in this chapter to compensate for the "twelfth test" in the ipsatised approach, does not alter the significance level.

This shows that it is possible to use a totally differential approach when classifying occupational groups, while losing scarcely any classificatory efficacy. It should be stressed that there has been no intention in this study to show that the differential approach is "better" than the traditional approach in a predictive sense, but that its stress on factors other than absolute levels of scores make it an advantageous method to use in addition to traditional methods. This data centring method shows that having "more of X than of Y", irrespective of what other group members have, is of predictive importance, although it is probable that correlations will also exist between absolute scores and the criterion.

This ipsatising method may also be applied to multiple regression. This has the effect of measuring whether deviation scores alone, or differences from the mean, centred around the candidate's own mean, can predict achievement. Somewhat surprisingly, in view of the attainment-based criterion, it would seem that they can. This suggests that predicting success should mean taking into account a candidate's own pattern of strengths and weaknesses, rather than just her level of ability, or even her weighted level(s) of ability in the most relevant tests.

Deviation score study: multiple regression

The following studies were performed on the dataset of sixth form students described earlier.

Subjects:

648 Year 12 GCSE students

Dependent variables

GCSE grades in Maths, English Language, Physics and French

Independent variables:

The 12 tests of the Morrisby Profile.

Treatment A - traditional "raw score" treatment

Four multiple regression analyses were performed on the raw scores of the 12 tests of the Morrisby Profile (CST, Verbal, Numerical, Perceptual, Spatial, Mechanical, the four speeded personality measures, Manual Speed and Manual Skill). The dependent variables used were GCSE Maths, English Language, Physics and French, used as single criteria, in the traditional way, and then as "paired" criteria, following Horst's perception of differential theory, which focused on the ability of a battery to distinguish between performance on several criteria. This would be of particular use in the situation of a selector asked to classify recruits into different job areas, rather than simply selecting/rejecting on a single or composite criterion.

Full tables are given in Appendix A, but a summary may be found below:

Table 5.32

N = 396 (French)
N = 519 (English language)
N = 188 (Physics)
N = 517 (Maths)

Raw scores - single criteria (traditional approach)

Table 5.33

Criterion	Method	Multiple R
Maths	Raw scores	0.66
Physics	Raw scores	0.47
English Lang	Raw scores	0.50
French	Raw scores	0.54
Average multiple "R"		**0.54**
Average multiple "R" squared		**0.29**

N.B. The battery's capacity to predict performance on separate criteria is no indication of its differential capabilities, but is given in order that this traditional validation data may be compared with the other treatments described in this section.

Raw scores - paired criteria (Horst's differential approach)

Table 5.34

Criterion	Method	Multiple R
Maths/French	Raw scores	0.37
Physics/French	Raw scores	0.41
English/French	Raw scores	0.35
Maths/Physics	Raw scores	0.26
English/Physics	Raw scores	0.44
English/Maths	Raw scores	0.55
Average multiple "R"		**0.40**

(Average of variances (Horst's index of efficiency) 0.16)

Following Horst, this gives 0.17 as the index of differential efficiency in the sense of classifying individuals according to particular jobs - or, in this case, particular subject specialisations. It is notable that the highest correlation, as might be expected, comes from the pair English/Maths, which might be expected to show high performance respectively from very different candidates, coupled with the fact that, in compulsory subjects, the whole range is represented rather than a pre-selected group which would restrict the range of scores. In the Horst sense, the battery would appear to be performing as a differential instrument.

Although the battery does not perform as well in predicting criterion difference scores as it does on the single criteria making up these pairs, it is clearly able to predict differences in performance on paired criteria. As differences between scores are clearly orthogonal to the sum of the scores, it is reasonable to suppose that the indices of efficiency do not overlap, and therefore, following the Horst procedure, the absolute scores would seem to explain 0.29% of the variance, and the differences appear to explain 0.16%; a significant addition to the amount of information to be gained from the battery as a whole.

Treatment B - Ipsatised condition

Each predictor variable was entered as a purely "ipsative" score; that is, it was subtracted from the candidate's own mean ability score and then restandardised so as to remove all element of level and retain only the pattern of the ability scores.

The Horst indices of efficiency were also calculated.

A summary of the results is given below, in Tables 5.35 and 5.36 but full results may be found in Appendix A

Ipsatised scores - single criteria

Table 5.35

Criterion	Method	Multiple R
Maths	Ipsatised	0.34
Physics	Ipsatised	0.32
English Lang	Ipsatised	0.41
French	Ipsatised	0.41
Average multiple "R"		**0.37**
Average multiple "R" squared(Horst index)		**0.14**

Ipsatised scores - paired criteria

Table 5.36

Maths/French	Ipsatised	0.30
Physics/French	Ipsatised	0.35
English/French	Ipsatised	0.24
Maths/Physics	Ipsatised	0.23
English/Physics	Ipsatised	0.26
English/Maths	Ipsatised	0.25
Average multiple "R"		**0.27**
Average of variances (Horst index)		**0.07**

Although all element of score level has now been removed from the scores, it is still possible significantly to predict performance, both on the single criteria and between paired criteria, although, as was the case with the raw scores, the battery performs better in predicting performance on single than on paired criteria. Again following Horst's index of efficiency, 14% of the variance is explained by using profile shape alone with all criteria, and 7% of the variance is explained by being "doubly differential"; using criterion differences with profile shape alone.

Thus it may be seen that, even with a criterion as heavily dependent on examination levels as GCSE performance, significant predictions may be made with an approach wholly centred around the candidate's own mean. Although the validity of this approach cannot, with these criteria, match the levels of the coefficients that are gained with the raw scores, it is sufficiently high to suggest that the profile pattern is in itself predictive of performance, irrespective of score level.

The "Horst" treatment shows the relevance of using a battery to predict different levels of performance on different criteria. This supports the argument for a multidimensional view of abilities, in that it suggests that it is worth looking for high performers in particular fields, rather than assuming that high levels of performance in all areas can be predicted from general ability test performance. The ipsative treatment takes the argument further, and suggests that differences in profiles are themselves predictive of performance, irrespective of the absolute levels of scores obtained on the higher or lower within-candidate scores.

These significant multiple correlations between pure score patterns and the various academic criteria are notable, in view of the complete absence in this condition of any information relating to elevation, or levels of score differences between subjects, although there is still an element of relative levels within an individual's profile. These significant correlations are particularly interesting as a formal examination is the type of criterion which would be expected to correlate most highly with score levels on tests.

Again, it should be borne in mind that the intention is not to illustrate the superior predictive power of differentials, but to indicate they add additional information.

Detailed tables of these results may be found in Appendix A. It should be noted that this treatment has much the same effect as if within-candidate difference scores were used as predictor variables. This should be borne in mind during the discussion on MANOVA in the next section.

Although these results are encouraging in showing the usefulness of deviation scores in predicting performance, the use of the Horst indices or the average correlation coefficients does not illustrate its full potential as a differentiating battery, as paired correlational methods still emphasise actual score levels and do not readily allow patterns of scores to emerge. The use of paired variables still gives notional equal weight to all differences, whereas in practice the differential approach requires group distances from the mean rather than correlations to be examined.

Accordingly, MANOVA was examined as a possible technique for assessing the full battery, with both occupationally linked and performance related criteria.

Manova

In multivariate analysis of variance, or MANOVA, test scores may be entered as dependent variables, and the independent variables are the classifiers or criteria. The particular advantage of MANOVA is that it allows for the possibility that there may be a relationship between the independent variables which affects the association between them and the classifiers or criteria.

Tabachnik & Fidel (1989), describe the use of MANOVA in profile analysis. They show that the technique may be used to test the levels, flatness and parallelism of profiles, assuming linearity of the relationships between the dependent variables.

A test of the levels of the scores examines differences between the means of all groups combined over the dependent variables, without taking account of particular scores on any one variable. The test is therefore a one way univariate F test, equivalent to the between-subjects main effect in repeated measures ANOVA. Eta square is then used to evaluate the strength of association between the mean of the dependent variables and the classifying factor, much as it is used in discriminant function analysis.

$$\text{Eta squared} = \frac{SS_{bg}}{SS_{bg} + SS_{wg}}$$

SS_{bg} Sum of squares between groups

SS_{wg} Sum of squares within groups

If the data matrix is converted into score differences, or segments, each segment represents a slope between two original dependent variables. If there is a multivariate difference between them the slopes differ and so the groups are not parallel. Wilks' lambda tests the hypothesis of parallelism by evaluating the ratio of the determinant of the within-groups cross-products matrix to the determinant of the sum of the within and between-groups cross products matrices.

$$\text{Lambda} = \frac{S_{wg}}{S_{bg} + Swg}$$

Groups could, of course, show parallel profiles if no characteristic slope, or score differences, exist in the sample. However, in most cases the absence of parallelism will also mean the absence of flatness. Flatness of profiles may be tested using a multivariate generalisation of the one-sample t-test, usually Hotelling's T^2, or trace. The test examines whether, with groups combined, the segments deviate from zero. The test subtracts a set of hypothesised grand means representing the null hypothesis from the matrix of actual grand means.

$$T_2 = N(GM-0)' \, S_{wg-1}(GM-0)$$

where N is the total number of cases

S_{wg-1} is the inverse of the within-groups sum of squares and the cross-products matrix.

A measure of strength of association, eta squared, may be found through Hotellings T^2 that is related to lambda, which in turn may be used to find eta squared:

$$\text{Lambda} = \frac{1}{(1+T^{2)}}$$

Eta squared = 1-lambda. Eta squared then gives the percentage of variance in the combination of segments which is accounted for by non-flatness of the profile, collapsed over groups.

The parallelism test is normally used to establish that two sets of group scores do indeed come from different populations; for example, it may be used to show that profiles from two disparate groups with similar levels of scores (e.g. graduate engineers and psychologists) differ in terms of the shape of their profiles.. However, it could also be used to establish that profiles not necessarily identical in level of scores (e.g. those from majority/minority ethnic groupings) still show the same shape of profiles relative to job success.

The parallelism tests may be found in the within cells - type by abilities tables below. It should be noted that the adjustment for the extra degree of freedom has already been made by the MANOVA.

Three studies were performed in the MANOVA treatments, in various conditions, to examine the usefulness of this approach. Occupational groups, performance in the workplace and vocational interests were used as classifiers in the studies. The first was performed on a sample of engineers and teachers (N = 463); As the groups were very diverse in nature, but had roughly the same mean length of education, it was hypothesised that differences would relate to the pattern of scores rather than the mean levels, thus establishing a prima facie case for the structure of abilities described by the Morrisby Profile.

As described earlier, in chapter 3, as well as arguing for the differential predictive power within the GAT (general abilities) and practical (Shapes/Mechanical) blocks, Morrisby argued that the relative arrangement of the three "blocks" consisting of the Compound Series, the mean score of the Verbal, Numerical and Perceptual (information processing) ability scores, and the mean score of the Spatial and

Mechanical (the "practical" block,) also had predictive power. He suggested that the relative mean levels of these three blocks could be used to indicate the way in which a test candidate preferred to learn and solve problems. Accordingly, in Treatment B, the block scores are used as predictor variables, whereas, in Treatment A, the predictor variables are the deviation test scores for each of the 6 ability tests of the battery.

Manova Study A

Subjects:

344 engineers

205 teachers

Criterion:

Occupational membership

Predictor variables:

Deviation test scores

Block mean scores

Method

The predictor variables, the test scores, were entered first as deviation scores from the grand mean (Treatment A) and then as blocks of scores, (Treatment B) replicating the Morrisby block differential concept. In this block differential treatment, the compound series (abstract reasoning) score, the average of the general ability scores, and the average of the practical (spatial/mechanical) scores were entered as dependent variables, to establish whether their pattern predicted occupational membership.

Treatment A - Deviation scores

Table 5.37

"Levels" Test

Tests of Between-Subjects Effects.

Tests of Significance for OVERALL using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	671.11	461	3.62		
CONSTANT	.62	1	.62	.17	.681
GROUP	2.20	1	2.20	.61	.437

In this test, it can be seen that there is no significant univariate effect of the two occupational groups on the combined subtests. This means that the groups do not significantly differ in terms of the absolute levels of their scores, irrespective of shape, as might be expected from two graduate/managerial groups.

Table 5.38

"Parallelism" Test

EFFECT .. GROUP BY ABILITIES

Multivariate Tests of Significance

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.46	34.81	11.00	451.00	.0005
Hotellings	.85	34.81	11.00	451.00	.0005
Wilks	.54	34.81	11.00	451.00	.0005
Roys	.46				

This test shows significantly different profiles for the two occupational groups. The various multivariate tests of parallelism produce slightly different probability levels for alpha, all less than .0005. Wilks' lambda may be used for statistical evaluation and strength of association, (eta squared = 1- lambda = 0.46) or the canonical correlation (0.69) may be used. This indicates the amount of the within-segment variance (46%) which is accounted for by the difference in shape of the profiles.

Table 5.39

"Flatness" test

EFFECT .. ABILITIES

Multivariate Tests of Significance

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.19	.75	11.00	451.00	.0005
Hotellings	.24	9.75	11.00	451.00	.0005
Wilks	.81	9.75	11.00	451.00	.0005
Roys	.19				

The flatness hypothesis is rejected (p<.0005). All the multivariate criteria show essentially the same result, but Hotelling's criterion is most appropriately reported, as it is a test of a single group(both groups combined). Strength of association, however, is better shown through Wilks' lambda (1-.81) = .19). The canonical correlation is therefore 0.44. The rejection of the flatness hypothesis is only what might be expected, since the hypothesis of parallelism has already been rejected,

Table 5.40

Roots	Wilks L.	F	Hypoth. DF	Error DF	Sig. of F
1 TO 2	.58	22.21	10.00	696.00	.0005
2 TO 2	.98	1.74	4.00	349.00	.142

The canonical correlation, or the squared ratio of the between-groups sum of squares to the total sum of squares, (1-lambda) or 0.42, may be regarded as the index of efficiency of the method. This shows that about 41% of the variability in the segment scores is attributable to between-group differences. The value of Wilks' lambda (0.58, above) gives the amount of observed variability not explained by group differences.

Univariate F tests were carried out on contiguous pairs of dependent variables. Although significant associations between each variable and occupational membership can be seen in Table 5.41 (p <.0005) it should be noted that not all possible paired and block differences were included in these tests, and that therefore it is probable that the differences within the full battery might account for even more of the variance.

Table 5.41
Univariate F-tests with (2,352) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
CST & GAT-V	8.88	159.67	4.44	.45	9.79	.0005
GAT-V & GAT-N	3.72	112.96	1.86	.32	5.79	.0030
GAT-N & GAT-P	8.99	106.24	4.50	.30	14.90	.0005
GAT-P & SHAPES	83.55	169.89	41.78	.48	86.56	.0005
SHAPES & MAT	46.43	194.58	23.21	.55	42.00	.0005

Treatment B - Block Scores

Using the blocks as predictor variables, and the occupations as classifiers.

The deviation scores appear to predict occupational grouping effectively. However, it would be of equal interest to see whether the approach can distinguish groups with similar facility when the blocks of their scores, rather than the individual subtests, are used as predictor variables, in line with the concept of different problem solving techniques and preferences based on the relationship of the reasoning, processing and practical abilities.

Results for the same occupational groups, with blocks as predictor variables, are given below.

Table 5.42
"Levels" Test
Tests of Between-Subjects Effects.

Tests of Significance for OVERALL using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	806.36	461	1.75		
CONSTANT	7.53	1	7.53	4.30	.039
GROUP	26.89	1	26.89	15.37	.000

There is a significant difference between the levels of the blocks of tests between the two groups. In view of the facts that there is no difference between the levels of the separate subtests (Table 5.37), taken as a whole, this is an interesting finding, and shows the importance of examining the scores as blocks when actual size of scores is at issue.

