

**A research project to evaluate the use of psychometric testing in selection and
assessment procedures in an organisation**

Lynne Hindmarch
Birkbeck College, University of London

Paper submitted in partial fulfilment of the requirements for the M.Sc. in
Organizational Behaviour

August 1994

CONTENTS

	Page
Acknowledgement	2
Abstract	3
Section 1	
Background	4
Initial discussions with Human Resources Department	4
Literature review	6
Research question and hypotheses	19
Model of good testing practice	19
Research design	20
Section 2	
Methodology	21
Populations	21
Questionnaires	21
Pilot study	21
Data collection methods	21
Data analysis methods	22
Section 3	
Results	23
Evaluation of the results against the model	23
Hypotheses testing	36
Section 4	
Discussion	42
Conclusion and recommendations	47
References	49
Appendix A: Model of good testing practice	
Appendix B: Questionnaire for Test Administrators	
Appendix C: Questionnaire for District Team Managers	
Appendix D: Questionnaire for Human Resources Advisers	

Acknowledgement

The author of this report would like to thank the company for permitting the research project to be conducted within the organisation. Particular thanks are extended to the Director of Human Resources and members of the Human Resources Department, who provided valuable information in the early stages of the project, helped to distribute and collect the questionnaires and arrange the interviews. The author would also like to thank all those employees of the company who took part in the research, by completing questionnaires or agreeing to be interviewed.

Abstract

This paper evaluated the use of psychometric testing in selection and assessment procedures in an organisation. A model of good testing practice was developed; based on the model, three questionnaires were designed for specific populations involved in testing in the organisation. Results suggested that, while the organisation's use of tests followed the model of good testing practice in some respects, there were areas where test use could be improved. Recommendations for improvement were made, including the development of an in-house policy document on testing; improvements to test administration; clearer guidance on test weighting; improvements to provision of debriefing and feedback; and the introduction of monitoring and validation of test results. It was suggested that further research should be conducted to evaluate the effects of psychometric testing on those employees who have been tested.

SECTION 1.

Background

The company in which this study took place is a utility which was privatised in 1989. It employs approximately 7,000 people and is spread over a fairly wide geographical area, with 15 district offices in addition to its headquarters.

The company introduced psychometric testing about 10 years ago to improve its selection process. Initially testing was used during the selection of senior managers, but its use has increased over the years in two ways. Firstly, it may now be used during the selection of supervisors, team leaders and those employees who have contact with customers. Secondly, it may now be used for processes other than selection (eg for assessing training needs or promotion potential). In recent years the number of tests carried out has on occasion exceeded 1,000 per year.

The tests commonly used by the company fall into three categories. These are ability tests (eg tests of verbal or numerical reasoning); aptitude tests (only recently introduced and mainly focused on clerical checking tests); and personality questionnaires (Saville & Holdsworth's OPQ and Cattell's 16PF).

Initial discussions with the HR Department

Discussions with the HR Department revealed that the adopted procedure for the use of psychometric testing in the company, is for those HR advisers trained in testing to make recommendations to the district management teams on the appropriate tests for the task in hand. The managers who make up the district teams are not trained in test use. The administration of tests is carried out by HR assistants who have received appropriate training. Feedback on tests results is given to internal candidates if they request it; external candidates are not normally given feedback. The data from tests are considered to be valid for two years.

No policy guidelines or criteria on test use have been drawn up by the company, although an HR adviser has been identified as the specialist in psychometric testing (and has received appropriate training). Her role is one of providing advice and support on matters of testing rather than laying down policy. The company also employs a qualified occupational psychologist as a senior HR adviser.

During discussions the view taken by key HR personnel involved in testing was that the results of psychometric testing should carry the lowest weighting in the assessment process. They indicated that data from testing should be used to develop probes and to set up hypotheses for exploring during interview.

Discussions also revealed a number of suppositions about the use of testing in the company. Firstly, that test use varied between the organisation's district offices. Some districts appeared to use tests extensively, while others were hardly using them at all. Secondly, that this may be related to the confidence and experience in psychometric testing of the district team managers and the HR advisers. Thirdly, that the selection of a test was likely to be based on common practice rather than specific suitability for the task in hand. Fourthly, that the weighting given to the test battery varied between the district offices of the company.

It was felt that, although psychometric tests had been used in the company for a number of years, there was little information about how tests were used in practice in the various district offices. The company had recently introduced a quality initiative, and was committed to extending it to all aspects of organisational life, including the assessment process. It was in this context that the review of psychometric testing in the company took place. The four suppositions listed in the previous paragraph formed the basis of the hypotheses.

Literature review

Legislation and guidelines for test use

The relevant legislation in the UK is The Sex Discrimination Act 1975 and The Race Relations Act 1976. There are no regulatory guidelines on gender discrimination in selection testing, although the Code of Practice produced by the Equal Opportunities Commission gives general guidance under the definition of indirect discrimination.

This definition states that if a test results in a considerably larger percentage of one gender being rejected by comparison with another, the employer could be called upon to justify the use of the test. The Race Relations Act 1976 makes it unlawful to discriminate on racial grounds in recruitment and selection procedures. The Race Relations Code of Practice in Employment (Commission for Racial Equality, 1985) is a statutory code advising employers how to avoid direct or indirect racial discrimination in selection and testing.

In addition to the guides produced by the test publishers relating to specific tests, general guidelines on good testing practice have been produced by organisations which have an interest in ensuring that testing is carried out in a fair and ethical manner. Firstly, the British Psychological Society (BPS) has produced 'Psychological Testing: A Guide' (1992), which is not confined to use of tests in occupational settings, but includes their use in clinical and educational areas as well. This guide has been supplemented by a certification scheme of competence in occupational testing. Level A standards in occupational testing cover the general foundations of testing and the performance skills associated with test administration and interpretation for group ability tests. Development work is continuing on Level B standard, which is intended to cover personality assessment and the interpretation and use of personality questionnaires.

Secondly, the Institute of Personnel Management (IPM) produced a code (1988) on occupational testing, which is aimed at those who use tests for employment purposes, including personnel practitioners, consultants and line managers.

Thirdly, the Commission for Racial Equality (CRE) has published a guide for employers (1992) to draw attention to the effect which the design and use of psychometric testing may have on the equality of opportunity for people from ethnic minority groups.

Fourthly, the Equal Opportunities Commission (EOC) has produced a guide for employers (1988) which is designed to provide guidance in the use of selection tests without gender bias.

Finally it should be noted that the situation regarding legislation is somewhat different in the USA, where the Civil Rights Act 1964 demanded that organisations take 'affirmative action' in providing equal employment opportunities for various subgroups. By 1978 employers were required to conform to strict guidelines on selection issued by the US Equal Employment Opportunities Commission. Much of the research into psychometric testing has taken place in the USA, and the direction it has taken has been greatly affected by the legislation on employment.

Growth of Test Use

A number of surveys designed to gather data on the use of psychometric tests in UK organisations has taken place in the last 10 years, namely: Robertson and Makin (1986); Bevan and Fryatt (1988); Mabey (1989, 1990, 1991); the Local Government Training Board (LGTB) (1986, 1989); and Williams (1991).

Mabey's surveys, which focus on test use in recruitment and development in companies with 1,000 or more employees, indicate that 3 out of 4 large organisations use either

ability tests (including tests of aptitude and general intelligence) or personality tests. In particular, Mabey (1992) notes that there is a perceived growth in the use of personality questionnaires in these organisation from 47% of companies in the 1988 survey to 57% in 1991. During this period the penetration of ability tests has remained relatively stable, with 66% of organisations reporting use in 1989 and 63% in 1991.

The three surveys among local authorities in England and Wales undertaken by the LGTB and Williams found lower test usage than in business with 45% using personality questionnaires and 39% using cognitive tests (Williams, 1994). However, Williams noted some increased use of tests in responding LAs from 42% in 1986 to 51% in 1991.

Bevan and Fryatt's survey (1988), however, reported only 16% of their responding organisations using cognitive tests and 22% using personality questionnaires. Williams (1994) notes that this is 'astonishingly low'. However, he points out that Bevan and Fryatt's survey was of organisations of different sizes, but no data are presented to demonstrate the effect of size on the results. Mabey (1992) points out that it is likely that amount of test use is related to organisational size and the figures are likely to be lower for small and medium sized organisations.

The surveys by Bevan and Fryatt (1988) and by Williams (1991) give an indication of tests being used mainly for managers and professionals, though LAs are more likely to extend their use of cognitive tests to clerical grades.

Further evidence from the surveys by Robertson and Makin (1991) and Shackleton and Newell (1991) indicate increased use of cognitive tests and personality questionnaires in making managerial appointments. The survey by Williams (1991), although indicating a lower rate of test use in general among LAs compared to business, show that LAs make greater use of personality questionnaires than cognitive tests. Williams

(1994) suggests that this may show an over-optimistic belief among LAs of what tests are able to do, and their objectivity and fairness. Williams points out that tests are constructed to be objective and unbiased, but from the surveys it is impossible to say if this is true of their application. Indeed, a survey sponsored by the CRE in 1993 showed that few organisations are monitoring test results by ethnic origin and still fewer are engaged in validation.

Williams (1994) concludes that in general the surveys reveal increasing test use, particularly for employee selection. Pertinent to this is the role of the test publishers. Fletcher (1994) comments on the influence of commercial forces in this field, and debates the question of responsibility between publisher and test user for competent test use. Mabey (1992) notes that the increased use of personality questionnaires since 1988 has been accompanied by 'a marked increase in promotional activity' (p.8) and states that there is an 'ongoing need to ensure that high standards are applied in the training and support given to the now considerable number of personality questionnaire users' (p.8). Williams (1994) comments on the lack of evidence about **how** tests are used, which makes it difficult to estimate their impact in employment settings and calls for qualitative research as well as quantitative to gather information on current test use. In particular, as Fletcher (1994) points out, a broader consideration of the ethical framework for assessment practice is also essential.

Test use and ethics

Smith and Robertson (1993) state that 'ethical issues are important aspects which should permeate the whole selection process' (p.7). This is particularly true of psychometric testing, as Cronbach (1990) points out, because of the potentially intrusive nature of tests.

As has been mentioned earlier, both the BPS (1992) and IPM (1988) have produced ethical guidelines on how psychometric tests should be used, and these are

supplemented by guides from the CRE (1992) and EOC (1988). These guides variously cover all aspects of good testing practice in organisations. This includes the importance of job analysis, test selection, training of personnel in test administration and interpretation, responsibilities to test-takers (including provision of feedback), the validation and monitoring of test use (including monitoring results by ethnic origin and gender), and the development and implementation of an in-house code of good testing practice. In addition, it is generally recommended that tests are used as the basis for discussion in interview rather than used alone to make selection decisions.