Table 5.43

"Parallelism" test

EFFECT .. GROUP BY BLOCKS
Multivariate Tests of Significance

Test Name	Value	Approx.F	Hypoth. DF	Error DF	Sig. of F
Pillais	.14	36.27	2.00	460.00	.0005
Hotellings	.16	36.27	2.00	460.00	.0005
Wilks	.86	36.27	2.00	460.00	.0005
Roys	.14				

The groups are clearly not showing parallelism of profiles. ($p < .0005$)
The degree of association between the profile differences and the block differences (1-.86380) = 0.14, so where the blocks are used, they can only account for 14% of the variance in the differences between the blocks.

Table 5.44

Flatness

EFFECT .. BLOCKS
Multivariate Tests of Significance

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.04	10.15	2.00	460.00	.0005
Hotellings	.04	10.15	2.00	460.00	.0005
Wilks	.96	10.15	2.00	460.00	.0005
Roys	.04				

The hypothesis of flatness is rejected ($p < .0005$). The degree of association (1-.95772) (0.04) is again less than with the separate tests. This is not surprising, however, as the difference between the abstract reasoning score and the practical abilities was not examined in the contiguous differences design, and must be an important factor in determining profile differences.

A comparison of these results with those gained from the individual subtests shows that levels of significance are equally low, whereas the F-ratio is even larger when the blocks are used, pointing to the usefulness of such an approach.

Table 5.45

Univariate F-tests with (2,352) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
IND.PROC/REASON	.19	94.83	.09	.27	.35	.7070
REASON/PRAC	138.48	225.04	69.24	.64	108.30	.0005

As may be seen from these tables, there are much greater between group differences between the practical and reasoning scores than between the reasoning and processing scores. Again, the significance of the practical/processing difference was not tested, because of the contiguous segments constraint in this repeated measures design, although it is likely that it too was responsible for some of the variability.

Comparisons of the averaged tests of significance, shown below, show greater F ratios and slightly lower (more significant) levels of significance for the Blocks method, although it should be remembered that it required fewer degrees of freedom.

Table 5.46

Tests involving 'BLOCKS' Within-Subject Effect.

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	319.86	704	.45		
BLOCKS	3.45	2	1.72	3.79	.0230
TYPE BY BLOCKS	138.66	4	34.67	76.30	.0005

Finally, MANOVA was used with this group to show the usefulness of the ipsatised approach, and standardised difference scores were entered as predictor variables, with the same occupational grouping variable.

Table 5.47

Multivariate Tests of Significance

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.43	19.31	10.00	698.00	.0005
Hotellings	.73	25.20	10.00	698.00	.0005
Wilks	.57	22.21	10.00	698.00	.0005
Roys	.41				

A comparison of these results with those gained from the individual subtests and the blocks show similarly low levels of significance, and identical test values and F ratios to those gained from the ordinary standardised scores, as might have been expected.

Table 5.48
Univariate F-tests with (2,352) D. F.

Variable	Hypoth. SS	Error SS	Hypoth.MS	Error MS	F	Sig. of F
C&V	16.93	370.65	8.46	1.05	8.04	.0005
V&N	6.21	307.49	3.10	.87	3.55	.0300
N&P	19.82	353.40	9.91	1.00	9.87	.0005
P&S	186.48	389.14	93.24	1.11	84.34	.0005
S&M	89.09	377.52	44.54	1.07	41.53	.0005

As may be seen from these tables, all the differences between the deviation scores are significant, with slightly reduced, but still significant, levels on the verbal and numerical difference score. This significance figure ($p < 0.030$) is slightly less good than that obtained by the standardised scores ($p < .003$) but overall there seems little to choose between the two methods for predicting differences between groups.

Manova Study B

Subjects
28 insurance salespeople

Materials
Morrisby Profile scores
Sales figures

Classifier **Sales**
Predictor variables **Blocks**

Table 5.49

Levels

Tests of Between-Subjects Effects.

Tests of Significance for OVERALL using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	32.88	24	1.37		
CONSTANT	.12	1	.12	.09	.766
SALES	1.24	2	.62	.45	.641

There do not appear to be significant difference of level between higher and lower performing groups.

Table 5.50

Parallelism

Multivariate Tests of Significance
EFFECT .. SALES BY BLOCKS

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.35	2.59	4.00	48.00	.049
Hotellings	.44	2.39	4.00	44.00	.065
Wilks	.68	2.49	4.00	46.00	.056
Roys	.21				

However, even given the very low numbers, the groups appear to differ significantly in terms of the shape of their profiles. (Canonical correlation 0.57 $p < .05$)

It is clear from this study that the profile shapes, irrespective of level, can predict differences in performance at a significant level, and that, if the levels of the scores are examined, irrespective of shape, no such prediction could in this particular case be made.

Manova Study C

Subjects

Vocational Guidance clients: N = 13000

M = 7568 F = 5532

MANOVA was used in the third group of subjects to indicate the possibility that test results, in standardised, ipsatised and block form, could differentiate between those with interests in careers with People, Data/Information, and Things (Practical /Scientific /Technical).

The subjects all completed an interest inventory to establish the criterion scores, and their results were standardised and recoded into those 1 standard deviation and over above the mean score (High), the mean plus/minus one standard deviation (Medium) and those below one standard deviation below the mean (Low) to establish three groupings. It was hoped that, owing to the increasing use of this type of differential battery in vocational guidance, it could be shown that there is a clear link between score patterns and vocational interests.

However, it was expected that the criteria used would be subject to certain attenuating factors, as it is unusual for aptitudes alone to determine career interests. Gender (Cole and Hanson 1975; Tittle 1983) as well as background and family expectations, may play a role in career choice, and sex stereotyping, in particular, is not uncommon (White, Kruczek & Brown 1989). For this reason, separate studies were carried out with the males and females.

Detailed results may be found in Table 5.53 onwards, but in all cases a significant canonical correlation was found between the tests and the interest. A summary of the results is set out in Table 5.52.

Table 5.51

Interest	Canonical Correlation	Parallelism (Wilks)		Levels
Things (Males)	0.38	.85	p<.0001	p<.0005
Things (Females)	0.19	.96	p<.0005	p<.0005
	0.20	.96	p<.0005	p<.001

Treatment 1 - Males

Table 5.52

Dependent variable Things (Cat3)
Subjects Males

CAT3	LOW	36.76	14.42	2949
CAT3	MED	36.16	16.56	3563
CAT3	HIGH	34.05	13.81	947

Table 5.53
Levels

Tests of Significance for OVERALL using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	20210973.08	7456	2710.70		
CONSTANT	98910424.00	1	98910424	36488.90	.0005
CAT3	52301.12	2	26150.56	9.65	.0005

The hypothesis that the groups do not differ in terms of level is rejected (p = .0005)

Levels do distinguish between the high and low interest groups.

Table 5.54

Parallelism

EFFECT .. CAT3 BY INTEREST
Multivariate Tests of Significance (N = 3722)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.15	54.73	22.00	14894.00	.0005
Hotellings	.17	58.61	22.00	14890.00	.0005
Wilks	.85	56.67	22.00	14892.00	.0005
Roys	.14				

The hypothesis that the groups may have parallel profiles is rejected, showing that there are significant differences in profile shape between the groups, with a canonical correlation coefficient of 0.39 (p = .0005).

Table 5.55
Flatness

EFFECT .. INTEREST
Multivariate Tests of Significance

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.95	12834.11	11.00	7446.00	.0005
Hotellings	18.96	12834.11	11.00	7446.00	.0005
Wilks	.05	12834.11	11.00	7446.00	.0005
Roys	.95				

The hypothesis of flatness may be rejected. ($p < .0005$), as would be expected in the absence of parallelism in the profiles.

As might be expected, when combined, the three groups (low, medium and high interest) do not have a flat profile. The segments significantly deviated from the mean, with a very high canonical correlation coefficient (0.98), $p = .0005$, and so were notably low or high on several of the tests. This shows that interest in "Things" is not, after all, just a reflection of high or low "academic" ability, and therefore there are likely to be high and low performers in all the interest groups.

Treatment 2 - Females

Dependent variable:	Things (Cat3)
Subjects:	5517 females

Table 5.56

Levels
Tests of Between-Subjects Effects.

Tests of Significance for OVERALL using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	8128505.21	5514	1474.16		
CONSTANT	11256995.56	1	11256996	7636.22	.0005
CAT3	24054.57	2	12027.28	8.16	.0005

The hypothesis that the groups do not differ in terms of level is rejected ($p = .0005$) Levels do distinguish between the high and low interest groups. This may suggest that academically abler students actually be less practical, or that they may wish to perceive themselves as less practical than less able students.

Table 5.57

Parallelism

```
Effect.... CAT3 BY INTEREST
 Multivariate Tests of Significance
```

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.04	9.79	22.00	11010.00	.0005
Hotellings	.04	9.95	22.00	11006.00	.0005
Wilks	.96	9.87	22.00	11008.00	.0005
Roys	.04				

The hypothesis that the groups may have parallel profiles is rejected, showing that there are significant differences in profile shape between the groups (p = .0005), although the canonical correlation (0.20) is not as great as that of the males (0.39).

Table 5.58

Flatness

```
EFFECT .. INTEREST
Multivariate Tests of Significance
```

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.81	2098.13	11.00	5504.00	.0005
Hotellings	4.19	2098.13	11.00	5504.00	.0005
Wilks	.19	2098.13	11.00	5504.00	.0005
Roys	.81				

As might be expected, when combined, the three groups (low, medium and high interest) do not have a flat profile. The segments deviated from the mean to a highly significant degree (canon. corr. 0.899; p = .0005), and so were notably low or high on several of the tests. This shows that interest in "Things" is not just a reflection of high or low "academic" ability, and therefore there are likely to be high and low performers in all the interest groups.

It may be that this capacity to measure both elevation and parallelism can be used to study a different hypothesis; that score patterns predict with differential effect between groups. For example, it would be interesting to know whether relatively high achieving males and females on identical criteria also show identical profile patterns of test scores, or whether areas of ability have different predictive power for the two gender groups.

Manova Study D

Accordingly, a group of students, N =156, M =50 F =106, all of whom had taken and passed GCSE Mathematics at Grade A, were given the Morrisby Profile, and in a multivariate analysis of the score differences and the overall levels of the scores, the null hypothesis of no difference in either levels or shape was examined.

It was expected that there would be differences in both level and, particularly, in profile pattern, if gender affects the way in which individuals achieve in both tests and measures of academic performance.

Table 5.59
Test for Levels
Tests of Significance for OVERALL using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	248.54	154	1.61		
CONSTANT	179.20	1	179.20	111.03	.000
SEX	7.80	1	7.80	4.83	.029

This test was significant ($p < .029$) showing the groups differed in level when the tests of the battery were averaged across each group.

Examination of the means of the two groups showed that the difference was in favour of the female group ($p < .0005$).

Table 5.60
Test for Parallelism

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.33	6.50	11.00	144.00	.0005
Hotellings	.50	6.50	11.00	144.00	.0005
Wilks	.67	6.50	11.00	144.00	.0005
Roys	.33				

This test was strongly significant ($p < .0005$), showing a considerable difference in the shape of the two groups' score profiles. (Canonical correlation 0.57)

Table 5.61
Flatness test

Multivariate Tests of Significance

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.50	13.23	11.00	144.00	.0005
Hotellings	1.01	13.23	11.00	144.00	.0005
Wilks	.50	13.23	11.00	144.00	.0005
Roys	.50				

As might be expected, the two groups do not have flat profiles, when combined. There is considerable slope within the combined group. (Canonical correlation 0.71, $p < .0005$)

It appears, therefore, that males and females do not utilise the same aptitudes to achieve high performance grades. A number of factors might account for this. The male group might have been better motivated to work hard at all levels, so achieving high grades in subjects that were not their specialist areas, whereas the female group might have contained a higher proportion of girls whose grade in mathematics represented their specialist areas and so more closely followed their aptitudes. Alternatively, different patterns of aptitude may predict female and male achievement on different criteria. Although this was not the subject of this study, it seems there might be considerable interest in examining gender and minority group profiles to see if aptitudes matching achievement follow parallel patterns between the groups.

Detailed analysis of the results of these tests shows that MANOVA may be used as another useful indicator of the effectiveness of the battery as a classifying agent, and that the various significance tests may be taken as indications of its efficiency in differentiating both on the basis of actual score levels and as a profiling measure. As the technique is based on comparisons of means rather than on correlations, like discriminant analysis, which in many ways it resembles, it is less likely than multiple regression to be contaminated by spurious correlations or suppressor variables. Because the technique is used with score differences (segments) rather than deviation scores, it is not open to the same objections that may be levelled at the use of statistical techniques with ipsative data. It is, however, less useful in dealing with categorical variables, where discriminant function analysis is a more appropriate measure, as it is not dependent on analysing variance.

It may be seen from this section that there is considerable predictive and classificatory efficiency in the use of MANOVA in examining the patterns of profiles in addition to their overall level.

This chapter has sought to show that it is possible to distinguish effectively between groups by including differentials in the discriminating process. Discriminant function analysis, Morrisby differential, summative and multiple coefficients, the use of deviation scores and MANOVA have all been described, and their relative importance will be discussed in the final chapter.

Chapter 6. Discussion

The intention of this study has been to establish that a prima facie case exists for considering aptitude score patterns or profiles as important and predictive in themselves, and for using a differentiating battery that permits them to be examined, both separately and in conjunction with score levels. Thus it is suggested that in an assessment situation it may be undesirable merely to sum an individual's separate scores, whether weighted or not, on a battery of aptitude tests designed to assess different intellectual aspects, and assume that the final score in some way assesses the individual's overall intellectual ability.

Instead, it is suggested that there are many different kinds of intelligence, which in some people vary widely within the individual, and that there is added power and predictive value in possessing as one's own higher ability the aspect of intelligence regarded as most relevant to a particular job. This would not necessarily imply that score levels are irrelevant and should be ignored; only that differentials have a part to play in selection, guidance and development. However, it is also suggested that for a number of social and cultural reasons it is undesirable to concentrate on score levels to the exclusion of other information.

In order to test this position, a battery originally designed along differential lines was selected, and its performance in both differential and traditional conditions was compared. The central intention of the study has not been to validate the Morrisby Profile, but to examine the differential premise. However, it was clearly necessary as a part of the study to ensure that the Morrisby Profile was properly standardised and validated along traditional lines, so that comparisons of its performance in the traditional and differential conditions provided more than merely trivial confirmation of the usefulness of the differential approach.

The differential position is based on the assumption that individuals are likely to perform better on some aptitude tests than on others, according to their particular pattern of strengths, and that this has implications for selection and development, in that individuals are likely to perform best when in situations that match their strengths. Two versions of the differential approach have been described; using total or weighted aptitude test scores to classify individuals into jobs by distinguishing between paired criteria (the Horst position) and using the differences between pairs or groups of aptitude tests to distinguish levels of performance on a single criterion (the Morrisby position).

Although absolute score differences give some indication of these patterns of ability, it is considered that these are in themselves attenuated by score levels rather than patterns of achievement. Accordingly, although absolute score differences were examined in the study, both alone and in conjunction with score patterns, it was suggested that, for the differential premise to be properly examined, score levels should be removed from the equation, and the profile or shape of an individual's test scores should be examined without being affected by the absolute levels of scores. The Morrisby position has therefore been further developed by the use of profile patterns alone, irrespective of actual score levels, employing the deviation score approach, or the repeated measures MANOVA.

This entailed a number of different types of study. It was necessary first to assess the capacity of the battery to predict performance in the usual way, employing correlational methods to establish the construct validity of the subtests, and multiple regression with a variety of criteria to ensure that it functioned adequately in selection and guidance situations. Most of the fourth chapter consists of a description of this process.

It was also considered important, although not the central topic of the study, which was to consider profile differences in relation to one criterion, to examine the capacity of the battery to differentiate, using summed or weighted scores, between paired criteria (the Horst approach) This was discussed in the fifth chapter, in the section on deviation scores, when a multiple regression study was described which examined the capacity of both raw and deviation scores to predict performance on single and paired criteria.

This allowed both the Horst differential position (raw scores with paired criteria) and the profile difference position (deviation scores) to be considered and compared, as well as comparing prediction of single and paired criteria.

It was found that, although both the raw scores and the deviation scores predicted performance on single criteria better than on paired criteria, and that, in this condition, the raw scores outperformed deviation scores, both raw and deviation scores were still able significantly to distinguish between performance on different criteria, and in doing so there was little difference between the two differential approaches, (Average multiple R = 0.4(raw scores) and 0.37 (deviation scores)). This suggested that there is considerable scope for further study of the capacity of test batteries to distinguish between different criteria, which would be of benefit to selectors or guidance specialists asked to classify test candidates into different types of jobs, rather than the more common situation of selection/rejection.

The capacity of absolute score differences to predict performance (the John Morrisby position) was also examined in the section on the Morrisby Differential Coefficient, to be found in the fifth chapter. It was found that summative (traditional) methods of scoring normally outperformed differential methods, although in one study the reverse proved to be the case, but that in virtually all cases a combined coefficient was superior to either, suggesting that the differential approach does account for otherwise unexplained variance, and adds significantly to the amount of information gained by a selector or assessor.

The ability of the profile or shape of the battery to predict performance was examined in two ways; by means of deviation or "ipsatised" scores in discriminant function analysis, which allowed separate "deviation" or "differential" and "traditional" coefficients to be calculated according to whether raw or deviation scores were used as predictor variables, and by means of repeated measures MANOVA, which offers separate tests of parallelism ("shape") and "level" within a single study. Again, it was found that in many cases the information from the profile shape was of much more importance than the information from score levels, and, even when that was not the case, that useful information is added by examining the part played by profile pattern. It was suggested that, where categorical classifiers such as occupational grouping are used, discriminant function analysis should be employed, and where there is a linear progression within the classifying variable, MANOVA should be used.

Four questions were posed within the course of the study:

1. Can the differential battery predict group differences, using profile patterns and levels of scores?

2. Can a purely differential approach, examining score differences, predict group membership or performance?

3. If differentials are seen to add usefully to the process of selection, development or guidance, is it necessary for them to outperform score levels

as predictors of occupational success?

4. By what standards are we to assess the performance of a battery which takes the differential approach?

It seems evident that the answer to the first question is that a differential battery can predict performance, if both score differences and score levels are taken into account. The techniques presented in this study have all shown that, whether or not a purely differential approach outperforms a traditional method emphasising score levels, combinations of the methods result in better prediction and higher multiple coefficients.

Despite the fact that the battery reaches acceptable levels of reliability and validity in traditional terms, its performance is improved by utilising both the differential and the "summative" approach. In the studies described, MANOVA, discriminant function analysis, with categorical criteria, and the Morrisby multiple coefficients, with non-categorical criteria, have all shown improvements over summative coefficients, separate test criterion correlations or even multiple regression coefficients.

The second question has proved a little more intractable. In some of the studies, the differential coefficient has actually outperformed the summative coefficient, as, for instance, when success in sales was more effectively predicted using differentials than score levels. In the MANOVA studies, it was seen that most of the variance in the samples was accounted for by profile shape (lack of parallelism) while levels were only significant in one of the four studies, and this occurred when block scores, rather than separate test scores, were used as predictor variables. However, it is probable that there are circumstances in which levels would be of greater importance. The purpose of this study has not been to show the superiority of profile patterns to levels in all cases; only to show that they have a place in selection and development. Although it was not central to the study, it was interesting to note the different male and female profile shapes in a high achieving GCSE Maths group (MANOVA Study C) which suggests the possibility of further research in examining whether achieving members of gender or minority groups actually differ from majority achieving members in pattern, level or both of aptitude test scores.

The third question raised in the course of this study is whether it is necessary for differentials to outperform score levels in predicting performance, if they are seen to add usefully to the process of selection, development or guidance.

It is common for studies of psychometric methodology to suggest a new method of assessment and to justify its value by showing its superiority, in terms of quantifiable coefficients, to the existing methodology. In the case of differential profiling, however, the argument is not so much intended to be that it offers more efficiency than the traditional method, but, first, that when employed in conjunction with current methods efficiency is increased, and, secondly, that it is preferable in non-quantifiable ways to reduce the stress laid on absolute test score levels.

Psychometric tests were devised in a culture which was perceived as much more homogeneous than our own, when the premises of tests and examinations were rarely questioned. The author has frequently received anecdotal evidence from careers advisers that many people in search of employment feel the examination system has been tilted against them for social, rather than intellectual, reasons, and this sense of suspicion, often coupled with memories of failure against alien criteria, disadvantages them when they are confronted by aptitude tests. If this is so, without wishing to remove all the importance attached to levels of achievement, it would be socially just and commercially sensible to consider utilising the talents of such people despite their lower test scores.

In vocational guidance and development, where commercial constraints are either absent or less pressing, there would seem to be advantage in employing profiling methods of assessment more truly client centred than those dependent on normative scoring alone. In Chapter 4, in the section on Fairness, a study was described in which pupils at a special needs unit recognised the accuracy of reports on their abilities which had been based on "ipsatised" scores from their Morrisby Profile results. The students did not find the results either unrealistic or irrelevant because they made little or no mention of score levels; describing results in ipsative terms, centred around each student's own mean, was constructive and helpful for this particular student body.

It is suggested, therefore, that, if any of the variance can be explained using profile pattern which is unexplained using score levels as predictor variables, the use of profiling cannot be dismissed as irrelevant.

Establishing the usefulness of a differential battery with particular criteria requires an acceptable index of efficiency, and this is addressed by the last of the four questions: how the performance of a differential battery can be assessed.

Various methods have been examined in attempting to devise a useful index, which could take into account both score patterns and the scores themselves, and give due weight to each against a useful set of criteria. The methods employed have been discriminant analysis, the differential and multiple coefficients developed by John Morrisby, and MANOVA, which were described in the preceding chapter. Multiple regression with single and paired "Horst" criteria was examined in the chapter on traditional methods of validation. In addition, the use of deviation scores or the "ipsatising" technique has been described, which can, allowing for the extra degree of freedom, be applied to each of the techniques discussed.

Multiple Regression

Although multiple regression is clearly preferable to using separate criterion correlations as a method to use in evaluating a test battery, in its traditional form it cannot take account of score differences or non-linear criterion associations. However, using first raw and then deviation scores as predictor variables, regression can be used to produce separate differential and traditional indices of efficiency of a battery, provided there is only a small number of subtests in the battery, although, with a larger battery, the number of possible pairings makes this an unwieldy method to use.

It is also possible to use multiple regression in its traditional form to produce a differential index in the Horstian sense, if paired criteria are employed, as illustrated in Deviation Score Study: Multiple Regression .

Where it is required that a battery of tests is to be validated according to its capacity to produce a multiple composite score, as in many selection procedures, or to distinguish between paired criteria, as in the Horstian classification situation, it is suggested that multiple regression should be used, with the criterion (or, if appropriate, the paired criteria) as the dependent variable, but that both the differential and the traditional multiple correlation should be employed to address both level and score differences. This would allow due weight to be given to each, according to the selectors' view of the relative importance of levels and profiles. Following Morrisby's arguments, the single index of efficiency of such a battery would, in the case of a two-test battery, be the square root of the sum of the two squared coefficients, as the score differences and score sums would be orthogonal.

In a battery with more than two tests, the multidimensional nature of a full battery would preclude such orthogonality. In this case, following Horst and Brogden, the index of efficiency of a differential battery, where both the amount of the scores and their shape are to be taken into account, would therefore be the average multiple squared correlation, or the average of the criterion variances.

Although it is a simple matter to produce multiple correlation coefficients for a differential battery, assuming adequate methods of construction, using regression techniques, these may misrepresent the true usefulness of the differential battery, in that they assume that the user is searching for the most helpful composite score, rather than, for example, criterion matching, as used in vocational guidance. Including the ipsatised approach and creating a combined multiple correlation will mitigate this, but in a battery of more than two tests the potential number of differentials would mean that the real usefulness of most would be lost in multiple regression.

Although multiple regression has been suggested as the most simple method to employ when a composite test score is required, it is the contention of this study that composite test scores overemphasise levels of test scores and underplay the usefulness of test patterns. For this reason, the combined multiple correlation of differences and raw scores, although useful in certain circumstances, where crude levels of attainment in test scores are all that is sought, is probably not an ideal method to use in development or guidance. In such circumstances, where a more client-centred approach is required, a battery that has been shown to utilise score patterns would be preferable, and a method of comparing an individual's test scores to a variety of criteria that would take both level and pattern into account would seem to have more to recommend it.

Discriminant Function Analysis

Discriminant function analysis allows for comparison of group means, although in its usual form it does not take account of non-linear relationships between dependent and independent variables. However, using ipsatised scores it can be constrained into acting as a profiling tool, and it is suggested that it is the method of choice when classifiers are categorical, especially if groups do not follow a linear progression, as with occupational groupings, although it is of course possible, where groups do follow such a progression, to recast linear figures into categorical high/low terms.

Unless the number of tests in the battery is small, it is probably not desirable to include both raw and difference or deviation scores in a single discriminant function analysis, for the same reason as with multiple regression, as the number of variables involved would greatly attenuate the canonical correlation obtained. Instead, it would be preferable either to calculate the differential and the traditional canonical coefficients separately, and use the information provided by both, or to take the mean of their squared canonical correlations as the index of efficiency.

Morrisby Differential and Multiple Coefficients.

Although it might be regarded as a sufficient contribution to have developed a new model of abilities, Morrisby himself would not have been satisfied if he had failed to develop an apposite measure for validating or assessing a test battery based on that model.

The method he preferred was the computation of a differential coefficient for pairs of scores in association with a criterion, and he presented data to show that in some circumstances such a coefficient could have as much or more predictive power than that derived from the correlation of the criterion with the sum of the pair of scores.

He also showed that, in virtually all cases, a multiple coefficient, combining both pieces of information, was of greater value still, implying that differentials added to the predictive nature of the battery.

Although this approach was of value in showing the contribution of differentials to multiple correlation with the criteria, it did not in itself offer a single index of efficiency for the whole battery, taking into account level, direction and differences as the Morrisby theory would require.

In order to extrapolate such an index from his work, it would be necessary to compute separate differential and summative coefficients, using his formulae, for each of the possible pairs of the tests of a battery - 66 in the case of the Morrisby Profile - and devise multiple coefficients for each pair, which could then be used to compute an overall index of efficiency by using the mean of the criterion variances.

Although the results, shown in detail in the last chapter, indicate the effectiveness of the differential and, particularly, the multiple coefficients, which performs better than the more traditional summative coefficient, three objections may be made. First, it is difficult to ignore the fact that this method is, even with modern computing techniques, inordinately cumbersome to compute; secondly, it may be as subject to adventitious correlations as factor analysis itself, and, finally, it rests on paired differences rather than wider segmental differences, which seem to be the practical strength of Morrisby's position. The difficulty of examining all possible combinations of all tests in this way, although a relatively simple matter for parallel processing by a clinical interpreter, becomes quite impossible as a basis for computation. Using paired absolute differences also presupposes that, inevitably, original score levels are still important, which somewhat weakens the truly differential position.

It seems therefore that the method devised by John Morrisby, combining the differential and the summative correlation coefficient to produce multiple coefficients, is probably too complex computationally to use with batteries of several tests, as it involves devising separate coefficients for each test pair, and cannot deal with differences between multiple tests as opposed to pairs. Nevertheless, his concept of the two complementary coefficients seems a valuable one, and resembles the approach taken in the MANOVA studies, in which it was possible to examine the importance of both level and profile shape.

The original Horst equation, from which Morrisby's multiple coefficient was derived, was designed to establish the overall efficiency of the differential battery as a classifier between criteria. DFA and MANOVA separate and evaluate the parts played by elevation and shape. In practice other, more clinical or subjective methods are employed, such as expert evaluation of the likelihood of success on the basis of the levels and shape of a profile, measured against performance measures. This might well be the subject of a separate study, although it would be difficult to take into account external criteria which might shape the expert judgment, irrespective of the levels and shape that were intended to influence it.

MANOVA

The use of MANOVA in predicting group difference is little different from discriminant function analysis, unless it is necessary to control for several factors or covariates. Its real use, as evidenced by this study, is to examine both shape and level of profiles in the same analysis, and to provide an index for describing each of these results in the form of a canonical correlation or partial eta squared.

It is a process that would be of particular use both where the distinction between "levels" and "profile" differences is required, and where it would be useful to assess the performance of a battery according to its capacity to distinguish groups on the basis of appropriate, but not irrelevant or adversely discriminatory, factors. If the classifiers are not categorical, then repeated measures MANOVA, as described above, would be the method of choice.

In view of this, it would seem that the most fruitful line of development is an approach which allows for the pattern of scores as well as their levels to be taken into account. Both MANOVA and discriminant analysis will allow for comparison of group means, and MANOVA allows in addition for a measure of parallelism to be made which indicates group membership in those whose mean scores alone do not match those of the group. Using both these approaches, it has been apparent that the Morrisby Profile can identify group membership, whether the criterion for membership is vocational calling or high/low performance on the criterion, and that the "ipsatised" approach described earlier can predict group membership almost as well as more traditional methods, while improving on those methods when used in association with them.. The strength of the correlations achieved from ipsatised results shows that group membership, for example, depends virtually as much on individuals' personal patterns of cognitive strengths and weaknesses as on the absolute levels of their scores.

Any future user of a differential index of efficiency, therefore, should consider either discriminant function or multivariate analysis as validating tools. In the former case, a coefficient may be computed using the "ipsatised" deviation canonical correlation and the coefficient derived from the scaled scores. Again, as differences are orthogonally opposed to combined scores, there seems no reason to avoid the square root of the summed squares as the formula for obtaining this multiple coefficient. If the classifiers are not categorical, than repeated measures MANOVA, as described above, would be the method of choice.

In the case of MANOVA, the partial eta square figure derived from the Wilks' test of parallelism, or 1-lambda, based on the segments or difference scores, would provide the measure of the variance attributable to shape in the profile, and the overall canonical correlation would provide the measure of the levels.

This approach would allow selectors and development consultants to use a measure which recognises the multidimensional nature of performance, rather than relying on methods dependent on product-moment correlations which may have little functional use and overstress the element of level of score. The purpose of using the tests would affect the emphasis on traditional and differential methods. If selecting for jobs, the preceding study suggests that those with the highest scores on relevant tests should be selected first, and then that shape should be taken into account as necessary in fine-tuning the selection process. However, if there is no-one available with the "highest" levels, it is worth noting that an individual with lower scores, of the correct pattern, is likely to perform virtually as well. In career development terms, this suggests that someone with low abilities is going to perform very badly in a job which does not match the pattern of his abilities, or almost as well as a high flyer in a profile matched area. Allowing for the inevitable error variance in testing, which is likely to cast doubt on the absolute accuracy of score levels, it would seem to be useful to be able to bear in mind that lack of ability in doing cognitive tests need not mean lesser ability in job performance, provided that the job matches the profile of the performer.

Three points have been examined in the course of this work; the relevance of the differential approach; the need for a psychometric tool to measure performance perceived differentially, and the efficiency of a particular differential battery, the Morrisby Profile, as a measuring tool in this context. These points have necessitated addressing two related methodological issues; the problem of validating and devising an index of efficiency for a differential battery, and the question of the fallibility of tests based on a factor analytic, normative approach.

Although it would appear that, during the course of this study, a case has been made for each of these points, the third need not necessarily be sustained for the first two to be independently valid. Morrisby was in many senses a pioneer but his thinking was inevitably coloured by the time in which he was devising his battery and so, although he created a differential, multidimensional measuring tool, much of his own thinking was still constrained by the concept of intelligence as unitary and scalar in nature.