However, as mentioned earlier, little is known about how tests are used in organisations, and if the guidelines are followed in practice. Research into test practice sponsored by the CRE in 1991 showed that only one-third of the 52 responding organisations monitored test results by ethnic origin; they therefore had no knowledge as to whether or not tests were having an adverse impact on ethnic minorities. It should be noted that if tests can be shown to have adverse impact on ethnic minorities, then it leaves a company open to charges of racial discrimination - viz. the cases against London Underground Limited in 1990 and British Rail in 1991 (both of which were settled out of court).

A smaller survey of 13 companies commissioned by the EOC (1987) discovered similar findings with no monitoring carried out at all in certain organisations. Over half the organisations taking part in the research carried no evaluation or validation of test use. Many had also carried out no job analysis, making it difficult for those organisations to justify test use at all. The EOC guidance for employers (1988) makes the point that it was lack of validation which brought many US employers before the courts in the 1970s and led to the imposition of tight guidelines on employment testing in the USA.

Newell and Shackleton (1994) carried out a survey of how tests were being used in British industry. The results revealed that the BPS and IPM codes of ethics were being

broken by test users in organisations in a number of ways. Although there did not appear to be a problem of tests being used by untrained staff, only one-third of organisations using tests had input from a professionally trained occupational psychologist or consultant. (This is contrary to the guidelines which suggest that policy and organisational decisions about tests should be made with professional advice from a trained occupational psychologist.)

Newell and Shackleton's survey also revealed that a number of organisations were using scores as a definitive judgement on people rather than following the recommendation that tests should be used as the basis for discussion and further exploration.

Finally, only a third of companies consistently gave feedback to candidates, and even fewer made the time for a full discussion of the results. This is contrary to the BPS guidelines which recommends at the very least a debriefing of test results, and extended counselling where possible.

Fletcher (1994) looks at the wider aspects of test use and in particular the part test publishers could (or should) play in research into issues such as equal opportunities and fair testing. He concludes that an analysis of the field of psychometric testing from the perspective of business ethics might illuminate thinking about professional responsibilities and priorities.

Fairness, validity and utility

This complex issue will be briefly addressed here as much of the research into fairness in selection procedures has focused on psychometric tests. Indeed, objectivity and lack of bias (a possible definition of 'fairness') are often cited by organisations as the reason for using tests in selection.

The general conclusion of extensive reviews and meta-analyses (eg reviews in the USA by Muchinsky (1986), Schmitt, Gooding, Noe & Kirsch (1984) and Reilly and Chao (1982); and in the UK by Robertson and Makin (1986)), is that psychometric tests are fairer and have greater validity than many alternative forms of assessment. However, some racial and ethnic minorities do not perform as well as majority applicants on tests. Whatever the reason for this (and as Flaughner (1978) observes, it is likely to be at least partly connected to socio-economic status), the outcome is that if tests scores are the sole basis for selection, ethnic minority members are not likely to be selected at the same rate as the majority group. The difference in selection rate is called adverse impact.

Part of the debate has focused on what is meant by 'fairness'. One widely quoted model of fairness is that of Cleary (1968) who defines a predictor as fair if regression lines are equal for the relevant subgroups. Anastasi (1988), discussing different decision models for test use, indicates that a selection strategy may select individuals solely on the grounds of their predictor criterion scores (the Cleary model) and this strategy will maximise overall criterion performance. But this approach ignores other goals of the selection process. Anastasi points out that other decision models are possible which can have the effect of increasing the selection of people from minority groups. The differences between models are explained by the value judgements implicit in each model - ie a different definition of 'fairness'.

The 'other goals' of selection to which Anastasi refers are influenced to a large extent by the need in the USA to comply with the relevant legislation supporting 'affirmative action'. The effect of testing on minority sub-groups has led the US Federal Government to urge employers to seek alternative selection procedures to testing which are equally valid but have less adverse impact.

A similar view is taken in this country by the CRE. A survey of test practice carried out on behalf of the CRE and published in 1993 revealed a widespread failure to take some of the steps essential for fair test use. In the CRE's view 'evidence of validity is not a sufficient justification for continuing to use tests that have been found to have adverse impact, particularly when alternative selection measures may be available' (Towards Fair Selection, p.12).

Schmitt (1989) states that the primary reason for developing personnel selection procedures and conducting personnel selection research is the prediction of employee performance and the selection of those individuals whose predicted performance is best. Research into utility measurement (Boudreau (1983); Cascio (1982)) indicates that valid selection procedures can add substantially to organisational productivity.

However, affirmative action in organisations might involve special recruitment efforts directed at applicants from minority groups and special treatment in hiring decisions. In the USA, affirmative action has largely led to the abandonment of testing for selection in favour of other methods. Affirmative action in effect impacts directly on the relationship between predictor and criterion, and thus affects validity. As Hunter and Hunter (1984) state, high selection validity translates into considerable financial benefits for most organisations. Schmitt (1989) points out that this means that the goals of productivity and affirmative actions are often in conflict.

The difficulty of finding a suitable alternative to psychometric tests is shown by Reilly and Chao (1982) who carried out a review of the validity and fairness of some alternative employee selection procedures. They concluded that only biodata and peer evaluations had research evidence showing validity equal to that of tests, and it was not clear from the research if either of those techniques would reduce adverse impact. Reilly and Chao point out that test fairness research has, with few exceptions, supported the predictability of minority groups even though adverse impact exists

(Hunter, Schmidt & Hunter, 1979). They say that there is no reason to expect alternative predictors to behave differently.

Hunter and Hunter (1984) support the continued use of tests, given that for entry level jobs ability tests have the highest validity (.53). Using other predictors can involve considerable costs to the organisation in lost productivity. Hunter and Hunter discuss various ways of addressing the problem of balancing minimum adverse impact and maximum utility, and conclude that the optimum strategy is to hire on ability by quota.

A further difficulty in the area of fairness and validity is described by Robertson (1994). One strategy for increasing the test score of adversely affected minority groups is to give those groups the opportunity to practise tests and receive coaching prior to taking tests for selection. One aspect of this is that some individuals and sub-groups may have greater access to coaching programmes than others - an issue of 'fairness'. There is a further issue of validity. Robertson points out that coaching and practice **may** provide more accurate assessment, but it may inflate scores on the assessment procedure while not improving a candidate's position on the underlying construct. Construct validity is therefore damaged. This could ultimately lead to damage to criterion-related validity if it happens to some candidates but not to others. Robertson comments that the effects of coaching and practice on validity are virtually ignored and this may be an important area for further research.

This section has so far focused on the use of tests with ethnic minorities. The adverse impact of tests on women vis-a-vis men has received far less research attention.

Kaplan (1985) in his review of the use of psychological tests, concludes that the problem of gender bias arising from tests content is either non-existent or insignificant compared with the gains to fairness from using tests as opposed to not using them at all. A study by Pearn, Kandola and Mottram (1987), which looked at the impact of selection testing on the employment opportunities of women and men, supports that

view, although they found a number of examples of both good and poor testing practice in the organisations they studied. They make the important point (which applies equally to minority groups as well as women) that while it is good practice to create valid and fair selection procedures, if job performance assessment and promotion procedures are themselves subject to bias, disadvantaged groups will remain in the lower grades of the organisations. Testing, in other words, is only one aspect of the assessment procedures, and the concept of 'fairness' needs to be extended to the system as a whole.

Combining test results with other selection methods

Within the domain of occupational testing there is a debate regarding combining the scores from psychometric testing with the results obtained from other selection procedures. To a degree this debate is related to the question of 'fairness' and the choice of decision models identified by Anastasi (1988) as being available in the selection process.

Concern about the impact of selection procedures such as psychometric testing on women and minorities leads Pearn (1989) to state that the best strategy for employers is to use selection instruments which combine moderate to high validity with small or moderate subgroup differences. In their separate guidelines for test use in employment settings, the EOC (1989), the CRE (1992) and the IPM (1988) all state that tests should not be used in isolation but combined with data from other selection techniques, although they give no guidance on how this should be done.

Toplis, Dulewicz and Fletcher (1989) discuss how much weight should be given to test scores compared with other sources of assessment data (such as interviews and references). They say that the evidence on the validity of test results compared with the validity of the interview and other selection methods makes a strong case for

putting the greatest weight on test results, although they still advocate gathering assessment information from additional sources.

Kline (1993) takes a more decided view. He states that it is generally true to say that psychometric tests are the most valid methods of assessment. 'Where a valid test exists, such as in the field of ability, motivation and personality, there is no point in trying to use other methods of assessment' (p.393). He says that interviews, for instance, frequently 'add in' errors to the selection process, because they have lower validity. Hunter and Hunter (1984) follow a similar line of thought. Starting from the premise that for entry-level hiring, ability tests are the best predictor of performance, Hunter and Hunter explore the question of whether other predictors can be used alongside ability tests to increase validity. They make a similar point to Kline. 'Many companies now use several predictors for selection. However, predictors are not being combined in accordance with the actual validities of predictors, but are weighted equally. Under these conditions, the validity of the total procedure can be lower than the validity of the best single predictor' (Hunter & Hunter, 1984, p.91). This has clear implications for utility. However, Hunter and Hunter do concede that adding an appropriate second predictor could increase validity if it were used with the weighting appropriate to its validity. If the second predictor had less adverse impact, the composite strategy may have higher validity and lower adverse impact than the use of ability tests alone. The difficulty here, as Hunter and Hunter neatly encapsulate, is that alternative predictors have only been studied in isolation. Correlations between alternative predictors are virtually unknown and therefore it has not been possible to carry out generalised multiple regression.

Personality Questionnaires

Anastasi (1988) states that the use of personality questionnaires is beset with special difficulties over and above the common problems encountered in all psychometric tests. She mentions the problems of faking responses, the changing nature of

behaviour over time which is greater than that measured by ability tests, and the greater situational specificity of responses. Cronbach (1990) refers to the fact that personality questionnaires are more often regarded as intrusive than ability tests. The particular difficulties associated with personality questionnaires are recognised by the BPS, which requires further training to be undertaken by those using personality questionnaires, over and above that provided by the BPS Level A Certificate of Competence, to enable the user to understand more fully the background and processes involved in personality testing and test construction before attempting to evaluate and interpret results. A Level B Certificate is being developed by the BPS and will cover the competencies necessary for personality testing.

Muchinsky (1986) refers to the high degree of controversy over the use of personality questionnaires for personnel selection - in particular because of 'their questionable validity, their perceived capacity to invade privacy and their lack of face validity' (p.44). Muchinsky says that in the most part the reported validity coefficients for personality questionnaires have been unimpressive. Ghiselli (1973) found an average validity coefficient of .28 for personality questionnaires; Guion and Gottier (1965) reviewed the use of personality questionnaires in industry over a 12 year period and concluded that the evidence did not support their use for the making of selection decisions. Meta-analyses by Hunter and Hunter (1984) and Schmitt, Gooding, Noe and Kirsch (1984) revealed validity coefficients of .10 and .15 respectively for personality questionnaires. Schmitt (1989) confirmed the low validity of personality questionnaires but also indicated that their results showed small sub-group differences (ie they have low adverse impact).