His claims for the predictive power of differences did not fit very comfortably into the scalar view prevalent at the time, and he did not offer a theoretical framework for his tests; nor was such a framework readily available to him when he constructed the battery. This does not prevent his battery from presenting a multidimensional picture of a candidate more appropriate to present day needs than the scalar models of his contemporaries.

Nevertheless, had his battery lacked efficiency, that would not of itself deny the need for a multidimensional approach to assessment of abilities, and, were the multidimensional approach to be superseded tomorrow by a new model, the differential test battery, with its stress on the predictive power of differences, and denial of the supreme importance of score levels, might still have a place in the new model of assessment. Although this study would seem to have shown that the Morrisby Profile is indeed a useful tool, the intention of the study was to use it illustratively rather than to direct the study towards proving its efficacy. Nevertheless, an interesting although tangential result of the study would seem to be that a battery conceived along non-traditional, non-factor analytic lines can be an effective predictor of performance.

One contention of this study has been that the factor analytic basis on which a unitary concept of ability rests is flawed psychologically and mathematically. Factor analysis is no more capable than craniometry of defining intelligence satisfactorily or of explaining differences; variance is not explained by its groupings, merely organised. The usual method of factor analysis, or rather, of principal components analysis, isolates its first principal component by sending its first axis through the average of all intercorrelations, and so a general factor, if axes remain unrotated, is virtually bound to emerge, on which the subtests will preferentially load. Once the axes are rotated, so as to rest closer to groups of factors, the general factor disappears. Both these techniques are artificial methods of attempting to reconstruct data so that it makes subjective sense; neither can be used to reify a construct of intelligence.

Factor analysis uses the matrix of covariances for a set of tests and identifies sources of underlying variation in test scores, called factors, which are supposed to account for the observed variation in those scores. Each factor is presumed to represent a different human ability, and the theories differ mainly in the number of factors adduced and their geometric relationship to one another, in terms of orthogonality, obliqueness and so on. Factors are named according to the subjective criteria of the test constructor, who will examine the measures comprising each factor so as to arrive at what he or she regards as an appropriate label. Calling the first principal component "g", or mental energy, or intelligence, or test-taking skill, is a matter of personal choice, rather like naming a dog Champion or Wolf. Naming is not definition.

A related issue, also arising from the premises of the original Morrisby tests, has been that of past test constructors' excessive reliance on correlations for validation as well as on the correlation matrix for factor based test construction.

Traditionally, test batteries have been so constructed that, at least ideally, each component test correlated positively and equally with the criterion; the "ideal measure" of the quality under discussion. Provided that each test inter-correlated with none of the others, such a battery of ten tests could give virtually perfect prediction.

Because the traditional battery assumes that each aspect of ability is separate from each of the others and bears no relationship to any of them, scores on each test can be simply added together and a summative measure of ability produced.

The basic assumption of such a battery is that each test correlates positively with the aspect of ability which it measures, and has no significant relationship, positive or negative, with any other aspect of ability.

This procedure lends itself comfortably to the overwhelming dependence on correlation coefficients which has characterised most validity research, but ignores the possibilities both of non-linear associations and of a useful association between the pattern of scores and the criterion. This has considerable implications for test-driven definitions of intellectual abilities.

Although intelligence as a concept continues to be redefined, a variety of measures are used in education, guidance and selection which purport both to assess the abilities of those tested and to make value judgments about their performance in a number of fields, based on that assessment. Current measures of intelligence may be culturally biased and often show mean differences between majority and minority groups, and the latter are therefore less likely to achieve parity of employment and status while "intelligence", so measured, is the criterion for success.

However, intelligence may not be accurately measurable in such terms. It may be that the performance of individuals relates more closely to their pattern of abilities than to the mere sum of those abilities, which has, previously, been popularly seen as representing intelligence.

Reliance on the technique to isolate the nature of intelligence as a single construct has side-stepped the controversy as to the nature of intelligence. That controversy has centred around the question of its dimensionality; that is, whether intelligence is best conceived of as a single dimension along which individuals differ, or whether it should be perceived as a multidimensional ability, so that an individual's intelligence may only be described as superior or inferior to that of another individual, or, indeed, to their own, in relation to a particular aspect of intelligence.

This question is not of purely academic interest. If ability is unidimensional, then presumably it may be unidimensionally assessed, and a test which shows high levels of general ability will be able, all other things being equal, to predict high levels on another. This has implications for selection, vocational guidance and occupational development. It is much simpler to ask candidates to sit a single test of general ability and rank them along the single dimension of its results than to devise separate tests, differentially weighted, for different criteria. It is also much easier to classify clients for vocational guidance along a single dimension of "brightness", and allocate certain jobs to certain levels of ability, than to assess a wide spread of abilities and work approaches and go to the trouble of a complex interpretation process.

Nevertheless, the simplicity and organisational convenience of the unitary approach should not blind us to its limitations. Even if it is believed that individuals have "flat" abilities, there is little evidence to bear this out. Individuals may vary widely in their abilities, and the same individuals may perform at a high level in some areas and poorly in another. Examination of the lack of "flat" profiles in the last chapter shows that, even in similar occupational groups, there is commonly a significant slope between scores.

The psychometric view of intelligence assumes that it consists of an ability or abilities which can be assessed by the use of tests, and this theory receives internal validation from experimental test data showing the high proportion of the variance between individuals which can be ascribed to the particular abilities supposed to be assessed by the tests.

This view of abilities is particularly convenient for employers wishing to select able and intelligent employees, for colleges seeking the brighter pupils, and for those working in the field of career guidance and development, as a by-product of the approach is a flourishing body of psychometric tests which claim to assess that ability in individuals, and to rank them in order of desirability in relation to their competitors.

Taken to extremes, this approach can give numbers a spurious authority, as statistical truth replaces reality. Minority groups have some reason to be suspicious of psychometrics tests, which have not always been regarded by them as either relevant or fair, and which have sometimes been perceived as unnecessarily biased in their tendency to accord high ratings to those most comfortable with the competitive, analytical culture from which the tests largely come.

Nevertheless, in the absence of more generally acceptable methods of assessing human abilities, psychometric measures are likely to remain and perform a useful function, and it is the task of those promulgating psychometric theories to try to establish their internal and external validity. External validation is harder to come by, as, in the absence of a perfect criterion of either general intelligence or of separate abilities, only approximations can be employed. At present, validation is usually based on some form of correlation between the test results, or a weighted composite score of those results, and performance on the task of interest, such as scholastic grades in particular subjects, a supervisor's assessment of an employee's performance, or informal assessment criteria such as interview ratings.

Occasionally, as, for instance, in the case of sales figures, more easily quantifiable criteria for assessing certain abilities may be found, but these are also subject to contaminating factors such as area, external economic factors and so on.

However, as this study has demonstrated, it is possible to validate a battery by examining it as a whole. This may be done either in a more traditional way, using multiple regression or discriminant function analysis with scaled scores, or by examining the shape and pattern of test results, separately from or in conjunction with actual score levels. Thus a more comprehensive picture of the individual should emerge, covering as many aspects of ability and performance as the battery is able to sustain.

The battery examined in this study covers several abilities; fluid intelligence, knowledge acquisition components in the areas of verbal, numerical and perceptual ability, practical intelligence in the form of spatial and mechanical ability, planning style, problem solving preferences, and certain personality variables, as well as manual dexterity skills. All of these, in varying combinations, seem to have a part to play in predicting success or vocational choice. Careers advisers, personnel managers and selectors may differ in their views as to the relevance of the qualities tested for the criteria in question for a particular candidate, but a case would seem to be made for having a comprehensive battery that allows all the aspects to be assessed and decisions to be made in the light of a full picture of the evidence. This means that an approach which condemns a candidate to a percentile band has to be seen as limiting and inefficient.

Although in practice none of the major theories relating to intelligence takes quite such a bald view as the "percentile band" approach, there are wide differences between the range and number of more specific abilities which are allowed to exist by the major theoreticians in the field. A variety of measures is used in education, guidance and selection which purport both to assess the intelligence of those tested and to make value judgments about their performance in a number of fields, based on that assessment of "intelligence".

However, even if the association between pattern of abilities and criterion is less than the more traditional association of score levels/criterion, any substantial association should add to the accuracy of predictions of performance levels previously based only on one form of association, or at least provide an acceptable substitute in circumstances where dependence on pure levels of scores might disadvantage certain groups in the community whose test-taking skills do not match their actual abilities.

It is common for employers to dismiss a set of test results because they do not match the informal assessment of the candidate's intelligence made by them at interview. It is also common for test results to be over-interpreted to ensure that they accord with informal decision-taking. The psychometric view clearly is not one which always impresses the majority of lay selectors with its scientific accuracy or its conceptual truth, although it can be selectively used by those who are prepared to distort flexible data for their own ends.

This is possibly because psychometrics have, as the name implies, concentrated on measurement and limitation rather than on explanation and the development of potential. One of the intentions of the present work is to suggest a wider role for the psychometric approach, reducing the concentration on scores and numbers, which may have more to do with test-taking competence than high ability levels, and concentrating instead on the structure of abilities in a manner less dependent on absolute score levels and more concerned with patterns or profiles of scores.

However, those who adhere to the psychometric view of intelligence have not agreed among themselves as to the structure of intellect or regarded it as indivisible. Throughout this century, those who agreed in describing intelligence and intellectual abilities with reference to individual differences in test performance have disagreed profoundly as to the arrangement and relationship of those abilities, although the disagreement may appear to be more apparent than real, given the dependence of virtually all such theorists on the tool of factor analysis to distinguish intellectual functions.

Deprived of the factor analytic argument, there are no strong reasons why individuals should be ranked along a single intellectual scale beyond those of the convenience of their employers or, in more sinister terms, those of the prejudice of their rulers. However, the rationale for assessment remains.

In the first place, it is useful to conceive of ability as multi-faceted in order to reflect the evident differences in intellectual direction taken by individuals who would be startled to learn that their choices could be ranked hierarchically. If different skills and abilities are required for administrators than for engineers, it is hardly controversial to suggest that it seems sensible to examine what they are, and find people whose strengths run in those directions, rather than reduce the position to absurdity by ranking the jobs and then assigning individuals to those jobs on the basis of their own intellectual ranking. If intellectual ranking has no place in such a schema, there seems little sense in employing it; if it is relevant, we need to establish where it resides.

Morrisby argued for a vectorial approach, in which the level of an individual's scores would indicate intellectual power; their direction would show intellectual structure. The capacity of individuals to perform well or badly on tests, irrespective of true competence in the abilities apparently assessed by those tests, is well known, and the ability of the Morrisby Profile to predict performance in the absence of any information about levels of scores has also been shown. It performs almost as well in its ipsatised form as when using standardised scores, and better than either when both forms can be utilised in a single measure.

The importance of the levels of individuals' scores is not denied by this study, but it is suggested that absolute level does not map onto absolute intelligence. It is more probable that some information can be gained from the level of scores which relates to an individual's capacity to make sense of his or her own strengths; to maximise on direction of ability and to minimise weaknesses. Whether or not some of that capacity is innate, it is certainly susceptible to environment, practice and training.

One of the contentions in this study has been that validity coefficients resting on correlations of single tests with single criteria lack relevance to a multidimensional view of abilities. Validity coefficients based on the multiple regression technique rest on four premises: that higher scores are always superior to lower scores; that it is only worth including in the weighted composite score those tests shown by the beta weights to correlate positively and significantly with the criterion, that it is the level of scores, suitably weighted, that predicts performance, and not the actual pattern of those scores, and that the predictor variables - the test scores - should, so far as possible, be independent of one another.

Differential theory, by contrast, challenges each of these assumptions. Although, except in its purest form, it accepts the value of higher scores, and the predictive value of between subject differences in scores, it also assumes that predictor variables operate together, not independently, and therefore that differences in performance can be predicted by within subject score differences. This means that it can useful to include in a test battery measures which do not appear, on the face of it, to be related to the task in question, but which will, when paired with another, relevant, test, predict performance, on some criteria, by virtue of the difference between the two.

Thus, it is not only relevant for a journalist to have, for example, high verbal ability; it is also important that verbal ability is higher than other, less relevant abilities. If this were not the case, and a less relevant ability - e.g. mechanical ability - actually produced that candidate's highest score, the verbal ability would be correspondingly reduced in strength and the higher ability, if untapped, would result in frustration and poor performance. Obviously, this argument would not hold water if mechanical comprehension were to be shown to be positively related to job performance.

It would be possible to raise the objection that, by including any unlikely test of a rare ability in the battery, a correlation of the criterion with a positive difference between a relevant score and the unlikely test would probably emerge as a statistical artefact, as most high performers on the criterion would show such a difference, however irrelevant it might be to their performance. Differentialists would respond that this would not be the case; a rare ability would not differentiate between high and low performers, unless, as might well be the case, the higher performers showed such a difference precisely because the irrelevant ability, if higher than the relevant one, would otherwise reduce performance levels.

Including unnecessary tests in a battery, although theoretically sound, is practically unproductive, as it is rarely possible to develop a separate battery for each conceivable criterion, and so, in practice, the most parsimonious use of tests offers the most desirable approach in selecting tests for a battery which will be of predictive use in most circumstances, although the test differences will receive different emphases with different criteria.

If this ability to control and synthesise ability "is" intelligence, and it is this that is in some part measured by psychometric tests, then the tests become useful tools for diagnosis and development, and may be used in apportioning training and training style as well as in assessing how to develop a work force to best advantage. They may also be of far more use in schools as diagnostic instruments to indicate special programmes to repair the damage done to those most vulnerable to poor educational methods and a demotivating environment.

The question "Should I be a doctor or a lawyer?" or "Should this candidate be put into mechanical or electronic engineering?" is one that cannot be answered by measuring an individual's performance on tests against a single criterion, unless it is believed that ability is unitary and that high performance on one criterion would of necessity predict high performance on another, in which case both questions become redundant.

In fact, it is not necessary to show that differences between scores are always predictive of differences on a single criterion measure. As has already been shown, combined indices of efficiency, taking elevation and shape into account, can be powerful predictors of performance. The real usefulness of differential theory lies in the possibility that score differences within subjects can predict differences in performance between subjects on a variety of performance measures. This lifts psychometric measurement out of the realms of ranking on a single scale, and becomes a useful and constructive method for indicating which of several career paths an individual might pursue, in the light of pattern of abilities and working style. It allows selectors to take account of the particular demands of the job, while recognising that these may be met in a variety of ways, and then to consider each applicant on the basis of her particular strengths rather than on the basis of a theoretical composite score.

The value of this approach in guidance and development, as well as in large selection procedures, where several jobs are in question, is apparent: a single dimension along which jobs are ranked according to a candidate's composite score is less likely to yield results than a multidimensional pattern of scores, different arrangements of which are more or less appropriate for success in different areas.

As Sternberg suggested, training might be made available to those who, relatively, lack the ability to use their gifts to their full potential; different methods of training may be made available to those who need may, if they wish, be offered training , especially in the early years before adolescence.

There are several reasons to consider a profiling model in the context of selection and development, in addition to its efficiency as a measuring tool.

If psychometrics have a real contribution to make to the study of intellectual functioning, it would seem that it is not in the search for a reductionist scale which can generalise to all intellectual functions. Rather, it would be useful to return to Binet's premise that measurement has the optimistic function of diagnosis for development rather than damnation.

This would presuppose a model designed to indicate the possibilities in each individual's pattern of abilities, in preference to one intended to reduce all performance to a single digit. This overturns the usual reverent approach to the development of a theory (first catch your theory, then find some evidence for its eternal truth, then cast around for applications). Instead, the intention precedes the model, and test evidence is given due weight but not over interpreted as construct validation when the jury is still out on the nature of the construct.

As Sternberg reflects, in presenting his triarchic theory, "this theory, like every other theory of intelligence or any other psychological construct, will have only a limited half-life".