However, Tett, Jackson and Rothstein (1991) used meta-analysis to assess, *inter alia*, the overall validity of personality measures as predictors of job performance. In particular they found that if personality measures were related to job analysis (ie if job analysis was used to select trait scales) overall validity co-efficients of .38 could be

obtained. Tett et al believed their results supported the potential value of undertaking personality-based conceptual analyses of job performance, and that personality measures have a place in personnel selection research.

It should be said that research into personality questionnaires has not been without its critics, most notably in the UK, Blinkhorn and Johnson (1990), who said there was little evidence that even the best personality questionnaires predicted job performance and a good deal of evidence of poorly understood statistical methods being used. In a recent article, Johnson and Blinkhorn (1994) confirm their earlier opinion and in particular criticise Tett et al's 1991 study on the grounds of poor statistical method. They conclude that: 'There is no body of public knowledge relating scores on personality tests taken as part of a selection procedure to objective criteria of later performance sufficient to form a basis for routine use of the tests, despite 40 or more years of research' (p.170).

Recently a more widely accepted taxonomy for classifying personality traits has been developed. The 'Big Five' personality dimensions (Extraversion, Emotional Stability, Agreeableness, Conscientiousness and Openness to Experience) were studied by Barrick and Mount (1991) against three job performance criteria (job proficiency, training proficiency and personnel data) for five occupational groups. The results indicate that one dimension of personality, Conscientiousness, showed consistent relations with all job performance criteria for all occupational groups; Extraversion was a valid predictor for two occupations involving social interactions, management and sales; and Openness to Experience and Extraversion were valid predictors of training proficiency criterion across occupations.

More recently, Robertson and Kinder (1993) used *a priori* hypotheses to identify which personality scales should show criterion-related validity in specific criterion areas. Their research revealed criterion-related validity for personality variables of up

to .33 with the higher values for criteria such as creativity, analysis and judgement. They believe the results suggest that personality variables add unique criterion-related information beyond that provided by ability alone.

Research question and hypotheses

From initial discussions with the company, a research question and hypotheses were formulated.

Research question

How are psychometric tests used by those involved in selection and assessment procedures in the company?

Hypotheses

H1 1) The use of psychometric testing in selection and assessment procedures varies between the district offices of the company.

H1 2) The confidence and experience of the Human Resources Advisers and District Team Managers in psychometric testing determines the use of psychometric testing in selection and assessment procedures.

H1 3) Where psychometric testing is used, the selection of a test is determined by 'common practice' rather than by its suitability for the specific applicants and the job.

H1 4) Where psychometric testing is used, the weighting given to the test battery varies between the district offices of the company.

Model of good testing practice

Although guidelines regarding the use of psychometric testing have been produced by a number of organisations (viz., the British Psychological Society, the Institute of

Personnel Management, the Equal Opportunities Commission and the Commission for Racial Equality), not all the issues which are likely to be faced by businesses using psychometric testing are addressed by any one set of guidelines. Indeed, certain of the guidelines (for instance those produced by the BPS) are not specifically aimed at companies, and are of a more general nature.

For this reason a model of good practice in the use of psychometric tests in organisations was designed (see Appendix A), based on a study of the relevant literature.

Research design

The purpose of the research project was: a) to evaluate the current use of psychometric tests in the company against the model of good testing practice; and b) to test the hypotheses. In essence, the study was one of process evaluation, concerned with how psychometric tests were being used by those involved in selection and assessment procedures. To this end, the utility of the project was a major consideration; the intention from the outset was to make recommendations if necessary, to improve the use of testing in the organisation.

The unit of analysis for the project was the individual employee and the group.

SECTION 2.

Methodology

Populations

Three populations were identified for the project following initial discussions with key HR personnel involved in psychometric testing in the company. The populations were: senior managers who were part of the district management teams (DMTs); the human resources advisers (HRAs); and the test administrators (TAs). In each case, questionnaires were distributed to the total population; this comprised 104 DMTs, 19 HRAs and 8 TAs.

Questionnaires

Three questionnaires were developed for each of the populations; these were based on the model of good testing practice, but included additional questions particularly pertinent to the company's experience of testing. The terminology used in the questionnaires was modified to suit that in current use in the company; for instance, 'job analysis' was replaced by 'job description'. The questionnaires comprised 33 items for DMTs; 35 items for HRAs; and 24 items for TAs (see Appendices B, C and D). Space was designated at the end of the questionnaires for participants to add extra relevant information if they so wished.

Pilot study

The questionnaire for the DMTs was piloted with the senior managers in one of the company's district offices. The questionnaires for the HRAs and the TAs were not piloted because of the small populations in each case.

Data collection methods

Many of the employees taking part in the survey work away from their office base. Therefore the questionnaires were sent to the home addresses of the populations, with instructions on how to complete them and an indication of how long completion would

take. The purpose of the questionnaires was also explained. The questionnaires were accompanied by a letter from the company's Human Resources Director, who gave the background to the project and reinforced the purpose. It was requested that the questionnaires be returned by a specific date to a named person in the HR Department.

Additional data were obtained by semi-structured interviews by telephone with 10 senior managers at three of the district offices.

Data analysis methods

The quantitative data were analysed using frequency count and descriptive statistics.

The qualitative data obtained from additional comments in the questionnaires and from the telephone interviews were grouped into themes corresponding to the main sections of the questionnaire.

SECTION 3.

Results

From the 104 questionnaires distributed to the district team managers, 72 were returned (a response rate of 69%). 50 respondents used test results in making staffing decisions. Of that number, 82% found tests useful, and 18% said they sometimes found them useful. Of the 22 respondents who did not use tests, the vast majority (18) cited low staff turnover and low need for recruitment as the reason for not using tests. Three respondents did not use tests because they had no staff responsible to them. Only one respondent had not considered using tests and was unconvinced of their value. Eight of the respondents who were not using tests volunteered the information that if they were recruiting staff, they would consider using them.

From the 19 questionnaires distributed to the human resources advisers, 11 were returned (a response rate of 58%).

From the 8 questionnaires distributed to the test administrators, 7 were returned (a response rate of 88%).

Note: i) Each Table title also indicates the source of information (ie questionnaire type - HRA, DTM or TA - and relevant question number).

ii) For the purposes of analysis, unless otherwise indicated, the response 'sometimes' is regarded as a positive response in the Tables.

Evaluation of the results against the model of good testing practice

This section evaluates the results obtained from the questionnaires and telephone interviews against the model of good testing practice (Appendix A). The salient points from the model are listed by the relevant results.

Job Analysis

Good testing practice:

- i) A job analysis should be conducted before testing for recruitment and selection.*
- ii) The tasks and activities of the job and the person specification should be ascertained.*

Table 1. Conducting a job analysis. (HRA: Q.10, 11 and 12; DTM: Q.11, 12 and 13)

	HRA's n = 11			DTMs n = 50		
	Yes	No	Sometimes	Yes	No	Sometimes
Revision of job description	64%	0%	36%	50%	2%	48%
Main tasks identified	91%	0%	9%	98%	2%	0%
Personal skills identified	91%	0%	9%	82%	18%	0%

The results indicate that on at least half the occasions when recruitment and selection is taking place, the job description is revised. Only in a very small number of cases is it never revised. The results also indicate that it is likely that the main tasks of the job, and the personal skills required, are identified, although there is some indication that personal skills are less likely to be identified by the DTMs.

Telephone interviews with the DTMs indicated that the reason job descriptions were sometimes not revised was because the main tasks and duties had not changed.

Test Administration

1. Good testing practice:

Tests should only be administered by those trained in their use.

The results from the TAs indicated that they had all received appropriate training in test administration.

2. *Good testing practice:*

The tests should not be administered by post but should be held in a designated location.

Table 2. Location of testing. (TA: Q.7)

n = 7	Yes	No	Sometimes
By post	0%	100%	0%
At work	29%	14%	57%
Specified location	86%	0%	14%

The results show that no tests are administered by post. Most tests are administered at a specified location.

3. *Good testing practice:*

Candidates should be:

- i) told in advance the tests to be used .*
- ii) told in advance the purpose of the tests.*
- iii) asked if they have taken any similar tests recently.*
- iv) sent test descriptions and sample sheets.*
- v) given the opportunity to practise tests.*

Table 3. Briefing and preparation of candidates. (TA: Q.10, 11, 12 & 13)

n = 7	Yes	No	Sometimes
Test purpose explained	57%	0%	43%
Which tests to be used	71%	0%	29%
Similar tests recently taken	57%	14%	29%
Samples of tests sent	0%	57%	43%
Opportunity to practise	0%	57%	43%

The results indicate that on over half the occasions when tests are used, the purpose of the tests are explained, candidates are informed of the tests to be used, and are asked if

they have taken any similar tests recently. However, the results show that it is not usual custom to send out samples of tests or to give candidates an opportunity to practise tests in advance.

Extra information supplied on the questionnaires indicated that checks were made only on internal candidates to ascertain if they had taken similar tests recently. One DTM commented in telephone interview that he felt that some candidates became very anxious before taking tests and suggested that the opportunity to take a 'mock' test may help in this respect.

4. Good testing practice:

- i) A tape-player should be used to relay instructions prior to testing.*
- ii) Instructions should also be handed out in the form of a briefing sheet.*
- iii) Candidates should be given the opportunity to ask questions during briefing.*
- iv) The physical conditions for testing should be suitable.*
- v) Flexibility may be needed on timed tests to avoid adverse impact on minorities.*

Table 4. Conditions of testing. (TA: Q.16, 17, 18 & 19)

n = 7	Yes	No	Sometimes
Instructions - from script	100%	0%	0%
Instructions - orally	14%	72%	14%
Instructions - from handout	0%	100%	0%
Instructions - from tapeplayer	0%	100%	0%
Opportunity for questions	100%	0%	0%
Timed tests - timekeeping	100%	0%	0%
Suitable environment	72%	28%	0%

The results indicate that instructions are always read from a script. They may be supplemented by unscripted oral instructions. Neither handouts nor tape-players are used. The candidates always have an opportunity to ask questions. Time keeping is

strictly adhered to. Although the environment for testing is often suitable, on a substantial number of occasions it is not.

5. Good testing practice:

- i) Care should be taken over the security of test materials.*
- ii) Test results should be treated as confidential documents.*
- iii) Access to test results should be restricted to accredited test users.*

In response to questions regarding the security of test materials and records (TA: Q.14, 23 & 24), all the respondents indicated that the test material are kept secure, the test records are treated as confidential documents, and access to test results is restricted.

Use of test results

1. Good testing practice:

- i) Test results should not be used on their own but alongside other methods of selection.*
- ii) Ability and aptitude tests can be used to screen candidates before short-listing. Personality tests should not be used to screening. They should be used to provide additional criterion-related information.*

Table 5. Use of ability test results. (HRA: Q.15; DTM: Q.16)

	HRAs n = 11			DTMs n = 49		
	Yes	No	Sometimes	Yes	No	Sometimes
Initial screening	0%	82%	18%	6%	63%	31%
Alongside other methods	82%	0%	18%	90%	0%	10%

Table 6. Use of aptitude test results (HRA: Q.21; DTM: Q.22)

	HRAs n = 9			DTMs n = 32		
	Yes	No	Sometimes	Yes	No	Sometimes
Initial screening	11%	44%	45%	0%	69%	31%
Alongside other methods	78%	0%	22%	88%	0%	12%

Table 7. Use of personality test results. (HRA: Q.27; DTM: Q.28)

	HRAs n = 11			DTMs n = 49		
	Yes	No	Sometimes	Yes	No	Sometimes
Initial screening	0%	73%	27%	4%	73%	23%
Alongside other methods	91%	0%	9%	86%	0%	14%

The results indicate that aptitude tests (Table 6) are occasionally used for screening applicants, particularly by HRAs. There is less likelihood of ability tests (Table 5) being used for screening.

Although the results indicate that personality tests (Table 7) are rarely used regularly for screening applicants, there is some evidence that they are sometimes used in this way by both HRAs and DTMs.

The results also indicate that the test results are used in most cases alongside other methods of selection. Where the response 'sometimes' is recorded, this may indicate that test results are not used at all, rather than that they are used in isolation. How test results are weighted alongside other selection methods is indicated in Tables 8 to 10 below.

In additional information supplied on the questionnaire, 50% of the DTMs who use tests indicated that they rely on the HRAs for expert advice on choice, use and interpretation of tests. During the telephone interviews, 3 DTMs said they would use tests for screening if there were many candidates for a job, the job had

clear duties demanding specific technical competences, and if it was difficult to choose between candidates on the basis of qualifications and experience.

Table 8. Weighting given to ability tests - internal and external applicants. (HRA: Q.18 & 19; DTM: Q.19 & 20)

		HRAs n = 10			DTMs n = 43		
Applicant							
		Yes	No	Sometimes	Yes	No	Sometimes
zero weighting	Int.	10%	70%	20%	16%	74%	10%
	Ext.	0%	90%	10%	12%	5%	83%
10% weighting	Int.	10%	60%	30%	21%	47%	32%
	Ext.	10%	50%	40%	12%	64%	24%
20% weighting	Int.	40%	30%	30%	12%	56%	32%
	Ext.	30%	30%	40%	21%	49%	30%
30% weighting	Int.	0%	80%	20%	7%	72%	21%
	Ext.	10%	70%	20%	14%	65%	21%

Table 9. Weighting given to aptitude tests - internal and external applicants. (HRA: Q.24 & 25; DTM: Q.25 & 26)

		HRAs n = 9			DTMs n = 28		
Applicant							
		Yes	No	Sometimes	Yes	No	Sometimes
zero weighting	Int.	11%	89%	0%	14%	79%	7%
	Ext.	0%	88%	12%	14%	75%	11%
10% weighting	Int.	0%	67%	33%	21%	46%	33%
	Ext.	12%	44%	44%	14%	61%	25%
20% weighting	Int.	33%	33%	34%	11%	54%	35%
	Ext.	12%	44%	44%	14%	58%	28%
30% weighting	Int.	0%	89%	11%	7%	68%	25%
	Ext.	24%	44%	32%	14%	58%	28%

Table 10. Weighting given to personality tests - internal and external applicants.

(HRA: Q.29 & 30; DTM: Q.30 & 31)

		HRAs n = 10			DTMs n = 43		
Applicant							
		Yes	No	Sometimes	Yes	No	Sometimes
zero weighting	Int.	10%	70%	20%	21%	70%	9%
	Ext.	10%	80%	10%	19%	74%	7%
10% weighting	Int.	10%	50%	40%	19%	53%	28%
	Ext.	0%	60%	40%	9%	63%	28%
20% weighting	Int.	30%	30%	40%	14%	56%	30%
	Ext.	30%	30%	40%	21%	47%	32%
30% weighting	Int.	0%	70%	30%	7%	77%	16%
	Ext.	10%	60%	30%	7%	67%	26%

Tables 8 to 10 show the wide range of weighting given to all tests amongst both HRAs and DTMs.

In additional information given on the questionnaire, one HRA said that the weightings to be applied were not usually specified. In the telephone interviews, 4 of the DTMs said that the weighting given varied, describing it as being 'ad hoc', 'subjective', 'partly objective, partly subjective'. Two commented on the questionnaire that they used specialist advice from the HRAs on weighting to be applied. Four DTMs said during telephone interview that the weighting given to different tests would depend upon the nature of the job. Two DTMs said that they may rely more heavily on tests with external candidates because of the lack of other job-related data.

2. Good testing practice:

The use of minimum cut-off scores should be avoided. Scores should be used as a guide.

Figure 1. Use of minimum cut-off scores in ability tests. (HRA: Q.16; DTM: Q.17)

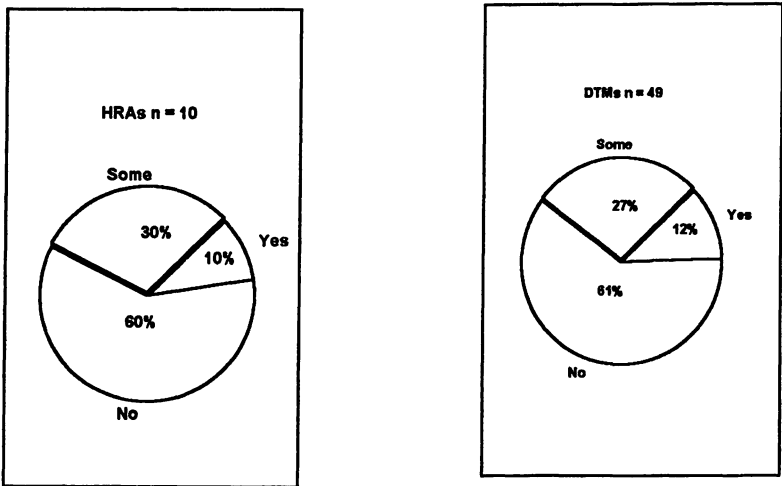
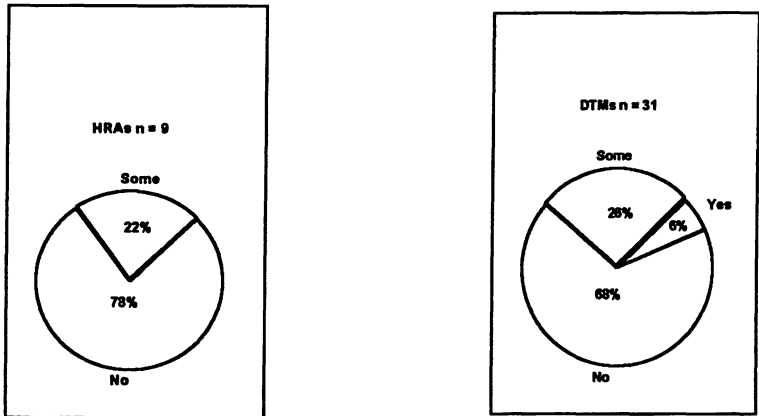


Figure 2. Use of minimum cut-off scores in aptitude tests (HRA: Q.22; DTM: Q.23)



The results show that minimum cut-off scores are not normally used, although there is some evidence that cut-off scores are more likely to be used in relation to ability test results by both HRAs and DTMs.

3. *Good testing practice:*

Test results should be used to instigate further discussion in interview.

Table 11. Use of ability test results to rank candidates and for exploration in interview. (HRA: Q.17; DTM: Q.18)

	HRAs n = 11			DTMs n = 49		
	Yes	No	Sometimes	Yes	No	Sometimes
To rank candidates	0%	36%	64%	20%	45%	35%
Exploration in interview	55%	0%	45%	51%	18%	31%

Table 12. Use of aptitude test results to rank candidates and for exploration in interview. (HRA: Q.23; DTM: Q.24)

	HRAs n = 9			DTMs n = 32		
	Yes	No	Sometimes	Yes	No	Sometimes
To rank candidates	22%	33%	45%	34%	38%	28%
Exploration in interview	78%	0%	22%	53%	31%	16%

Table 13. Use of personality test results to compare with personal skills and attributes identified in the job description, and for exploration in interview. (HRA: Q.28; DTM: Q.29)

	HRAs n = 10			DTMs n = 49		
	Yes	No	Sometimes	Yes	No	Sometimes
To compare with job description	60%	0%	40%	63%	16%	21%
Exploration in interview	60%	0%	40%	61%	16%	23%

The results indicate that while ability test results (Table 11) are not used regularly to rank candidates, they are used sometimes, particularly by HRAs. There is a higher likelihood of aptitude test results (Table 12) being used to rank candidates. Personality test results (Table 13) are used frequently to compare the candidates profile with the person specifications identified in the job description.

Tables 11, 12 and 13 show that the results from all three types of test are likely to be used for exploration in interview, although there is some evidence that DTMs are less likely to use the results in this way. Responses from the questionnaire (TA: Q.15)

indicated that test results were nearly always (84%) passed to the test interpreter on the same or following day.

One DTM commented during telephone interview that the results from the tests were used by the HRA to formulate probe questions for use in interviewing the candidate. Three DTMs said that they did not always receive the test interpretation in time to use the results in interviews with candidates. Four DTMs commented on the questionnaire that they used the test results to verify other parts of the selection process. An HRA and said that personality tests were used to identify warning signals which might affect a candidates job performance or 'fit' in the team. The use of personality tests to identify whether or not a candidate would 'fit in' was mentioned by 4 DTMs. One DTM commented that if a person was perceived not to 'fit in' during the interview, the test results didn't matter. One HRA said that DTMs seemed to want general information about candidates from the test results rather than specific attributes related to the job.

De-briefing and feedback to candidates

Good testing practice:

At the very least, candidates should be offered a de-briefing session. Ideally, a confidential session to give feedback and discuss test results should be given.