This leads directly to the perception that intelligence is not the same for all peoples and for all time (Cole & Means 1981) and that intelligent functioning is more likely to depend on a synthesis of several general abilities than on the possession of one. This suggests the possibilities inherent in the work of pattern theorists, and the implications of true differential theory, with its emphasis on differences, their relationships and their synthesis.

One of the approaches suggested in this study has been a more widespread use of a version of the ipsative model for guidance, selection and development. This approach allows for the pattern of an individual's scores to be examined without reference to actual score levels, which may then be taken into account, if appropriate, at a later stage. This would be of particular interest when working with groups who traditionally have found it difficult, when faced with speeded psychometric tests, to perform at very high levels, either because of language and cultural difficulties or because early experiences of examinations and tests have left them insecure and nervous in a test situation, which prevents them from performing at their best.

The knowledge that their pattern of scores is of interest; that differences within an individual's test profile will be given due weight and that there are strengths in these differences may serve to alleviate some of the tension faced by these reluctant test-takers, and, in any case, the positive feedback that can be given after such an "ipsative" test session, where discussion can focus on maximising strengths without being distracted by score levels, can only be advantageous to candidates.

Since there is little loss of predictive value in the ipsative approach, and no loss, with some advantage, in employing it in addition to traditional methods, selectors might be encouraged to concentrate in their preliminary job analyses and job specifications on desirable ability patterns and combinations, rather than, as at present, concentrating on ways of extracting the candidate with the highest composite score. The normative use of the battery need not be abandoned, but emphasis could usefully shift from the level to the shape of profile.

It has been argued that ability measures may be "ipsatised" by the use of standardised deviation scores without loss of statistical integrity. Unlike the forced choice, paired personality measure usually thought of as the ipsative test, this method does not impose either forced choices or paired scales, except in so far as the final scale score can be predicted from the score on the rest; a circumstances that should be allowed for by an extra degree of freedom.

This method allows for a criterion to be compared with the pattern or shape of a candidate's scores, which may be of particular advantage when individual preferences or strengths are of more importance than overall levels of achievements. In addition, such an approach allows selectors and developers to consider the value of such strengths rather than focusing on weaknesses that may be artificially induced by the testing situation.

If a criterion may be as closely or more closely associated with a differentiated pattern of test results than with a composite test score, or even if the association is significant, this might cast doubts on g-based cognitive theories of intelligence, and suggest instead a vectorial or geometrical pattern of abilities as being closer to reality than the underlying factor of general intelligence which underpins more hierarchical theories.

It would also lend some weight to the possibility that ability has more to do with the individual's capacity to synthesise his or her particular structure of abilities than with absolute quantities or levels of ability. This recalls Sternberg's concept of the part by the environment in an individual's structure of abilities: the capacity to shape one's environment can easily be extended to include the capacity to make best use of one's own particular cognitive mix, which is one of the most important environmental factors likely to affect an individual's development and perception of herself.

The possibility that how an individual makes use of her particular cognitive mix has more to do with success than the quantities of ingredients which go to make up that mix would be of particular interest in the context of selection and development, and, yet again, suggests the need to assess more than merely linear relationships of test scores and performance criteria.

The multidimensional approach allows for the differences within and between individuals to be taken into account in selection and developmental decision making, and permits the discussion of wider opportunities in guidance counselling. A battery which effectively predicts performance and vocational interest from such a multidimensional standpoint would be a powerful aid in organisational development, and might with advantage affect our view of human abilities so as to enable us to reduce the part played by score levels and concentrate instead on pattern and diversity.

REFERENCES

Anastasi A. (1958) *Differential psychology* 3rd ed. New York, Macmillan.

Bennett G.K. & Doppelt J.E.(1947) *Differential Aptitude Test: manual and eight tests*. New York. The Psychological Corporation.

Bennett G.K. Doppelt J.E.(1948) The evaluation of pairs of tests for guidance use. *Educational and Psychological Measurement* 1948 8,319-325

Binet A.(1909) *Les idées modernes sur les enfants (with a preface by Jean Piaget)*. Paris: Flammarion.

Brogden H.E. (1951) Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. *Educational and Psychological Measurement* 11 173-196.

Brotherton C.(1980) Paradigms of selection validation: Some comments in the light of British Equal Opportunities Legislation. *Journal of Occupational Psychology*, 53, 73-9.

Burt C.L. (1937) Correlations between persons. *British Journal of Psychology*, 28, 59-96.

Burt C. (1939) Lines of possible reconcilement. *British Journal of Psychology* 30 84-94.

Burt C.(1940) *The factors of the mind.* London: University of London Press.

Burt C. (1944) Mental Abilities and Mental Factors. *British Journal of Educational psychology*, 14. 85-94.

Burt C. (1955) The evidence for the concept of intelligence. *British Journal of Educational Psychology*, 25. 158-177.

Campione J.C. Brown A.L. & Bryant N.R. (1984) Individual differences in learning and memory. In Sternberg R.J.(Ed) *Human abilities: an information-processing approach* . New York: Freeman.

Carroll J.B. (1972) Stalking the wayward factors: Review of J.P. Guilford & Ralph Hoepfner, *The analysis of intelligence. Contemporary Psychology*, 1972, 17, 321-324.

Carroll J.B. (1982) The measurement of intelligence. In Sternberg R.J. (Ed.) (1982) Handbook of Human Intelligence. Cambridge University Press, Cambridge.

Carroll J.B. (1986) *A critical synthesis of knowledge about cognitive abilities.* Final technical report, NSF Grant BNS 8212486, University of North Carolina, Chapel Hill, NC.

Carroll J.B. (1988) The measurement of intelligence. In R.Sternberg (Ed) *Handbook of Human intelligence.* Cambridge, Cambridge University Press.

Carroll J.B. (1986) Defining abilities through the person characteristic function. In S.E.Newstead,S.H.Irvine & P.L.Dann (Eds), *Principles of Modern Psychological measurement.* Hillsdale,N.J.: Erlbaum.

Carroll J.B. (1991) No demonstration that'g'is not unitary, but there's more to the story: Comment on Kranzler and Jensen. *Intelligence* 15, 423-436).

Carroll J.B. (1991) Still no demonstration that'g'is not unitary: Further comment on Kranzler and Jensen. *Intelligence* 15, 449-453.

Cattell R.B. (1943) The Measurement of Adult Intelligence. *Psychological Bulletin*, 40. 153-193.

Cattell R.B.(1949) r_p And other Coefficients of pattern similarity. *Psychometrika* 14, 4, 279-298.

Cattell R.B. (1965) Higher order factor structures and reticular vs. hierarchical formulae for their interpretation. In C.Banks & P.L.Broadhurst (Eds.) *Studies in psychology.* London: University of London Press.

Cattell R.B., Eber H.W. & Tatsuoka M.M.(1969) *The Sixteen Personality Factor Questionnaire* (3rd. edit.) Champaign, Illinois: Institute for Personality and Ability Testing.

Cattell R.B. (1969) The profile similarity coefficient r_p in vocational guidance and diagnostic classification. *Psychometrika* 34, 131-142.

Cattell R.B. (1977) *Personality and Mood by Questionnaire.* San Francisco, Jossey-Bass.

Cole N.S. & Hanson J.R. (1975)Impact of interest inventories on career choice. In E.E.Diamond (Ed.), *Issues of sex bias and sex fairness in career interest measurement,* Washington D.C.: National Institute of Education.

Cole M. & Means B. (1981) *Comparative Studies of how people think.* Harvard Univ. Press.

Crawley B., Pinder R., Herriot P.(1990) Assessment centre dimensions, personality and aptitude. *Journal of Occupational Psychology* 63,3,211-216.

Cronbach L.J. & Gleser G.F. (1946) *Similarity between persons and related problems of profile analysis*. Urbana, University of Illinois, 1952 (Technical report No. 2).

Cronbach L.J. & Gleser G.F. (1953) Assessing similarity between profiles. *Psychological Bulletin* 50, 6, 458-473.

Derrick C. (1953) *Three aspects of reading comprehension as measured by tests of different lengths*. Princeton,N.J.: Educational Testing Service, Research Bulletin , 53-58. p.176.

Detterman D.K. (1987) What does reaction time tell us about intelligence? In P.Vernon (Ed) *Speed of information processing and intelligence*. Norwood, NJ, Ablex.

Dressel P.& Mayhew L. (1954) *General education: Exploration in evaluation*. Final report of the Co-operative Study of Evaluation in General Education. Washington,D.C. American Council on Education.

Elliott C.D. (1983) *British Ability Scales*: Technical Handbook. Windsor: NFER-Nelson.

Estes W.K.(1982) Learning, memory and intelligence. In R.J.Sternberg (Ed.) *Handbook of Human Intelligence*. Cambridge: Cambridge University Press.

Eysenck H.J. (1988) The concept of "intelligence": Useful or useless? *Intelligence* 12, 1-16.

Fox G.D. & Brotherton C (1991) Objective Personality Testing - the way forward. *Guidance and Assessment Review*,7,3-5.

Frederiksen N. (1986) Towards a Broader Conception of Human Intelligence, *American Psychologist*,41, 445-452.

French J.W. (1955) The logic of and assumptions underlying differential testing. *Psychometrika*, 20, 321-330.

Ghiselli E.E.,(1973) The Validity of Aptitude Tests in Personnel Selection. *Personnel Psychology*, 26, 461-477.

Guilford J.P.(1952) When not to factor analyze. *Psychological Bulletin* 49 26-37.

Guilford J.P. (1967) *The nature of human intelligence*. New York: McGraw-Hill.

Guilford J.P. & Zimmerman W.S. (1947-1956). *Guilford-Zimmerman Aptitude Survey*. Orange, California: Sheridan Psychological Services.

Guyote M.J. & Sternberg R.J. (1981) A transitive-chain theory of syllogistic reasoning. *Cognitive Psychology*, 13. 461-525.

Hakstian A.R. & Cattell R.B. (1976) *Comprehensive ability battery*. Champaign, Ill: Institute for Personality and Ability Testing.

Hakstian A.R.& Cattell R.B. (1978) Higher stratum ability structures on a basis of twenty primary abilities. *Journal of Educational Psychology*, 70. 657-669.

Halpern D.F. (1986) *Sex differences in Cognitive Abilities*. Hillsdale N.J. Erlbaum.

Hambleton R.K.(1982) Review of Differential Aptitude Tests (Forms V and W). *Educational and Psychological Measurement*, 44, 504-505.

Hammond D. Dingley J. (1989)Sex differences in career preferences of lower sixth formers in Belfast grammar schools. *Journal of Occupational Psychology* 62,3, 263-264).

Heim A.W. (1970) *Intelligence and Personality*: Penguin Books.

Hendrickson A.E. (1982) The Biological basis of intelligence. Part 1: Theory. In H.J.Eysenck(Ed.) *A model for intelligence*. Berlin: Springer-Verlag.

Herriot P. & Ecob R.(1979) Occupational choice and expectancy-value theory: Testing some modifications. *Journal of Occupational Psychology*, 52, 311-324.

Herriot P. (1984) *Down from the Ivory Tower: Graduates and their jobs*. Chichester: John Wiley.

Holzinger K.J., Harman H.H. (1938) Comparison of two factorial analyses. *Psychometrika* 3. 45-60.

Horn J.L.(1961) Significance tests for use with r_p and related profile statistics. *Educational & Psychological Measurement* XX1,2, 363-319.

Horn J.L. (1961) Significance tests for use with r_p and related profile statistics. *Educational & Psychological Measurement* XX1,2, 363-319.

Horn W.O. (1973) *International Primary Factors Test Battery*. Stevens Point, Wis: International Tests.

Horst P. *The Prediction of Personal Adjustment* (1941). New York. Social Science Research Council. Bulletin 48 431-6.

Horst P. (1954) A Technique for the Development of a Differential Prediction Battery, *Psychological Monographs, Whole and Applied*, 68,9,1-31 (Whole No. 380).

Horst P. (1955) A Technique for the Prediction of a Multiple Absolute Prediction Battery, *Psychological Monographs: General and Applied,* 69,5,1-22(Whole No. 390).

Howe M. (1976) Good learners and poor learners. *Bulletin of the British Psychological Society,* 29. 16- 19.

Howe M.(1988) Intelligence as an explanation. *British Journal of Psychology,* 79. 349-360.

Howe M. (1988) The hazards of using correlational evidence as a means of identifying the causes of individual ability differences. *British Journal of Psychology,* 79. 539-546.

Hunt E. Frost N. & LunneborgC. (1973) Individual differences in cognition. In G.Bower (Ed.) *Advances in learning and motivation (Vol 7).* New York: Academic Press.

Irvine S.H., Dann P.L. & Anderson J.H. (1990) Towards a theory of algorithm determined cognitive test construction. *British Journal of Psychology,* 81. 173-196.

James L.R. & White J.F.(1983) Cross-situational specificity in managers' perceptions of subordinate performance, attributions and leader behaviours. *Personnel Psychology,* 36, 809-856

Jensen A.R.(1986) g: Artefact or reality? *Journal of Vocational Behaviour, 29,* 301-33.

Jensen A.R. (1969) How much can we boost IQ and scholastic achievement? *Harvard Educational Review* 33,1-123).

Jensen A.R. (1980) *Bias in Mental testing.* New York Free Press.

Jensen A.R.(1987) Process differences and individual differences in some cognitive tasks. *Intelligence,* 11, 107-136.

Johnson C.E., Wood R., Blinkhorn S.F.(1989) Spuriouser and spuriouser: the use of ipsative personality tests. *Journal of Occupational Psychology* 61,153-162.

Johnson C., Zeidner J., Scholarios D. (1990) *Improving the classification efficiency of the armed services vocational aptitude battery through the use of alternative test selection indices.* Alexandria, Institute for Defense Analyses.

Johnson-Laird P.N. (1984) Deductive reasoning ability. In Sternberg R.J. (Ed) *Human abilities: an information-processing approach* . New York: Freeman.

Journal of Educational Psychology, Intelligence and its Measurement: A symposium (1921) *Journal of Educational Psychology* 12, 123-47, 195-216,217-275.

Joynson R.B. (1989) *The Burt Affair*, Routledge: London.

Kelley T. L.(1923) A new method for determining the significance of differences in intelligence and achievement test scores. *Journal of Educational Psychology,* 1923 14 321-333.

Kline P. (1977) Cross-cultural personality testing. In Warren N.(Ed.) *Studies in Cross-cultural Psychology*. London, Academic Press.

Kline P.(1993) *The Handbook of Psychological Testing*,London,Routledge.

Kranzler J.H. & Jensen A.R. (1991) The nature of psychometric g: Unitary process or a number of independent processes? *Intelligence* 15, 397-422.

Kranzler & Jensen (1991) Unitary g: Unquestioned postulate or empirical fact? *Intelligence* 15, 437-448.

Landis R. (1976) The psychological dimensions of three measures of critical thinking and twenty-four structures-of-intellect tests for a sample of ninth grade students. *Dissertation Abstracts,* 5705-A.

Landy F.J. Farr J.L.(1980) Performance rating. *Psychological Bulletin*,87, 72-107.

Lane J. & Herriot P. (1990) Self-ratings, supervisors ratings, positions and performance. *Journal of Occupational Psychology* 63, 77-88.

Maccoby E.E. & Jacklin C.N. (1974) *The Psychology of Sex Differences,* Stanford Ca. Stanford Univ. Press.

Mitchell T.R. & Wood R.E. (1980) Supervisors' responses to subordinate poor performance: A test of an attributional model. *Organizational Behaviour and Human Performance*. 25, 123-13.

Mollenkopf W.G.(1950) Predicted differences and differences between prediction. *Psychometrika 15.409-41.*

Morrisby J.M.(1955) *The Theory of the Differential Test Battery,* Hemel Hempstead, Educational & Industrial Test Services Ltd.

Newmark C.S. et al. (1980) Using discriminant function analysis with clinical, demographic and historical variables to diagnose schizophrenia. *British Journal of Medical Psychology* 53, 365-373.

Nunnally J. (1962) The Analysis of Profile Data *Psychological Bulletin* 59 4, 311-319.