Figure 3. Giving of feedback by HRAs. (HRA: Q.32)

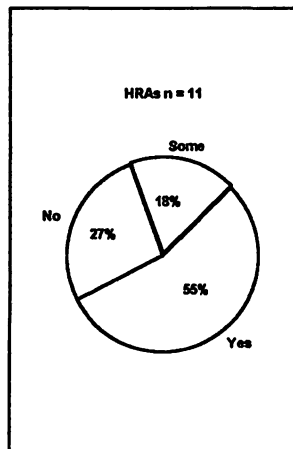


Table 14. Giving of feedback by DTMs or HRAs. (DTM: Q.33)

DTM n = 50	Yes	No	Sometimes
Feedback given by DTM	26%	40%	34%
Feedback given by HRA	58%	8%	34%

Table 15. De-briefing and feedback. (TA: Q.20, 21 & 22)

n = 7	Yes	No	Sometimes
Debriefing offered	57%	29%	14%
Confidential feedback offered	57%	14%	29%
Written report given	29%	57%	14%

The results show that in just over half the occasions where tests are used, the candidates are given feedback. The results indicate that this is a confidential session (Table 15). Table 14 shows that HRAs generally feed back the results, although DTMs are sometimes involved.. Written reports are not given to candidates on a regular basis (Table 15).

In extra information on the questionnaire, all the TAs commented that only internal candidates were offered de-briefing and feedback, although during

telephone interview a DTM said that he offered feedback to both external and internal candidates. Of the DTMs who are involved in giving feedback, 3 gave it jointly with the HRA; the HRA fed back the test results and the DTM gave more general feedback. One DTM gave feedback on his own, including test results, although he said that more recently the feedback had been carried out by an HRA. During the telephone interview, one DTM commented that he thought (internal) candidates did not get 'a good deal' out of the testing process. He said that feedback was given by an HRA, whom the candidate would be unlikely to know, and he felt the manager should be present too. He also felt that there was little follow-up with test results to aid individual development, particularly for those at lower levels in the company who were not included in Development Centres.

Monitoring and validation

Good testing practice:

- i) Validation studies should be carried out in the organisation to establish if the test results are good performance predictors.*
- ii) the relationship between test and job should be regularly monitored for relevance.*
- iii) the test results should be monitored for unfair discrimination against subgroups.*

Table 16. Monitoring test use. (HRA: Q.33, 34 & 35)

HRA's n = 11	Yes	No	Sometimes
Test & work/job performance	0%	73%	27%
Changes in tasks	27%	46%	27%
Test results & discrimination	9%	91%	0%

The results indicate that the relationship between test results and job performance is not usually evaluated, nor are the results normally monitored for effects such as

adverse impact. There is some indication that jobs are monitored for changes in tasks which may affect test suitability.

In extra information on the questionnaire, a DTM commented that he thought tests have proved to be reasonably accurate and a valuable support to the assessment process. However, he said that he didn't think that the tests used had changed much, despite changes in management culture and business objectives.

Hypotheses testing

Hypothesis 1

H0: There is no difference between the district offices of the company in the use of psychometric testing in selection and assessment procedures.

H1: The use of psychometric testing in selection and assessment procedures varies between the district offices of the company.

Table 17. Percentage of team managers within each district using tests for staffing decisions. (DTM: Q.6)

District No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n =	5	4	4	5	4	4	4	5	5	7	4	5	4	6	6
Percentage using tests	80%	100%	100%	60%	75%	75%	75%	60%	80%	43%	25%	40%	75%	67%	100%

Table 18. Percentage of team managers within each district using tests for purposes other than selection. (DTM: Q.32).

District No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n =	4	4	4	3	3	3	3	3	4	3	1	2	3	4	6
Promotion potential	75%	50%	25%	33%	33%	33%	33%	67%	100%	67%	0%	50%	33%	75%	67%
Training needs	25%	0%	25%	0%	67%	67%	33%	33%	75%	0%	100%	0%	0%	100%	33%
Career development	50%	25%	50%	33%	0%	33%	33%	33%	100%	33%	0%	50%	0%	100%	33%
Restructuring	100%	50%	50%	33%	33%	67%	0%	33%	50%	67%	100%	0%	33%	100%	50%

Table 17 indicates a range from 25% to 100% in terms of test use; Table 18 indicates a range from 0% to 100% in terms of test use for purposes other than selection. Because of the small sample sizes it has not been possible to apply a statistical test. Therefore, although the results appear to demonstrate that test use varies in both intensity and nature from district to district, and the alternative hypothesis is substantiated, the results should be treated with caution.

Hypothesis 2

H0: There is no difference in the use of psychometric testing and selection procedures between those human resources advisers and district team managers who indicate confidence and experience and those who do not.

H1: The confidence and experience of human resources advisers and district team managers determines the use of psychometric testing in selection and assessment procedures.

Table 19. Frequency of test use according to HRA training and confidence. (HRA: Q. 6, 8 & 9)

n = 11		Weekly	Monthly	Less
Trained	Very confident	2	0	0
Untrained	Very confident	0	0	0
Trained	Fairly confident	1	1	1
Untrained	Fairly confident	0	0	4
Trained	Not confident	0	0	0
Untrained	Not confident	0	0	2

Table 20. Confidence levels and experience of test use among district team managers. (DTM: Q. 7 & 9)

n = 50	Average no of years test use
Always confident (n = 34)	4.9 years
Sometimes (n = 11)	4.1 years
Never (n = 5)	3.2 years

Note: Information regarding the frequency of use of tests for selection purposes by district team managers was restricted because of the extremely low levels of recruitment activity at present.

While Table 19 indicates that there are HRAs untrained in test use who are fairly confident in their use, the figures indicate that more frequent use of tests involves trained and confident HRAs. Table 20 indicates that, among DTMs, those with longer experience tend to be more confident in using tests. As explained above, it was not possible to relate this to test use. Low sample sizes meant that it was not possible to apply a statistical test. The alternative hypothesis appears to be partially substantiated, but caution needs to be exercised in the reliability of this result.

Hypothesis 3

H0: Where psychometric testing is used, the selection of a test is determined by its suitability for the specific applicants and the job.

H1: Where psychometric testing is used, the selection of a test is determined by 'common practice' rather than by its suitability for the specific applicants and the job.

Table 21. Selection of ability tests by HRAs and DTMs. (HRA: Q.14; DTM: Q.15)

	HRA	n = 11	DTM	n = 49
	Yes	No	Yes	No
Always use same tests	0%	100%	16%	84%
Vary tests - seniority	100%	0%	78%	22%
Vary tests - task/activities	91%	9%	73%	27%

Table 22. Selection of aptitude tests by HRAs and DTMs. (HRA: Q. 20; DTM: Q.21)

	HRA	n = 9	DTM	n = 32
	Yes	No	Yes	No
Always use same tests	11%	89%	6%	94%
Vary tests - seniority	78%	22%	88%	12%
Vary tests - task/activities	89%	11%	88%	12%

Table 23. Selection of personality tests by HRAs and DTMs. (HRA: Q. 26; DTM: Q. 27)

	HRA	n = 10	DTM	n = 44
	Yes	No	Yes	No
Always use same tests	80%	20%	50%	50%
Vary tests - seniority	20%	80%	43%	57%
Vary tests - nature of job	20%	80%	48%	52%

Tables 21 and 22 indicate that in both ability and aptitude tests the greater proportion of HRAs and DTMs vary the tests according to the seniority of the job and/or the tasks and activities of the job. The situation is less clear for personality tests (Table 23),

although this may be a feature of the limited choice of personality tests in use in the company. The alternative hypothesis appears not to be substantiated for ability and aptitude tests; for personality tests it appears to be substantiated, subject to the rider mentioned above. However, because of small sample sizes it has not been possible to apply a statistical test; the results should therefore be treated with caution.

Hypothesis 4

H₀: There is no difference between the district offices of the company in the weighting given to the test battery.

H₁: Where psychometric testing is used, the weighting given to the test battery varies between the district offices of the company.

Table 24. The weighting given to ability tests for internal and external applicants by district offices. (DTM: Q. 19 & 20)

District No		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n =		4	3	1	3	3	3	3	2	4	3	1	2	2	4	6
	Appl.															
zero weighting	Int.	25%	67%	0%	0%	0%	67%	67%	50%	25%	0%	100%	50%	50%	0%	0%
	Ext.	25%	0%	0%	0%	0%	33%	67%	50%	0%	0%	100%	50%	50%	0%	0%
10% weighting	Int.	50%	67%	0%	67%	67%	67%	0%	50%	50%	33%	100%	50%	50%	75%	50%
	Ext.	25%	67%	0%	33%	33%	0%	0%	50%	50%	33%	100%	50%	50%	50%	33%
20% weighting	Int.	75%	0%	100%	33%	67%	33%	0%	50%	25%	33%	100%	0%	50%	50%	67%
	Ext.	50%	67%	100%	67%	67%	67%	0%	50%	25%	33%	100%	0%	50%	50%	67%
30% weighting	Int.	25%	0%	100%	33%	67%	0%	33%	0%	25%	0%	100%	0%	50%	25%	33%
	Ext.	75%	0%	100%	0%	67%	33%	33%	0%	25%	0%	100%	0%	50%	25%	50%

Table 25. The weighting given to aptitude tests for internal and external candidates by district offices. (DTM: Q. 25 & 26)

District No		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n =		4	0	0	2	2	1	3	2	3	2	1	1	1	2	4
	Appl.															
zero weighting	Int.	25%	-	-	0%	0%	0%	100%	50%	0%	0%	100%	0%	0%	0%	0%
	Ext.	25%			0%	50%	0%	100%	50%	0%	0%	100%	0%	0%	0%	0%
10% weighting	Int.	50%	-	-	50%	50%	100%	0%	0%	33%	50%	100%	100%	100%	100%	75%
	Ext.	25%			50%	50%	0%	0%	0%	33%	50%	100%	100%	100%	50%	50%
20% weighting	Int.	75%	-	-	50%	50%	0%	0%	0%	33%	50%	100%	0%	100%	50%	75%
	Ext.	50%			50%	50%	100%	0%	0%	33%	50%	100%	0%	100%	0%	75%
30% weighting	Int.	50%	-	-	50%	100%	0%	0%	50%	0%	0%	100%	0%	100%	0%	25%
	Ext.	50%			0%	100%	0%	0%	50%	0%	0%	100%	0%	100%	50%	50%

Table 26. The weighting given to personality tests for internal and external candidates by district offices. (DTM: Q.30 & 31)

Personality Tests - Internal and External Applicants

District No		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n =		4	3	1	3	3	2	3	3	4	3	1	2	3	4	6
	Appl.															
zero weighting	Int.	25%	67%	0%	0%	0%	50%	100%	33%	0%	0%	100%	50%	33%	50%	0%
	Ext.	25%	0%	0%	0%	0%	0%	100%	33%	0%	0%	100%	50%	33%	25%	0%
10% weighting	Int.	75%	33%	0%	67%	67%	50%	0%	33%	25%	33%	100%	50%	33%	50%	50%
	Ext.	50%	33%	0%	33%	33%	50%	0%	33%	50%	33%	100%	50%	33%	50%	50%
20% weighting	Int.	50%	33%	100%	33%	67%	50%	0%	33%	50%	33%	100%	0%	33%	50%	50%
	Ext.	75%	67%	100%	67%	67%	0%	0%	33%	50%	33%	100%	0%	33%	50%	100%
30% weighting	Int.	25%	0%	100%	0%	67%	0%	0%	33%	25%	0%	100%	0%	33%	0%	33%
	Ext.	50%	0%	100%	33%	100%	50%	0%	33%	25%	0%	100%	0%	33%	25%	50%

The results in Tables 24, 25 and 26 indicate a wide difference in the weightings applied to ability, aptitude and personality tests by district offices. The alternative hypothesis appears to be substantiated. However, because of the small sample size it has not been possible to apply a statistical test. Therefore the results should be treated with caution.

SECTION 4.

Discussion

In discussing the results of the survey, two factors must be taken into account; firstly, the low sample sizes, particularly those obtained at group (district) level; and secondly, the current lack of recruitment in the company, which has affected the opportunity to use tests in selection. Both these factors have implications for the reliability of the results.

The results indicate a great deal of support in favour of using psychometric testing, both among those who currently use testing, and those who, for various reasons, do not. There is a wide variation in the way tests are used in the company, both between districts, and between individuals. The main use for testing is in selection processes; the use of testing for other purposes, such as in assessing employees for promotion potential, for training needs or for career development, is much more varied.

The HR Department seems to have largely succeeded in its aim to present itself as the in-house specialist in testing; it employs a qualified occupational psychologist, it has an identified person within the Department as the company expert in psychometric testing, and it has a number of HR advisers who are fully trained in psychometric testing. As the results indicate, half the DTMs volunteered additional information on the questionnaire in which they acknowledged that they relied on HRAs to provide expert advice on choice, use and interpretation of tests. However, the relationship between the HR Department and the DTMs in the use of tests appears to be solely based on shared understanding; despite 10 years of test usage, no policy document has been drawn up outlining the use of psychometric testing in the company.

Job analysis is fundamental to the selection of appropriate psychometric tests. Part of the process of choosing tests involves job analysis to identify important job components. The results of the study indicate that the company places some

importance on carrying out a revision of the job description prior to selection and recruitment. The job description usually identifies the main tasks of the job, and the personal skills required to carry out the job. There is also some evidence that jobs are monitored for changes in tasks which may affect test suitability. As Bethell-Fox (1989) states, it is especially important, at a time when great emphasis is placed on equal opportunities in employment, that organisations are able to demonstrate that the tests they use are clearly linked to the requirements of particular jobs. Careful job analysis can achieve this. Additionally, where there may be difficulty collecting sufficient data to derive predictive validity coefficients, job analysis may be the only means of establishing the validity of psychometric tests. Given the current lack of monitoring and validation of test use in the company, it is particularly important that job analysis is conducted to justify the choice and interpretation of tests.

In two areas, the results of this survey follow those found by Newell and Shackleton (1994). Both surveys indicate that ethical guidelines are generally followed in the use of trained staff for administering and interpreting tests. Both surveys also indicate a less clear-cut picture in the provision of feedback to candidates. The results of this survey indicate that de-briefing and feedback are offered in just over half the occasions when tests are used, and then generally only to internal candidates. Newell and Shackleton's survey indicates that only one-third of companies are routinely feeding back results to candidates who have taken tests as part of a selection process. It seems that the ethical guidelines for training staff in test use have been adopted by companies; one might speculate that the role of the test publishers may have played a part in this respect in the promotion of training courses for particular tests. But the ethical responsibilities of companies in providing feedback have yet to be fully adopted. De-briefing and feedback have particular implications, of course, for staff resources. However, as Herriot (1984) states, it is incumbent on companies to supply information from the selection process; given the relatively powerless position of the candidate in the selection process it is unlikely that many would request it.

There may also be practical reasons why feedback should be given to candidates. Although Herriot (1989) states that little psychological research has been carried out into the impact of selection and assessment procedures on candidates, Smith and Robertson (1993) say that there is evidence to suggest that being assessed (even without feedback) is an important event for an individual and may have some psychological impact. Schmitt, Ford and Stutts (1986) found that when assessment is coupled with feedback in a selection/promotion decision, the possibility of impact on a candidate is enhanced. Iles and Robertson (1989) point out that the treatment of candidates during the selection process may affect their subsequent commitment to, and performance in, the organisation. In this respect, feedback plays an important part in the selection process, both on ethical grounds, and for practical considerations as well.

The company has particular difficulties in test use because it is a multi-site operation. This presents a challenge in terms of ensuring that tests are administered in a standardised form across different district offices. The results indicate that improvements could be made in this area. Although low sample sizes mean that caution should be used in interpretation, the results also indicate that the weighting given to tests varies between districts. However, this may not be a function of the company's multi-site operation; even between individual HRAs and DTMs the weighting given to tests appears to vary considerably. It seems that the likely explanation of this occurrence is that the application of weighting is, at least partly, subjective. This is perhaps not surprising given the lack of guidance in this respect from the various guidelines on test practice. These guidelines in turn presumably reflect the lack of research into combining predictors for selection (Hunter and Hunter, 1984). However, given the company's position on testing (that it should carry the lowest weighting in the assessment process) clearer guidance on which weighting should apply in which circumstances may help to overcome the subjective approach currently taken.

The company's intention in using tests is that the results should form the basis of probes to be used during interview. This intention is reflected to an extent in the results, which indicate that test results are fairly frequently used in this way, particularly by HRAs. It is a little surprising, perhaps, given the clear purpose of the use of tests in the company, that the survey results do not show a greater use of test results in this way. This may be partly explained by the additional information from telephone interviews with DTMs, which indicates that on occasions the test results do not reach them before the interviews begin. Further information supplied on the questionnaire also shows that DTMs may be using the test results to verify what they have found from other parts of the selection process, rather than using them to form probes for testing in interview. Clearer guidelines may be needed on the use of test results to formulate probe questions in interview.

Bethell-Fox (1989) points out that in terms of utility, test use can sometimes bring immediate savings. An example of this is when aptitude and ability tests are used to 'screen out' candidates who do not meet minimum job requirements, thereby saving the expensive time of interviewers. The results of the survey indicate that the company only occasionally uses ability and aptitude tests for screening, which may be a reflection of the low level of recruitment taking place in the company at the moment. There is some evidence that personality tests are used occasionally for screening, which is not part of recommended practice.

The comments made by a number of HRAs and DTMs regarding the use of personality test results to see if a person would 'fit' into the team or company are interesting in the light of Schneider's (1987) view that organisations maintain their identity and culture by attracting and selecting people like themselves; those who do not 'fit' tend not to stay. Herriot's (1989), in commenting about the continuing popularity of the interview, also states that employers want to know 'what the person is really like', and this is normally done by interview. Herriot reflects that the relative popularity of personality

tests, despite their low validity, may be viewed as another means of getting to know what the person is like.

There is little evidence that the company is addressing the problem of possible indirect discrimination. Although tests appear to be related to job requirements, there is no monitoring of test results by ethnic origin or gender, which means there is no evidence as to whether tests are discriminating against those groups. These results follow research by both the EOC (1987) and the CRE (1991) which indicates that monitoring is rarely carried out by companies. Additionally, the company makes no provision when administering tests for samples of test materials to be sent to candidates in advance, or for candidates to be given the opportunity to practise 'mock' tests, which may disadvantage minorities. Strict adherence to time keeping in timed tests, while on the whole laudable, may also have implications for minorities if applied inflexibly.

The results from the survey indicate that in the company the relationship between test results and job performance is not normally validated. There is no doubt that this is an area where organisations have particular difficulties. As Pearn et al (1987) point out, validation studies are very difficult to undertake, being both time-consuming and demanding in resources. Groups of at least 100 are recommended (Toplis et al, 1987) which presents difficulties in companies with only small numbers of people doing the same type of jobs. Similar problems may be experienced in following the CRE's (1992) recommendations that test validity should be established for all the main racial groups expected to take it. However, as the IPM guide (1988) states, if expertise is available and numbers are sufficient, in-house validation should be carried out. If a company does not evaluate or validate the relationship between tests and job performance (or some other appropriate criterion) there is no conclusive evidence that tests are helping the selection process.

Conclusion and recommendations

This study was undertaken to evaluate the use of psychometric testing in the company, and to make recommendations for improvement where necessary.

In certain respects, test use in the company follows the model of good testing practice which was developed for the study. There is evidence that job descriptions are revised; appropriate training of both HR advisers and test administrators appears to take place; importance is placed on proper test administration; great care is taken over the security of test materials and access to test results. In line with the model of good practice, it appears that test results are used alongside other methods of selection and minimum cut-off scores are not frequently used.

However, there are areas where test use in the company could be improved. In this respect the following recommendations are suggested:

General

A policy document is developed stating the aims and procedures of the use of psychometric testing in the company.

Test administration

1. Tape-recorded instructions are used in briefing candidates, which should be supplemented by printed handouts.
2. The purpose of the tests, and the tests to be used, are explained to candidates prior to the test situation.
3. Samples of test materials should be available to candidates, and the opportunity to practise 'mock' tests if necessary.
4. Attempts should be made to ascertain, from external as well as internal candidates, whether they have taken similar tests recently.
5. Steps should be taken to ensure that the physical conditions of testing are always suitable.
6. The effect on minorities of strict time-keeping in timed tests should be monitored.

Test use

1. Personality tests should not be used for screening candidates.
2. Clearer guidelines should be devised on the weighting to be used on tests when combining them with other selection procedures.
3. Clearer guidelines should be devised on using test information to formulate probe questions in interview.
4. Steps should be taken to ensure that test interpretation always reaches the DTMs prior to interviewing.

Feedback

De-briefing (and ideally full feedback) should be offered to all candidates, internal and external, regardless of whether they are successful in their application or not.

Monitoring and validation

1. Validation should be carried out of test use in the company.
2. Test results should be monitored for discrimination against minorities and women.
3. Jobs should be regularly monitored for changes which may affect test suitability.

Further research

Given the growing body of evidence that assessment has a psychological impact on candidates (Schmitt, Ford and Stutts, 1986; Iles and Robertson, 1989; Smith and Robertson, 1993), it is suggested that further research should be conducted to evaluate the effect of psychometric testing on those employees who have been tested.