Osgood C.E. & Suci G. (1952) A measure of relation determined by both mean difference and profile information. *Psychological Bulletin* 49, 251-262.

Pellegrino J.W. (1984). Inductive reasoning ability. In R.Sternberg (Ed) *Handbook of Human intelligence.* Cambridge, Cambridge University Press.

Piaget J. (1921) Une forme verbale de la comparison chez l'enfant. *Archives de Psychologie*, 141-142.

Pollock V.E., Schneider L.S. Chui H.C., Henderson V., Zemansky M.& Sloane R.B.(1989) Visual evoked potentials in dementia: A meta-analysis and empirical study of Alzheimer's disease patients. *Biological Psychiatry*,8,1003-1-13.

Powers D.E. and Swinton S.S. (1981) Extending the Measurement of Graduate Abilities Beyond the Verbal and Quantitative Domains. *Applied Psychological Measurement,* 52,141-158.

Quereshi M.Y. (1972). The Differential Aptitude Test. In Buros O.K.(Ed.) *The VIIth Mential Measurement Yearbook.* Highland Park, Gryphon Press.

Saigh P.A. (1986) Religious Symbols and the WISC-R performance of Roman Catholic Junior High School Students, *Journal of Genetic Psychology* 147(3), 417-418.

Saville P. Willson E. (1991)The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology,* 64,3,219-238.

Rasch,G. (1960) *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Nielson.

Ree M.J.& Earles J.A.(1991) Predicting training success: not much more than g: *Personnel Psychology, 4*4, 321-332.

Schmid J. & Leiman J.M. (1957) The development of hierarchical factor solutions. *Psychometrika,* 22.53-61.

Schmidt F.L. (1988) The problem of group differences in ability test scores in employment selection. *Journal of Vocational behaviour* 33(3) 272-292.

Schmidt F.L. Hunter J.E. & Larson M.(1988) *General cognitive ability versus general and specific aptitudes in the prediction of training performance: Some preliminary findings.* Navy Personnel Research and development Center: San Diego, CA.

Segel D. (1934) *Differential Diagnosis*: Baltimore Warwick and York.

Spearman, C. (1904) "General Intelligence," objectively determined and measured. *American Journal of Psychology,* 15, 201-293.

Spearman, C. (1927) *The abilities of man.* London: Macmillan.

Sternberg R.J. (1977) *Intelligence, information-processing and analogical reasoning: The componential analysis of human abilities.* Hillsdale, NJ. Erlbaum.

Sternberg R.J. (1988) Explaining away intelligence: a reply to Howe. *British Journal of Psychology,* 78. 527-534.

Sternberg R.J. (1980) Representation and process in linear syllogistic reasoning. *Journal of Experimental Psychology*: General. 109. 119-159.

Sternberg R.J. & Gardner M.K. (1983) Unities in inductive reasoning. *Journal of Experimental Psychology:General* , 112. 80-116.

Sternberg R.J.(1985) *Beyond IQ- a triarchic theory of human intelligence.* Cambridge University Press.

Sternberg R.J. & Suben J. (1986) The socialisation of intelligence. In M. Perlmutter (Ed) *Perspectives on intellectual development: Minnesota symposium on child psychology*(Vol 19 201-235) Hillsdale N.J: Lawrence Erlbaum Associates.

Tabachnik B.G. Fidell L.S. (1989)*Using Multivariate Statistics* (2nd edition) New York Harper & Row.

Tatsuoka M.M. & Tiedeman D.V.(1954) Discriminant analysis *Review of Educational Research.* 24 402-420.

Thomson G.(1950) *The factorial analysis of human ability* (4th ed.) London. Univ of London Press.

Thorndike E.L. (1924.) The measurement of intelligence:Present status. *Psychological Review,* 31, 219-252.

Thurstone L.L. (1924) *The nature of intelligence.* London & New York: Harcourt,Brace.

Thurstone L.L. (1938) *Primary mental abilities.* Psychometric Monographs, 1.

Thurstone L.L.(1947) *Multiple factor analysis.* Chicago: Univ. of Chicago Press.

Thurstone L.L. (1940) Current issues in factor analysis. *Psychological Bulletin,* 37. 189-236.

Thurstone L.L. & Thurstone T.G. (1962) *Tests of Primary Mental Abilities(Revised edition).* Chicago: Science Research Associates.

Tittle C.K. (1983) Studies of the Effects of Career Interest Inventories: Expanding Outcome Criteria to include Women's Experiences. *Journal of Vocational Behaviour*, 22, 148-158.

Vernon P.(1947) Psychological Tests in the Royal Navy, Army and ATS. *Occupational Psychologist*, 21, 53-74.

Vernon P. (195O) *The structure of human abilities*. London: Methuen.

Vernon P. (1958) Educational Testing and Test-form factors. Princeton,N.J.: Educational Testing Service, *Research Bulletin* 58-3. 77.

Wherry (1946) Test Selection and Suppressor Variables. *Psychometrika* 11, 4 239-47.

White M.J., Kruczek T.A., Brown M.T. (1989) Occupational Sex Stereotypes among college students, *Journal of Vocational Behaviour*, 34, 289-298.

Witkin H.A. & Berry J.W. (1975) Psychological differentiation in cross-cultural perspective. *Journal of Cross-Cultural Psychology*, 6. 4-6.

Woodruffe C. (1990) *Assessment Centres.* Institute of Personnel Management.

Zeidner J. & Johnson C.(1990) *Classification Utility: measuring and Improving Benefits in Matching Personnel to Jobs*. Alexandria: Institute for Defense Analyses.

Zubin J. (1936) A technique for pattern analysis *Psychological. Bulletin,* 6,33,733.

APPENDICES

Appendix A.1 - multiple regression with selection/promotion criteria

Multiple regression to show the ability of the 12 tests (ability/personality) and 6 tests (ability only) to predict selection and promotion. N = 135 (selection), 18 (promotion). Subjects: engineering technicians.

Table A.1

Dependent variable	Selection
N	135
Test	Full 12 tests of profile
Multiple R	.64556
R Square	.41675
Adjusted R Square	.09862
Standard Error	.97826

Analysis of Variance	
Multiple R	.36997
R Square	.13688
Adjusted R Square	.05056
Standard Error	.77215

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	11.34604	.94550
Residual	120	71.54617	.59622

F = 1.58583 Signif F = .1045

Table A.2

Dependent variable	Selection
N	135
Test	6 ability tests
Multiple R	.28947
R Square	.08379
Adjusted R Square	.04051
Standard Error	.77623

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	6	6.99825	1.16638
Residual	127	76.52192	.60253

F = 1.93578 Signif F = .0799

Table A.3

Multiple R	.53517
R Square	.28640
Adjusted R Square	.26405
Standard Error	1.17650

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	212.77082	17.73090
Residual	383	530.13430	1.38416

F = 12.80984 Signif F = .0000

Dependent variable	Internal engineering tests & assessments
N	18
Tests	12 Morrisby Profile tests

Multiple R	.84182
R Square	.70866
Adjusted R Square	.00945
Standard Error	1.02551

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	12.79061	1.06588
Residual	5	5.25832	1.05166

F = 1.01352 Signif F = .5349

Table A.4

Dependent variable	Internal engineering tests & assessments
N	18
Tests	6 Morrisby ability tests

Multiple R	.64556
R Square	.41675
Adjusted R Square	.09862
Standard Error	.97826

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	6	7.52193	1.25365
Residual	11	10.52700	.95700

F = 1.30998 Signif F = .3297

Appendix A.2 - multiple regression with single and paired GCSE criteria.

A multiple regression was performed on the raw scores of the 12 tests of the Morrisby Profile (CST, Verbal, Numerical, Perceptual, Spatial, Mechanical, the four speeded personality measures, Manual Speed and Manual Skill).

Table A.5

N = 396 (French)
N = 519 (English language)
N = 188 (Physics)
N = 517 (Maths)

The dependent variables, GCSE Maths, English Language, Physics and French were successively entered, and the multiple correlations calculated.

Table A.6

Equation Number 1 Dependent Variable.. MATHS

Multiple R	.66138
R Square	.43743
Adjusted R Square	.42403
Standard Error	.92545

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	335.63648	27.96971
Residual	504	431.65926	.85647

F = 32.65708 Signif F = .0000

Table A.7

Dependent Variable PHYSICS

Multiple R	.47147
R Square	.22229
Adjusted R Square	.16896
Standard Error	1.04018

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	54.11885	4.50990
Residual	175	189.34391	1.08197

F = 4.16825 Signif F = .0000

Table A.8

Dependent Variable.. ENGLISH LANGUAGE

Multiple R	.50031
R Square	.25031
Adjusted R Square	.23253
Standard Error	.78116

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	103.09261	8.59105
Residual	506	308.76825	.61021

F = 14.07875 Signif F = .0000

Table A.9
Dependent Variable.. FRENCH

Appendix A.3.
Multiple regression performed with the same subjects, using the 12 ability tests, with six pairs of four criterion differences scores as dependent variables

Table A.10
Dependent Variable.. Maths/French

Multiple R	.37004
R Square	.13693
Adjusted R Square	.10975
Standard Error	1.16074

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	81.44306	6.78692
Residual	381	513.33213	1.34733

F = 5.03732 Signif F = .0000

Table A.11
Dependent Variable.. Physics/French

Multiple R	.41217
R Square	.16989
Adjusted R Square	.09555
Standard Error	1.19016

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	38.84479	3.23707
Residual	134	189.80827	1.41648

F = 2.28529 Signif F = .0112

Table A.12
Dependent Variable.. English Language/French

Multiple R	.35107
R Square	.12325
Adjusted R Square	.09557
Standard Error	1.10963

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	65.77442	5.48120
Residual	380	467.88198	1.23127

F = 4.45167 Signif F = .0000

Table A.13

Dependent Variable.. Maths/Physics

Multiple R	.25779
R Square	.06646
Adjusted R Square	.00207
Standard Error	.83002

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	8.53332	.71111
Residual	174	119.87310	.68893

F = 1.03220 Signif F = .4215

Table A.14
Dependent Variable.. English Language/Physics

Multiple R	.43604
R Square	.19013
Adjusted R Square	.13396
Standard Error	1.00301

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	40.85977	3.40498
Residual	173	174.04346	1.00603

F = 3.38457 Signif F = .0002

Table A.15
Dependent Variable.. English Lang/Maths

Multiple R	.55016
R Square	.30267
Adjusted R Square	.28580
Standard Error	.94840

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	193.64434	16.13703
Residual	496	446.13430	.89946

F = 17.94071 Signif F = .0000

Multiple regression performance with the same subjects and GCSE grades as dependent variables, but using the profile scores in the deviation, "ipsative" condition. A note has been made of the effect of allowing for the extra degree of freedom, where it altered significance levels.

Table A.16

Dependent Variable.. MATHS

Multiple R	.33531
R Square	.11243
Adjusted R Square	.09489
Standard Error	1.16013

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	86.26695	8.62670
Residual	506	681.02878	1.34591

F = 6.40958 Signif F = .0000

Table A.17
Dependent Variable.. PHYSICS

Multiple R	.31671
R Square	.10031
Adjusted R Square	.04948
Standard Error	1.11244

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	24.42136	2.44214
Residual	177	219.04141	1.23752

F = 1.97341 Signif F = .0387

Table A.18
Dependent Variable.. ELANG

Multiple R	.41089
R Square	.16883
Adjusted R Square	.15247
Standard Error	.82090

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	69.53531	6.95353
Residual	508	342.32555	.67387

F = 10.31881 Signif F = .0000

Table A.19
Dependent Variable.. FRENCH

Multiple R	.42210
R Square	.17817
Adjusted R Square	.15682
Standard Error	1.25929

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	132.36410	13.23641
Residual	385	610.54103	1.58582

F = 8.34672 Signif F = .0000

Multiple regression tables with the same subjects in the ipsatised condition, but under the paired criterion conditions.

Table A.20
Dependent Variable.. Maths/French

Multiple R	.29584
R Square	.08752
Adjusted R Square	.06370

Standard Error 1.19039

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	52.05454	5.20545
Residual	383	542.72065	1.41703

F = 3.67351 Signif F = .0001

Table A.21

Dependent Variable.. PHYSICS/FRENCH

Multiple R	.35207
R Square	.12395
Adjusted R Square	.05954
Standard Error	1.21362

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	28.34235	2.83423
Residual	136	200.31072	1.47287

F = 1.92429 Signif F = .0468
(Allowing the extra degree of freedom alters the significance level to .05)

Table A.22

Dependent Variable.. English Lang/French

Multiple R	.23494
R Square	.05519
Adjusted R Square	.03046
Standard Error	1.14887

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	29.45489	2.94549
Residual	382	504.20151	1.31990

F = 2.23160 Signif F = .0155
(Allowing for the extra degree of freedom alters the significance level to 0.5)

Table A.23

Dependent Variable.. Maths/Physics

Multiple R	.23006
R Square	.05293
Adjusted R Square	-.00088
Standard Error	.83124

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	6.79613	.67961
Residual	176	121.61029	.69097

F = .98357 Signif F = .4595

Table A.24

Dependent Variable.. English Language/Physics

Multiple R	.26063
R Square	.06793
Adjusted R Square	.01467
Standard Error	1.06986

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	14.59784	1.45978
Residual	175	200.30539	1.14460

F = 1.27536 Signif F = .2476

Table A.25

Dependent Variable.. English Language/Maths

Multiple R	.25101
R Square	.06301
Adjusted R Square	.04419
Standard Error	1.09716

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	10	40.31140	4.03114
Residual	498	599.46724	1.20375

F = 3.34882 Signif F = .0003

Appendix B - Manova tables for vocational guidance study

MANOVA tables showing the effects of levels, parallelism and flatness on 2400 clients for vocational guidance. Dependent variables all 12 Morrisby tests. Classificatory factor vocational interest: Science/technical, practical, artistic/literary.

Table B.1

PRACTICAL

Tests of Significance for OVERALL using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	1783.53	637	2.80		
CONSTANT	8.15	1	8.15	2.91	.089
PRAC	56.37	1	56.37	20.13	.000

Table B.2

Multivariate Tests of Significance

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.08537	11.81730	5.00	633.00	.000
Hotellings	.09334	11.81730	5.00	633.00	.000
Wilks	.91463	11.81730	5.00	633.00	.000
Roys	.08537				

Table B.3

Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.
1	.093	100.000	100.000	.292

Table B.4

Univariate F-tests with (1,637) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
CNV	1.22629	374.86112	1.22629	.58848	2.08383	.149
VNN	.99817	314.71988	.99817	.49407	2.02031	.156
NNP	1.66005	245.68118	1.66005	.38568	4.30416	.038
PNS	9.70855	334.16639	9.70855	.52459	18.50679	.000
SNM	22.04672	319.56508	22.04672	.50167	43.94649	.000

Table B.5

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	1588.99	3185	.50		
PRACTIC	5.45	5	1.09	2.19	.053
PRAC BY PRACTIC	35.64	5	7.13	14.29	.000

ARTISTIC/LITERARY

Table B.6

Tests of Significance for OVERALL using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	1805.90	637	2.84		
CONSTANT	8.18	1	8.18	2.89	.090
ARTLIT	34.01	1	34.01	12.00	.001

Table B.7

Multivariate Tests of Significance

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.03988	5.25786	5.00	633.00	.000
Hotellings	.04153	5.25786	5.00	633.00	.000
Wilks	.96012	5.25786	5.00	633.00	.000
Roys	.03988				

Table B.8
Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.
1	.042	100.000	100.000	.200

Table B.9
Univariate F-tests with (1,637) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. of F
CNV	1.30976	374.77766	1.30976	.58835	2.22616	.136
VNN	5.20361	310.51444	5.20361	.48746	10.67486	.001
NNP	.00559	247.33564	.00559	.38828	.01440	.905
PNS	.00541	343.86953	.00541	.53983	.01003	.920
SNM	7.12091	334.49089	7.12091	.52510	13.56097	.000

Table B.10

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	1610.99	3185	.51		
ARTIST	3.33	5	.67	1.32	.254
ARTLIT BY ARTIST	13.65	5	2.73	5.40	.000

Table B.11

SCIENCE/TECHNICAL
Tests of Significance for OVERALL using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	1681.84	637	2.64		
CONSTANT	10.51	1	10.51	3.98	.046
ST	39.03	1	39.03	14.78	.000

Table B.12
Multivariate Tests of Significance

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.09543	6.01339	11.00	627.00	.000
Hotellings	.10550	6.01339	11.00	627.00	.000
Wilks	.90457	6.01339	11.00	627.00	.000
Roys	.09543				

Table B.13
Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.
1	.105	100.000	100.000	.309

Table B.14

Univariate F-tests with (1,637) D. F.