References

- Anastasi, A. (1988). *Psychological Testing*. New York: Macmillan.
- Barrick, M. R. and Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, **44**, 1-26.
- Bethell-Fox, C. C. (1989). *Psychological Testing*. In P. Herriot (ed), *Assessment and Selection in Organisations*. Chichester: Wiley
- Bevan, S. and Fryatt, J. (1988). *Employee Selection in the UK*. Falmer, Sussex: Institute of Manpower Studies.
- Blinkhorn, S. and Johnson, C. (1990). The insignificance of personality testing. *Nature*, **348**, 671-672.
- Boudreau, J. W. (1983). Economic considerations in estimating the utility of human resource improvement programmes. *Personnel Psychology*, **35**, 551-576.
- British Psychological Society (1992). *Psychological Testing: A Guide*. Leicester: BPS.
- Cascio, W. F. (1982). *Costing human resources: the financial impact of behaviour in organisations*. Boston, MA: Kent.
- Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, **5**, 115-124.
- Commission for Racial Equality (1992). *Psychometric Tests and Racial Equality. A Guide for Employers*. London: CRE
- Commission for Racial Equality (1993). *Towards Fair Selection*. London: CRE.
- Cronbach, L. J. (1990). *Essentials of Psychological Testing*. London: Harper Collins.
- Equal Opportunities Commission (1988). *Avoiding Sex Bias in Selection Testing: Guidance for Employers*. Manchester: EOC.
- Flaugher, R. L. (1978). The many definitions of test bias. *American Psychologist*, **33**, 671-679.
- Fletcher, C. (1994). Validity, test use and professional responsibility. *The Psychologist*, **7** (1), 30-31.
- Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology*, **26**, 461-467.
- Guion, R. M. and Gottier R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, **18**, 135-164.
- Herriot, P. (1984). *Down from the Ivory Tower: Graduates and their Jobs*. Chichester: Wiley.
- Herriot, P. (1989). Interactions with clients in personnel selection. In P. Herriot (ed), *Assessment and Selection in Organisations*. Chichester: Wiley.
- Herriot, P. (1989). The selection interview. In P. Herriot (ed), *Assessment and Selection in Organisations*. Chichester: Wiley.
- Hunter, J. E., Schmidt, F. C. and Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, **86**, 721-735.
- Hunter, J. E. and Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, **96** (1), 72-98.
- Iles, P. A. and Robertson, I. T. (1989). The impact of personnel selection procedures on candidates. In P. Herriot (ed), *Assessment and Selection in Organisations*. Chichester: Wiley.

- Institute of Personnel Management (1988). *IPM Code on Psychological Testing*. London: IPM.
- Johnson, C. and Blinkhorn, S. (1994). Desperate measures: Job performance and personality test validities. *The Psychologist*, 7 (4), 167-170.
- Kline, P. (1993). *The Handbook of Psychological Testing*. London: Routledge.
- Mabey, B. (1989). The majority of large companies use psychological tests. *Guidance and Assessment Review*, 5 (3), 1-4.
- Mabey, B. (1992). The growth of test use. *Selection and Development Review*, 8 (3), 6-8.
- Muchinsky, P. M. (1986). Personnel selection methods. In C. L. Cooper and I. Robertson (eds), *International Review of Industrial and Organizational Psychology*. Chichester: Wiley.
- Newell, S. and Shackleton, V. (1994). The use (and abuse) of psychometric tests in British industry and commerce. *Human Resource Management Journal*, 4 (1), 14-23.
- Pearn, M. A., Kandola, R. S. and Mottram, R. D. (1987). *Selection Tests and Sex Bias: The Impact of Selection Testing on the Employment Opportunities of Women and Men*. London: H.M.S.O.
- Pearn, M. A. (1989). Fairness in employment selection: a comparison of UK and USA Experience. In M. Smith and I. Robertson (eds), *Advances in Selection and Assessment*. Chichester: Wiley.
- Reilly, P. R. and Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, 35, 1-62.
- Robertson, I. T. and Makin, P. J. (1986). Management selection in Britain: A survey and a critique. *Journal of Occupational Psychology*, 59, 45-57.
- Robertson, I. T. and Kinder, A. (1993). Personality and job competences: The criterion-related validity of some personality variables. *Journal of Occupational and Organizational Psychology*, 66 (3), 225-244.
- Robertson, I. T. (1994). Personnel selection research: Where are we now? *The Psychologist*, 7 (1), 17-21.
- Schmitt, N., Ford, J. K. and Stults, D. M. (1986). Changes in self-perceived ability as a function of performance in an assessment centre. *Journal of Occupational Psychology*, 59, 327-335.
- Schmitt, N. E., Gooding, R. Z., Noe, R. A. and Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407-421.
- Schmitt, N. E. (1989). Fairness in employment selection. In M. Smith and I. T. Robertson (eds), *Advances in Selection and Assessment*. Chichester: Wiley.
- Schmitt, N. and Noe, R. E. (1986). Personality selection and equal employment opportunity. In C. L. Cooper and I. T. Robertson (eds), *International Review of Industrial and Occupational Psychology*. Chichester: Wiley.
- Schneider, B. (1987). The people make the place. *Personnel Psychology*, 40, 437-453.
- Shackleton, V. and Newell, S. (1991). Management selection: A comparative survey of the methods used in top British and French companies. *Journal of Occupational Psychology*, 64, 23-26.
- Smith, M. and Robertson, I. T. (1993). *The Theory and Practice of Systematic Personnel Selection*. London: Macmillan.
- Tett, R. P., Jackson, D. N. and Rothstein, M. (1991). Personality measures as

- predictors of job performance: A meta-analytic review. *Personnel Psychology*, **44**, 703-742.
- Toplis, J., Dulewicz, V. and Fletcher, C. (1987). *Psychological Testing: A Practical Guide*. London: IPM.
- Williams, R. S. (1991). *Psychological testing and management selection practices in local government: Results of a survey, June 1991*. Luton, Beds: Local Government Management Board.
- Williams, R. S. (1994). Occupational testing: Contemporary British practice. *The Psychologist*, **7** (1), 11-13.

APPENDIX A

A model of good practice in the use of psychological tests in UK organisations

Introduction

The term 'psychological test' (also known as an 'occupational test' or 'psychometric test') can be defined as 'any procedure on the basis of which inferences are made concerning a person's capacity, propensity or liability to act, react, experience, or to structure or order thought or behaviour in particular ways'. (BPS, 1992)

Broadly speaking, in an organisational context, psychological tests are used to measure individual differences in personality, aptitude or ability in the areas of recruitment and selection, training and development, and counselling. This model of good practice is written within this contextual framework.

Prior to choice of test

Relevance

If the situation is one of recruitment, there is a need to establish if the use of tests is relevant to the prevailing labour market and recruitment conditions (ie there needs to be an excess of applicants over number of job vacancies to make the use of tests in recruitment appropriate).

It also needs to be established if the test will produce any relevant information which could not be obtained by other means.

In situations where tests are used, they should not be the only source of information upon which decision-making is based. They should be used alongside other methods of collecting data about the applicant, and should generally be used to instigate further discussion. There is a need to establish, therefore, the form the other methods of assessment will take, and at what stage tests should be incorporated into the assessment process.

Job analysis

Where testing is to be used for recruitment and selection, a job analysis should be conducted to ascertain the tasks/activities of the job and hence the person specification (ie the qualities required for successful job performance). This will establish what is to be assessed and what is not to be assessed, and will determine the type of test to be used. As test results should not be used as the only method for making selection decisions, the other methods to be employed in the selection process can be identified at this stage.

Choosing the test

The choice of test can be broadly divided into two categories. The principles of good testing practice apply equally to both.

'Customised' tests

These are tests developed within an organisation for a particular job or cluster of jobs. These should only be designed and developed by a chartered occupational psychologist. They should be trialed and validated to the same standards as tests purchased 'off the shelf' from reputable publishers.

'Off the shelf' tests

These should be supplied by a reputable publisher only to those trained in its use. If there is uncertainty over which test to use, a chartered occupational psychologist should be consulted.

A test should have clear instructions for administration, scoring and interpretation (including norms). The manual should contain objective scientific data to demonstrate the internal consistency and test re-test reliability of the test, and its construct and concurrent or predictive validity. There should be evidence of the test performance of subgroups (eg women, ethnic minorities, people with disabilities) to indicate that the test does not have adverse impact. The appropriateness of the test for its intended use should be considered, with particular reference to the norm groups used and whether

the test actually measures attributes directly relevant to the employment situation. Its acceptability (face validity) to applicants should also be considered to avoid alienation.

Local validation studies

Ideally, evidence of validity should not be accepted solely from the test manuals.

Before results from the test are used, validation studies should be carried out in the organisation to establish if the test results are good performance predictors both for the majority group and each subgroup (ethnic minorities, women, those with disabilities).

The validation study should involve at least 100 people from each of the groups, and should establish either predictive validity or, if this is not feasible, concurrent validity.

Where a local validation study is not possible, then the data in the test manual should be cross-checked from other sources.

Test administration

The test should only be administered by those trained in its use, and the test instructions in the test manual should be followed correctly. The test should not be administered by post. Candidates should be told in advance that they will be tested, which tests will be used and their purpose. In a selection process, candidates should be reassured that test results are not used in isolation in reaching a decision, and that other selection methods are referred to.

A check should be made to ascertain if candidates have been through a similar process recently, and if so, which tests were used.

Before testing, candidates should be sent test descriptions and sample question sheets to increase acceptability and reduce discrimination. Access training programmes in testing for multi-ethnic groups should be provided to enable them to practise tests.

The conditions of testing should be standardised in order that they remain the same whenever and wherever the test is administered. In particular, multi-site operations need to monitor the maintenance of good testing practice.

The group to be tested should comprise not more than 20 people without help from extra staff, and care should be taken with the physical conditions of the test (room size, heating, lighting, etc).

To establish consistency, a tape-recording of instructions should be played. The instructions should also be handed out in the form of a briefing sheet. There should be an opportunity to ask questions before the test begins.

The length of time taken to complete the test may be important, but the needs of ethnic minorities should be treated sensitively, as strict adherence to timing can lead to adverse impact.

Care should be taken over the security of the test materials.

Test interpretation

Test results should be scored and interpreted only by qualified staff. The results should be evaluated only in relation to appropriate employment criteria.

Feedback

Ownership of the testing process should rest jointly with the employer and the applicant. At the very least, therefore, candidates should be offered a 'de-briefing' session where they can discuss their feelings about the testing process. Ideally, a confidential session to give feedback and discuss the results of the test should be given by someone trained in both testing and counselling. It should cover both test interpretation and implications for the person's behaviour at work, relationships, development needs, career progress etc. This should take place once the outcome of the total assessment process is known.

Use of results

In selection, test results can be used as a method of screening candidates before short-listing or for selecting candidates at a later stage as part of an assessment process.