Variable	Hypoth. SS	Error SS	Hypoth. MS	Error MS	F	Sig. F
CNV	3.79986	372.28756	3.79986	.58444	6.50172	.011
VNN	2.93511	312.78294	2.93511	.49103	5.97752	.015
NNP	3.57831	243.76292	3.57831	.38267	9.35082	.002
PNS	.38287	343.49207	.38287	.53923	.71003	.400
SNM	.21158	341.40022	.21158	.53595	.39478	.530
MS1	11.55591	397.37331	11.55591	.62382	18.52442	.000
S1S2	21.63635	616.48285	21.63635	.96779	22.35643	.000
S2S3	5.30872	405.51530	5.30872	.63660	8.33915	.004
S3S4	6.75014	438.18743	6.75014	.68789	9.81279	.002
S4S5	7.55452	380.78834	7.55452	.59778	12.63755	.000
S5S6	3.99769	560.67982	3.99769	.88019	4.54185	.033

Table B.15

AVERAGED Tests of Significance for MEAS.1 using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	4412.75	7007	.63		
SCITECH	6.46	11	.59	.93	.507
ST BY SCITECH	67.71	11			

Appendix C - Factor analyses with MBTI dimensions as marker variables.

With the Myers Briggs measures as marker variables, four analyses were performed, using all of the Morrisby variables and one MBTI variable each time.

1. Extraversion/Introversion

Using Extraversion/Introversion as a marker, and the OBLIMIN (oblique) method of factor rotation, the following factors were extracted:

Structure Matrix:

	FACTOR 1	FACTOR 2
STR1ITF	-.78333	.15863
STR2PMO	.12565	.75802
STR3NEC	.82246	.48848
STR4CLMH	.80892	.40952
MBS1EI	-.20370	-.71064

Factor Correlation Matrix:

	FACTOR 1	FACTOR 2
FACTOR 1	1.00000	
FACTOR 2	.19802	1.00000

The first factor appeared to identify introversion with a) a lack of overall confidence, and b) to associate extraversion with inflexibility and introversion with flexibility.

The second factor shows an association between introversion and an "inner confidence" or "inner conviction" in preference to outward confidence or decisiveness.

B. Sensing/Intuitive

Using Sensing/Intuition as a marker, the following factors were extracted, using oblique rotation.

	FACTOR 1	FACTOR 2
STR1ITF	-.55325	-.60247
STR2PMO	.43157	.01018
STR3NEC	.86204	.13545
STR4CLMH	.90986	-.07959
MBS2SN	-.11377	.89702

The first factor appears to associate the intuitive approach with the Morrisby measure of awareness - being more aware of the surroundings than confident with others. The second factor, which is not particularly strong, links Sensing with Flexibility.

C. Thinking /Feeling

	FACTOR 1	FACTOR 2
STR1ITF	-.48079	.70729
STR2PMO	.45194	.03906
STR3NEC	.88859	-.04089
STR4CLMH	.88609	-.15090
MBS3TF	.19149	.86029

Factor Correlation Matrix:

	FACTOR 1	FACTOR 2
FACTOR 1	1.00000	
FACTOR 2	-.06800	1.00000

The first factor links thinking with awareness and a lack of confidence. The second factor appears to link thinking with inflexibility.

D. Judging/Perceiving

	FACTOR 1	FACTOR 2
STR1ITF	-.64855	.03627
STR2PMO	.30488	.70922
STR3NEC	.85416	.22545
STR4CLMH	.89414	.10184
MBS4JP	-.10425	.80966

Factor Correlation Matrix:

	FACTOR 1	FACTOR 2
FACTOR 1	1.00000	
FACTOR 2	.10159	1.00000

The first factor associates perceiving, rather than judging, with a low level of confidence and high awareness. The second factor associates perceiving with outward confidence and decisiveness.

Appendix D1- program for computing discriminant analyses and multiple regressions with ipsatised and raw scores

```
SPSSPC program for computing discriminant analyses with occupational
groups
DATA LIST FILE 'CSMANTEC.dat' FREE / no age type sex SURNAME (A15)
nAME (A12)  mp1 mp2 mp3 mp4 mp5 mp6 st1 st2 st3 st4 st5 st6
 mps1 mps2 mps3 mps4 mps5 mps6  sts1 sts2 sts3 sts4 sts5 sts6 .
missing value mp1 to st6 (00).
value labels type 1 'Guidance' 2 'Technicians' 3 'Managers'.
DESCRIPTIVES/VARIABLES MP1 TO ST6/OPTIONS 3.
COMPUTE MEAN=((ZMP1 + ZMP2 + ZMP3 + ZMP4 + ZMP5 + ZMP6)/6.)
compute smean = ((ZST1 + zst2 + zst3 + zst4 +ZST5 + ZST6)6).
COMPUTE IMP1 = (ZMP1-MEAN).
COMPUTE IMP2 = (ZMP2-MEAN).
COMPUTE IMP3 = (ZMP3-MEAN).
COMPUTE IMP4 = (ZMP4-MEAN).
COMPUTE IMP5 = (ZMP5-MEAN).
COMPUTE IMP6 = (ZMP6-MEAN).
COMPUTE IST1 = (ZST1-sMEAN).
COMPUTE IST2 = (ZST2-sMEAN).
COMPUTE IST3 = (ZST3-SMEAN).
COMPUTE IST4 = (ZST4-sMEAN).
COMPUTE IST5 = (ZST5-sMEAN).
COMPUTE IST6 = (ZST6-sMEAN).
DESCRIPTIVES/VARIABLES IMP1 TO IST6/OPTIONS 3.
DSCRIMINANT/GROUPS type (1,3)/variables Zmp1 zmp2 zmp3 zmp4 zmp5
zmp6 zst1 zst2  zst3 zst4 zst5 zst6/priors size/options 6
7/statistics 1 2 6 7 12 13 15.
dscriminant/group type (1,3)/variables imp1z imp2z imp3z imp4z imp5z
imp6z ist1z ist5z /analysis Zmp1z  to zst5z/PRIORS SIZE/options 6
7/statistics 1 2 6 7 11 12 13/ analysis  IMP1z to ist5z/prIORS
SIZE/options 6 7 /statistics 1 2 6 7 11 12 13.
```

Appendix D.2 - Program for computing coefficients with paired "Horst" criteria

```
Program for computing multiple correlation coefficients with paired
and single GCSE criteria
DATA LIST FILE 'vgsexam.dta' FREE / id schid sex age cs v1 v2 n1 n2
p1 p2 sh  mat st1 st2 st3 st4 st5 st6 elang elit hist geog econ
busstud fr ger lat class  re blank1 blank2  maths addmeths phys chem
biol scil sci2 comput homecon  cdt art music  blank3 op st  soc
artlit comm prac reward interest security  pride autonomy .
MODIFY VARS/KEEP CS TO ST6 ELANG FR MATHS PHYS.
MISSING VALUE ALL (00).
descriptives/variables all/options 3.
compute verbAL = (V1 + V2).
COMPUTE NUMERIC = (N1 + N2).
COMPUTE PERCEPT = (P1 + P2).
regression/variables cs verbal to percept sh mat st1 to st6 maths
 /missing pairwise/dependent maths/method enter.
regression/variables cs verbal to percept sh mat st1 to st6 phys
 /missing pairwise/dependent phys/method enter.
regression/variables cs verbal to percept sh mat st1 to st6 ELANG
 /missing pairwise/dependent elang/method enter.
regression/variables cs verbal to percept sh mat st1 to st6 FR
 /missing pairwise/dependent FR/method enter.
COMPUTE MF = (MATHS-FR).
COMPUTE PF = (PHYS-FR).
COMPUTE EF = (ELANG-FR).
COMPUTE MP = (MATHS-PHYS).
COMPUTE EP = (ELANG-PHYS).
COMPUTE EM = (ELANG-MATHS).
regression/variables cs verbal to percept sh mat st1 to st6 MF
 /missing pairwise/dependent MF/method enter.
regression/variables cs verbal to percept sh mat st1 to st6 PF
 /missing pairwise/dependent PF/method enter.
regression/variables cs verbal to percept sh mat st1 to st6 EF
 /missing pairwise/dependent EF/method enter.
regression/variables cs verbal to percept sh mat st1 to st6 MP
 /missing pairwise/dependent MP/method enter.
regression/variables cs verbal to percept sh mat st1 to st6 EP
 /missing pairwise/dependent EP/method enter.
regression/variables cs verbal to percept sh mat st1 to st6 EM
 /missing pairwise/dependent EM/method enter.
COMPUTE MEAN = ((ZCS + ZVERBAL + ZNUMERIC + ZPERCEPT + ZSH +
ZMAT)/6). COMPUTE SMEAN = ((ZST1+ ZST2 + ZST3 + ZST4 + ZST5 +
ZST6)/6).
COMPUTE ICS = (ZCS-MEAN).
COMPUTE IV = (ZVERBAL-MEAN).
COMPUTE IN = (ZNUMERIC-MEAN).
COMPUTE IP = (ZPERCEPT-MEAN).
COMPUTE IS = (ZSH-MEAN).
COMPUTE IM = (ZMAT-MEAN).
COMPUTE IS1 = (ZST1-SMEAN).
COMPUTE IS2 = (ZST2-SMEAN).
COMPUTE IS3 = (ZST3-SMEAN).
COMPUTE IS4 = (ZST4-SMEAN).
COMPUTE IS5 = (ZST5-SMEAN).
COMPUTE IS6 = (ZST6-SMEAN).
descriptives/variables ics to is6/options 3.
regression/variables ZICS TO ZIS6 maths
 /missing pairwise/dependent maths/method enter.
```

```
regression/variables ZICS TO ZIS6 phys
 /missing pairwise/dependent phys/method enter.
regression/variables ZICS TO ZIS6 ELANG
 /missing pairwise/dependent elang/method enter.
regression/variables ZICS TO ZIS6 FR
 /missing pairwise/dependent FR/method enter.
regression/variables ZICS TO ZIS6 MF
 /missing pairwise/dependent MF/method enter.
regression/variables ZICS TO ZIS6 PF
 /missing pairwise/dependent PF/method enter.
regression/variables ZICS TO ZIS6 EF
 /missing pairwise/dependent EF/method enter.
regression/variables ZICS TO ZIS6 MP
 /missing pairwise/dependent MP/method enter.
regression/variables ZICS TO ZIS6 EP
 /missing pairwise/dependent EP/method enter.
regression/variables ZICS TO ZIS6 EM
 /missing pairwise/dependent EM/method enter.
```

Appendix E - Multiple regression tables for engineering candidates

Multiple regression with 6 tests (ability tests) and 12 tests (ability and personality tests). Dependent variables Selection and Internal promotion. N = 135 (Selection) and 18 (promotion). The subjects were all engineering technicians applying for posts. Of those selected, some were assessed for promotion on the basis of engineering and process tests, panel interview and personality assessment after a team building course.

Table E.1

Dependent variable	Selection
N	135
Test	Full 12 tests of profile

Multiple R	.36997
R Square	.13688
Adjusted R Square	.05056
Standard Error	.77215

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	11.34604	.94550
Residual	120	71.54617	.59622

F = 1.58583 Signif F = .1045

Table E.2

Dependent variable	Selection
N	135
Test	6 ability tests

Multiple R	.28947
R Square	.08379
Adjusted R Square	.04051
Standard Error	.77623

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	6	6.99825	1.16638
Residual	127	76.52192	.60253

F = 1.93578 Signif F = .0799

Table E.3

Dependent variable	Internal engineering tests & assessments
N	18
Tests	12 Morrisby Profile tests

Multiple R	.84182
R Square	.70866
Adjusted R Square	.00945
Standard Error	1.02551

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	12	12.79061	1.06588
Residual	5	5.25832	1.05166

F = 1.01352 Signif F = .5349

Table E.4

Dependent variable Internal engineering tests & assessments

N 18

Tests 6 Morrisby ability tests

ce

	DF	Sum of Squares	Mean Square
Regression	6	7.52193	1.25365
Residual	11	10.52700	.95700

F = 1.30998 Signif F = .3297

Appendix F - special needs group questionnaire

Questionnaire sent to special needs group described in Section 3:
Time limits: none. Instructions: "Please read the questions and tick the answer that you think is most true for you".

N = 17

Q1. Why did you go to the Morrisby session?

41% 7 were told to go
12% 2 wanted better self-assessment
47% 4 had other reasons

Q2. Did you take the session seriously?

88% 15 did the best they could
6% 1 did just about what they were told
6% 1 mucked around

Q3. Did you know you would get a careers guidance report?

71% 12 knew they would get a report
29% 5 were unaware a report would follow

Q4. Would you have tried harder if you had known about the report?

77% 13 tried anyway
23% 4 would have tried harder

Q5. Did you find the instructions easy or difficult to understand?

47% 8 found the instructions easy
53% 9 found the instructions mostly easy
0% 0 found the instructions difficult

Q6. Did you find any of the tests impossible to do?

94% 16 could do some of each test
6% 1 found one test too hard
0% 0 found all the tests too hard

Q7. What did you think of the look of the report?

77% 13 thought the report looked quite good
18% 3 thought the report looked ordinary
0% 0 thought it looked messy
(1 MISSING ANSWER - candidate missed second side of sheet)

Q8. What did you think of what the report said about you?

35% 6 thought it was mostly helpful
53% 9 thought that there were some helpful things in it
6% 1 thought the report was no help
0% 0 said the report made them feel depressed
(1 MISSING ANSWER)

Q9. Did you find the words and diagrams of the report easy to understand?

77% 13 found both text and diagrams easy to understand
18% 3 found text easy, diagrams hard
0% 0 found diagrams easy, text hard
0% 0 found both hard
(1 MISSING ANSWER)

Q10. Did you agree with most of what the report said about your ability?

41% 7 agreed with most of what was said
53% 9 said that some of what was said was right
0% 0 said it was completely wrong
(1 MISSING ANSWER)

Q11. Did you agree with most of what the report said about the type of person you are?

53% 9 agreed with most of it
35% 6 said that some of it was right
6% 1 said it was completely wrong
(1 MISSING ANSWER)

Q12. Did you learn anything new about yourself from the report?

24% 4 said they had learned some helpful things about themselves
54% 9 said that one or two things in the report might be worth thinking about
17% 3 said they had learned nothing new
(1 MISSING ANSWER)

Q13. Did the job suggestions seem to you to be helpful?

24% 4 found the job suggestions very helpful
59% 10 found the job suggestions quite useful
11% 2 found the job suggestions not helpful
(1 MISSING ANSWER)

Q14. Did you ask your careers teacher to talk to you about the report?

6% 1 talked to careers teacher about ideas in report
83% 14 did not talk to careers teacher
(2 MISSING ANSWERS)

Q15. Did you ask your careers officer to talk to you about the report?

12% 2 talked to the careers officer
82% 14 did not talk to careers officer
(1 MISSING ANSWER)

Q16. Did you find it better to talk to someone about the report rather than just read it on your own?

24% 4 said talking it over made it much clearer
53% 9 said they understood it better by talking it over with someone
18% 3 said they understood it without any help
(1 MISSING ANSWER)

Q17. Do you think it would be a good idea for people to take away something like this report when they go to see a careers officer for careers advice?

35% 6 said yes, definitely
59% 10 said people should be able to have it on request
0% 0 said there was no need for such a report
(1 MISSING ANSWER)

Q18. Do you think other students should have the chance to do the tests and get the reports?

94% 16 said Yes, others should be able to do the tests
0% 0 said others should not do the tests
(1 MISSING ANSWER)

Appendix G - chisquare calculations based on Morrisby Block scores.

In an attempt to examine the relative predictive power of test scores and block categories, a group of engineering technician applicants were studied, whose Morrisby Profile scores were available as well as details of whether they were accepted or rejected on the basis of a stringent selection procedure including engineering tests, interview and past experience.

The mean age of the applicants was 32.44 (min. 16, max 48, SD 6.85). Almost all were male (M = 124, F = 11). The educational background of the applicants is unknown, but the majority had entered engineering through the craft/technician apprentice routes, and all were taking part in a genuine selection procedure in the hope of being selected for a job as an engineering technician.

Mean scores for the ability tests were computed, and the groups were categorised as above or below the mean score. The groups were also assigned to "block differential" groups, to see whether, as might be expected, the group characterised as "Beta" by John Morrisby (practical and reasoning skills greater than information processing abilities) would in fact prove to be the more successful.

None of the applicants belonged to the "Gamma" type of profile - that is, reasoning strong, practical and information processing weak. All were categorised as in either the alpha (information processing stronger than practical) or beta (practical and reasoning stronger than information processing) profiles. The Morrisby hypothesis would be that the beta group would contain the majority of successful candidates, assuming an ideal selection procedure and all other things being equal. In addition, for the purposes of this study, it was assumed that sheer size of profile would be a less effective determinant.

In fact, this proved to be the case. In Table F1 "Block by Success", beta candidates are in the majority in the successful group, (p <.06), and it can also be seen that there is a majority of "alpha" candidates in the unsuccessful group.

Table F1
BLOCK by SUCCESS

BLOCK	SUCCESS NO	YES	Total
alpha	33	8	41 51.3
beta	24	15	39 48.8
Column Total	57 71.3	23 28.8	80 100.0

Chi-Square	Value	DF	Significance
Pearson	3.50368	1	.06123

It may be seen from this table that there is a clear association (p<0.06; N = 80) between success in the assessment process, and test patterns. This may be compared with the following table, which compares total ability scores with success. Table F2 shows thatthere is no significant association (p<0.4) between sheer size of ability scores and success in application, despite the fact that the group is a little larger, and the relationship might therefore be expected more easily to reach apparent significance.

Table F2
SCORE by SUCCESS

SCORE (Count)	SUCCESS YES	NO	Row Total
LOW	31	10	41 50.4
HIGH	21	18	39 49.6
Column Total	52 64.4	28 35.6	80 100.0

Chi-Square	Value	DF	Significance
Pearson	.61332	1	.43354

Minimum Expected Frequency - 23.822

These tables are of particular interest in that they suggest a need to examine profile patterns as an integral part of the selection procedure.

Appendix H - program for computing Morrisby coefficients

Program devised for this study for computing Morrisby summative, differential and multiple coefficients for score pairs, for a battery of 10 variables, using a single criterion.

```
SPSS/PC+ The Statistical Package for IBM PC

DATA LIST FREE / cstv cstn cstp cstsh cstm csts1 csts2 csts3 csts4
vn vp vsh vm vs1 vs2 vs3 vs4 np nsh nm ns1 ns2 ns3 ns4 psh pm ps1
ps2 ps3 ps4 shm shs1 shs2 shs3 shs4 ms1 ms2 ms3 ms4 s1s2 s1s3 s1s4
s2s3 s2s4 s3s4 cs vs ns ps shs ms s1s s2s s3s s4s.
BEGIN DATA.
(Here the appropriate intercorrelations are inserted from the
intercorrelation matrix, ending (cs to s4s) with the correlations of
the variables with the criterion.)
END DATA.
compute sumcv = ((cs + vs)/SQRT(2*(1 + cstv))).
compute sumcn = ((cs+ns)/SQRT(2*(1 + cstn))).
compute sumcp = ((cs+ps)/SQRT(2*(1 + cstp))).
compute sumcsh = ((cs+shs)/SQRT(2*(1 +cstsh))).
compute sumcm = ((cs+ms)/SQRT(2*(1 +cstm))).
compute sumcs1 = ((cs+s1s)/SQRT(2*(1 +csts1))).
compute sumcs2 = ((cs+s2s)/SQRT(2*(1 + csts2))).
compute sumcs3 = ((cs+s3s)/SQRT(2*(1 + csts3))).
compute sumcs4 = ((cs+s4s)/SQRT(2*(1 +csts4))).
compute sumvn = ((vs+ns)/SQRT(2*(1 + vn))).
compute sumvp = ((vs+ps)/SQRT(2*(1 + vp))).
compute sumvsh = ((vs+shs)/SQRT(2*(1 + vsh))).
compute sumvm = ((vs+ms)/SQRT(2*(1 +vm))).
compute sumvs1 = ((vs+s1s)/SQRT(2*(1 +vs1))).
compute sumvs2 = ((vs+s2s)/SQRT(2*(1 + vs2))).
compute sumvs3 = ((vs+s3s)/SQRT(2*(1 + vs3))).
compute sumvs4 = ((vs+s4s)/SQRT(2*(1 +vs4))).
compute sumnp = ((ns+ps)/SQRT(2*(1 +np))).
compute sumnsh = ((ns+shs)/SQRT(2*(1 + nsh))).
compute sumnm = ((ns+ms)/SQRT(2*(1 +nm))).
compute sumns1 = ((ns+s1s)/SQRT(2*(1 +ns1))).
compute sumns2 = ((ns+s2s)/SQRT(2*(1 +ns2))).
compute sumns3 = ((ns+s3s)/SQRT(2*(1 + ns3))).
compute sumns4 = ((ns+s4s)/SQRT(2*(1 +ns4))).
compute sumpsh = ((ps+shs)/SQRT(2*(1 +psh))).
compute sumpm = ((ps+ms)/SQRT(2*(1 + pm))).
compute sumps1 = ((ps+s1s)/SQRT(2*(1 +ps1))).
compute sumps2 = ((ps+s2s)/SQRT(2*(1 +ps2))).
compute sumps3 = ((ps+s3s)/SQRT(2*(1 + ps3))).
compute sumps4 = ((ps+s4s)/SQRT(2*(1 +ps4))).
compute sumshm = ((shs+ms)/SQRT(2*(1 +shm))).
compute sumshs1 = ((shs+s1s)/SQRT(2*(1 +shs1))).
compute sumshs2 = ((shs+s2s)/SQRT(2*(1 +shs2))).
compute sumshs3 = ((shs+s3s)/SQRT(2*(1 + shs3))).
compute sumshs4 = ((shs+s4s)/SQRT(2*(1 +shs4))).
compute summs1 = ((ms+s1s)/SQRT(2*(1 +ms1))).
compute summs2 = ((ms+s2s)/SQRT(2*(1 +ms2))).
compute summs3 = ((ms+s3s)/SQRT(2*(1 + ms3))).
compute summs4 = ((ms+s4s)/SQRT(2*(1 +ms4))).
compute sums1s2 = ((s1s+s2s)/SQRT(2*(1 +s1s2))).
compute sums1s3 = ((s1s+s3s)/SQRT(2*(1 + s1s3))).
```

```
compute sums1s4 = ((s1s+s4s)/SQRT(2*(1 +s1s4))).
compute sums2s3 = ((s2s+s3s)/SQRT(2*(1 + s2s3))).
compute sums2s4 = ((s2s+s4s)/SQRT(2*(1 +s2s4))).
compute sums3s4 = ((s3s+s4s)/SQRT(2*(1 +s3s4))).
list /variables sumcv to sums3s4.
compute difcv = ((cs - vs)/SQRT(2*(1 - cstv))).
compute difcn = ((cs-ns)/SQRT(2*(1 - cstn))).
compute difcp = ((cs-ps)/SQRT(2*(1 - cstp))).
compute difcsh = ((cs-shs)/SQRT(2*(1 -cstsh))).
compute difcm = ((cs-ms)/SQRT(2*(1 -cstm))).
compute difcs1 = ((cs-s1s)/SQRT(2*(1 -csts1))).
compute difcs2 = ((cs-s2s)/SQRT(2*(1 - csts2))).
compute difcs3 = ((cs-s3s)/SQRT(2*(1 - csts3))).
compute difcs4 = ((cs-s4s)/SQRT(2*(1 -csts4))).
compute difvn = ((vs-ns)/SQRT(2*(1 - vn))).
compute difvp = ((vs-ps)/SQRT(2*(1 - vp))).
compute difvsh = ((vs-shs)/SQRT(2*(1 - vsh))).
compute difvm = ((vs-ms)/SQRT(2*(1 -vm))).
compute difvs1 = ((vs-s1s)/SQRT(2*(1 -vs1))).
compute difvs2 = ((vs-s2s)/SQRT(2*(1 - vs2))).
compute difvs3 = ((vs-s3s)/SQRT(2*(1 - vs3))).
compute difvs4 = ((vs-s4s)/SQRT(2*(1 -vs4))).
compute difnp = ((ns-ps)/SQRT(2*(1 -np))).
compute difnsh = ((ns-shs)/SQRT(2*(1 - nsh))).
compute difnm = ((ns-ms)/SQRT(2*(1 -nm))).
compute difns1 = ((ns-s1s)/SQRT(2*(1 -ns1))).
compute difns2 = ((ns-s2s)/SQRT(2*(1 -ns2))).
compute difns3 = ((ns-s3s)/SQRT(2*(1 - ns3))).
compute difns4 = ((ns-s4s)/SQRT(2*(1 -ns4))).
compute difpsh = ((ps-shs)/SQRT(2*(1 -psh))).
compute difpm = ((ps-ms)/SQRT(2*(1 - pm))).
compute difps1 = ((ps-s1s)/SQRT(2*(1 -ps1))).
compute difps2 = ((ps-s2s)/SQRT(2*(1 -ps2))).
compute difps3 = ((ps-s3s)/SQRT(2*(1 - ps3))).
compute difps4 = ((ps-s4s)/SQRT(2*(1 -ps4))).
compute difshm = ((shs-ms)/SQRT(2*(1 -shm))).
compute difshs1 = ((shs-s1s)/SQRT(2*(1 -shs1))).
compute difshs2 = ((shs-s2s)/SQRT(2*(1 -shs2))).
compute difshs3 = ((shs-s3s)/SQRT(2*(1 - shs3))).
compute difshs4 = ((shs-s4s)/SQRT(2*(1 -shs4))).
compute difms1 = ((ms-s1s)/SQRT(2*(1 -ms1))).
compute difms2 = ((ms-s2s)/SQRT(2*(1 -ms2))).
compute difms3 = ((ms-s3s)/SQRT(2*(1 - ms3))).
compute difms4 = ((ms-s4s)/SQRT(2*(1 -ms4))).
compute difs1s2 = ((s1s-s2s)/SQRT(2*(1 -s1s2))).
compute difs1s3 = ((s1s-s3s)/SQRT(2*(1 - s1s3))).
compute difs1s4 = ((s1s-s4s)/SQRT(2*(1 -s1s4))).
compute difs2s3 = ((s2s-s3s)/SQRT(2*(1 - s2s3))).
compute difs2s4 = ((s2s-s4s)/SQRT(2*(1 -s2s4))).
compute difs3s4 = ((s3s-s4s)/SQRT(2*(1 -s3s4))).
list/variables difcv to difs3s4.
compute multcv = (SQRT((sumcv * sumcv)+(difcv * difcv))).
compute multcn = (SQRT((sumcn * sumcn)+(difcn * difcn))).
compute multcp = (SQRT((sumcp * sumcp)+(difcp * difcp))).
compute multcsh = (SQRT((sumcsh * sumcsh)+(difcsh * difcsh))).
compute multcm = (SQRT((sumcm * sumcm)+(difcm * difcm))).
compute multcs1 = (SQRT((sumcs1 * sumcs1)+(difcs1 * difcs1))).
compute multcs2 = (SQRT((sumcs2 * sumcs2)+(difcs2 * difcs2))).
compute multcs3 = (SQRT((sumcs3 * sumcs3)+(difcs3 * difcs3))).
compute multcs4 = (SQRT((sumcs4 * sumcs4)+(difcs4 * difcs4))).
compute multvn = (SQRT((sumvn * sumvn)+(difvn * difvn))).
compute multvp = (SQRT((sumvp * sumvp)+(difvp * difvp))).
compute multvsh = (SQRT((sumvsh * sumvsh)+(difvsh * difvsh))).
```

```
compute multvm = (SQRT((sumvm * sumvm)+(difvm * difvm))).
compute multvs1 = (SQRT((sumvs1 * sumvs1)+(difvs1 * difvs1))).
compute multvs2 = (SQRT((sumvs2 * sumvs2)+(difvs2 * difvs2))).
compute multvs3 = (SQRT((sumvs3 * sumvs3)+(difvs3 * difvs3))).
compute multvs4 = (SQRT((sumvs4 * sumvs4)+(difvs4 * difvs4))).
compute multnp = (SQRT((sumnp * sumnp)+(difnp * difnp))).
compute multnsh = (SQRT((sumnsh * sumnsh)+(difnsh * difnsh))).
compute multnm = (SQRT((sumnm * sumnm)+(difnm * difnm))).
compute multns1 = (SQRT((sumns1 * sumns1)+(difns1 * difns1))).
compute multns2 = (SQRT((sumns2 * sumns2)+(difns2 * difns2))).
compute multns3 = (SQRT((sumns3 * sumns3)+(difns3 * difns3))).
compute multns4 = (SQRT((sumns4 * sumns4)+(difns4 * difns4))).
compute multpsh = (SQRT((sumpsh * sumpsh)+(difpsh * difpsh))).
compute multpm = (SQRT((sumpm * sumpm)+(difpm * difpm))).
compute multps1 = (SQRT((sumps1 * sumps1)+(difps1 * difps1))).
compute multps2 = (SQRT((sumps2 * sumps2)+(difps2 * difps2))).
compute multps3 = (SQRT((sumps3 * sumps3)+(difps3 * difps3))).
compute multps4 = (SQRT((sumps4 * sumps4)+(difps4 * difps4))).
compute multshm = (SQRT((sumshm * sumshm)+(difshm * difshm))).
compute multshs1 = (SQRT((sumshs1 * sumshs1)+(difshs1 * difshs1))).
compute multshs2 = (SQRT((sumshs2 * sumshs2)+(difshs2 * difshs2))).
compute multshs3 = (SQRT((sumshs3 * sumshs3)+(difshs3 * difshs3))).
compute multshs4 = (SQRT((sumshs4 * sumshs4)+(difshs4 * difshs4))).
compute multms1 = (SQRT((summs1 * summs1)+(difms1 * difms1))).
compute multms2 = (SQRT((summs2 * summs2)+(difms2 * difms2))).
compute multms3 = (SQRT((summs3 * summs3)+(difms3 * difms3))).
compute multms4 = (SQRT((summs4 * summs4)+(difms4 * difms4))).
compute mults1s2 = (SQRT((sums1s2 * sums1s2)+(difs1s2 * difs1s2))).
compute mults1s3 = (SQRT((sums1s3 * sums1s3)+(difs1s3 * difs1s3))).
compute mults1s4 = (SQRT((sums1s4 * sums1s4)+(difs1s4 * difs1s4))).
compute mults2s3 = (SQRT((sums2s3 * sums2s3)+(difs2s3 * difs2s3))).
compute mults2s4 = (SQRT((sums2s4 * sums2s4)+(difs2s4 * difs2s4))).
compute mults3s4 = (SQRT((sums3s4 * sums3s4)+(difs3s4 * difs3s4))).
list/variables multcv to mults3s4.
```