Tests can also be used in assessing training needs, or to establish an individual's potential for promotion. They may also be used as an aid to career counselling.

However, they should always be used in conjunction with other methods of assessment.

The use of absolute cut-off scores should be avoided. The minimum level below which a candidate may be considered unsuitable should be appropriate to the job level. A top-down approach can also lead to adverse impact when used across the whole group, and consideration should be given to taking a top-down approach with individual sub-groups. Generally speaking, the test score should be used as a guide rather than as an absolute.

Records

Records should be kept of all test scores (preferably with answer sheets as well) to facilitate follow-up studies. The records are confidential documents and access to them should be tightly controlled. They should be held in the organisation's personnel department, with access restricted to accredited users only.

The 'shelf-life' of the data should be no more than 2 years. After that it should not be used in making assessment decisions. The data should be destroyed at this point unless it is to be kept for validation studies (eg identification of management potential among graduates).

Cognitive ability tests

For less complex jobs, tests of psychomotor ability may yield higher validity than tests of cognitive ability. As job complexity increases, or if receiving training is an important part of a new job, tests of cognitive ability produce higher validity.

Personality tests

In addition to the guidelines for good testing practice outlined above, certain procedures should be followed when personality questionnaires are used.

Generally they should not be used alone to screen candidates. They should be used to provide additional criterion-related information over and above that provided by other psychometric tests and other methods of selection.

The trait measures to be used should be selected on the basis of a job analysis aimed specifically at identifying the personality characteristics necessary for a given job.

The use of cut-off scores is not appropriate in the interpretation of personality tests. If the job analysis has identified certain traits as particularly necessary for a given job, it may be more appropriate to place candidates on a continuum according to how much of a particular trait they display.

Monitoring test use

Candidates' experience of the testing process should be monitored. The test results should also be monitored for unfair discrimination against subgroups. If candidates from particular subgroups appear to be failing validated selection tests disproportionately, the reasons should be investigated and the substitution of other measures of potential considered.

The relationship between the test and the job should also be regularly monitored for relevance, particularly if the job has been the subject of organisational change.

Monitoring should aim to ensure that high standards are maintained in the use of tests in the organisation.

The use of tests in assessment in organisations comes within the scope of three Acts of Parliament, namely:

Sex Discrimination Act 1975

Race Relations Act 1976

Data Protection Act 1984 (where test scores are stored on computers).

Sources:

- British Psychological Society (1992). **Psychological Testing: A Guide.**
- Commission for Racial Equality (1992). **Psychometric Tests and Racial Equality.**
- Cronbach, L.J., (1990). **Essentials of Psychological Testing.** New York: Harper Collins.
- Equal Opportunities Commission (1988). **Avoiding Sex Bias in Selection Testing: Guidance for Employers.**
- Institute of Personnel Management (1989). **The IPM Code on Occupational Testing.**
- Newell, S. & Shackleton, V. (1992). Are psychometric tests being used ethically in British industry and commerce? **Selection and Development Review**, Vol.8, No.6, 5-7.
- Robertson, I.T. & Kinder, A. (1993). Personality and job competences: the criterion-related validity of some personality variables. **Journal of Occupational and Organizational Psychology**, Vol.66, Part 3, 225-244.
- Smith, M. & Robertson, I.T. (1993). **The Theory and Practice of Systematic Personnel Selection.** London: Macmillan.
- Tett, R.P., Jackson, D.N. & Rothstein, M. (1991). Personality measures as predictors of job performance: a meta-analytic review. **Personnel Psychology**, 44, 703-42.
- Toplis, T., Dulewicz, V. & Fletcher, C. (1987). **Psychological Testing. A Practical Guide.** London: IPM.

APPENDIX B

Questionnaire for Test Administrators

PERSONAL DETAILS

1. Name:
2. Position in company:
3. Location:
4. Length of service in company:
5. How long have you been in your current position?
6. How long ago were you trained in test administration?

TEST ADMINISTRATION

7. Are tests administered:-

- | | |
|---|------------------|
| a) by post | YES NO SOMETIMES |
| b) at the candidate's place of work (for internal candidates) | YES NO SOMETIMES |
| c) under supervision at a specified location | YES NO SOMETIMES |
| d) other (please specify) | |
| | |

8. How often do you administer tests?

WEEKLY FORTNIGHTLY MONTHLY LESS OFTEN THAN MONTHLY

9. How confident are you in administering tests?

VERY CONFIDENT FAIRLY CONFIDENT NOT VERY CONFIDENT

10. Are candidates told in advance:-

- | | |
|-------------------------------------|------------------|
| a) the purpose of the tests | YES NO SOMETIMES |
| b) which tests are going to be used | YES NO SOMETIMES |

11. Are candidates asked in advance if they have taken any similar tests recently?

YES NO SOMETIMES

- | | |
|--|------------------|
| 12. Are samples of test questions sent to candidates in advance? | YES NO SOMETIMES |
| 13. Are candidates given the opportunity to practise tests in advance? | YES NO SOMETIMES |
| 14. Is the test material kept in a secure place? | YES NO SOMETIMES |
| 15. Are all test results returned to the interpreter on the same or the following day? | YES NO SOMETIMES |

CONDITIONS OF TESTING

- | | |
|--|------------------|
| 16. Are instructions given by:- | |
| a) reading out a prepared script? | YES NO SOMETIMES |
| b) orally, but without a prepared script? | YES NO SOMETIMES |
| c) by giving out a handout for candidates to read? | YES NO SOMETIMES |
| d) by playing tape recorded instructions? | YES NO SOMETIMES |
| e) other (please specify) | |

.....

.....

- | | |
|---|------------------|
| 17. Are candidates given the opportunity to ask questions prior to completing the test? | YES NO SOMETIMES |
| 18. In timed tests is strict adherence of timekeeping kept? | YES NO SOMETIMES |
| 19. Is the test environment (room etc.) always suitable? | YES NO |

DEBRIEFING AND FEEDBACK

- | | |
|--|------------------|
| 20. Are all candidates offered a debriefing where they can express their feelings about the testing process? | YES NO SOMETIMES |
| 21. Are all candidates offered a confidential session where their results are fed back to them? | YES NO SOMETIMES |
| 22. For personality tests, are all candidates given a written report? | YES NO SOMETIMES |

KEEPING OF RECORDS

23. Are the records of test results treated as confidential documents? YES NO SOMETIMES

24. Is access to the test results restricted to accredited test users? YES NO SOMETIMES

THANK YOU FOR COMPLETING THE QUESTIONNAIRE. IF YOU WISH TO ADD ANY FURTHER INFORMATION, PLEASE DO SO BELOW.

APPENDIX C

Questionnaire for District Team Managers

PERSONAL DETAILS

1. Name:
2. Position in company:
3. Location:
4. Length of service in the company:
5. How long have you been in your current position?

EXPERIENCE OF TEST USE

6. Do you use tests for making staffing decisions?

YES ☐ Please go to question 7.

NO ☐ There is no need for you to continue with the questionnaire.
However, please give your reasons for not using tests:

.....
.....
.....

7. How long have you been using tests for making staffing decisions?
8. How frequently per month, on average, do you use tests for making staffing decisions?
9. Are you confident in your ability to use test results for making staffing decisions? YES NO SOMETIMES
10. Are you confident that test results make a useful contribution to the quality of staffing decisions? YES NO SOMETIMES

USE OF JOB DESCRIPTION

11. Prior to selection for a particular job, is the job description revised? YES NO SOMETIMES

12. Does it identify the main tasks and activities of the job? YES NO N/A

13. Does it identify the personal skills and attributes of the person to fill the job? YES NO N/A

USE OF TESTS IN SELECTION

14. For selection purposes, do you use:-

- | | |
|--|------------------|
| a) ability tests (eg tests of verbal or numerical reasoning) | YES NO SOMETIMES |
| b) aptitude tests (eg clerical checking tests) | YES NO SOMETIMES |
| c) personality tests (eg OPQ, 16PF) | YES NO SOMETIMES |

Ability Tests (eg tests of verbal or numerical reasoning)

15. When you use **ability** tests for selection, do you:-

- | | |
|--|--------------------------|
| a) always use the same tests | YES NO |
| b) vary the tests depending on seniority of job to be filled | YES NO |
| c) vary the tests depending on tasks/activities of job | YES NO |
| d) don't use ability tests for selection | <input type="checkbox"/> |

(If you do not use ability tests for selection, please continue from Question 21)

16. At what stage in the selection process would you use **ability** tests:-

- | | |
|---|------------------|
| a) for initial screening of candidates | YES NO SOMETIMES |
| b) alongside other selection procedures | YES NO SOMETIMES |

17. In **ability** tests, are minimum cut-off scores used? YES NO SOMETIMES

18. Is the **ability** test score used:-

- | | |
|---|------------------|
| a) to rank candidates for suitability for the job | YES NO SOMETIMES |
| b) as a guide for exploration in interview | YES NO SOMETIMES |
| c) other (please specify) | |
| | |

19. When deciding which candidate to appoint from **internal** applicants, are the results from **ability** tests:-

- | | |
|---|------------------|
| a) given zero weighting alongside other methods of selection | YES NO SOMETIMES |
| b) given a 10% weighting alongside other methods of selection | YES NO SOMETIMES |
| c) given a 20% weighting alongside other methods of selection | YES NO SOMETIMES |
| d) given a 30% or more weighting alongside other methods of selection | YES NO SOMETIMES |

20. When deciding which candidate to appoint from **external** applicants, are the results from **ability** tests:-

- | | |
|---|------------------|
| a) given zero weighting alongside other methods of selection | YES NO SOMETIMES |
| b) given a 10% weighting alongside other methods of selection | YES NO SOMETIMES |
| c) given a 20% weighting alongside other methods of selection | YES NO SOMETIMES |
| d) given a 30% or more weighting alongside other methods of selection | YES NO SOMETIMES |

Aptitude Tests (eg clerical checking tests)

21. When you use **aptitude** tests for selection, do you:-

- | | |
|--|--------------------------|
| a) always use the same tests | YES NO |
| b) vary the tests depending on seniority of job to be filled | YES NO |
| c) vary the tests depending on tasks/activities of job | YES NO |
| d) don't use aptitude tests for selection | <input type="checkbox"/> |

(If you do not use aptitude tests for selection, please continue from Question 27)

22. At what stage in the selection process would you use **aptitude** tests:-

- | | |
|---|------------------|
| a) for initial screening of candidates | YES NO SOMETIMES |
| b) alongside other selection procedures | YES NO SOMETIMES |

23. In **aptitude** tests, are minimum cut-off scores used? YES NO SOMETIMES

24. Is the **aptitude** test score used:-

- | | |
|---|------------------|
| a) to rank candidates for suitability for the job | YES NO SOMETIMES |
| b) as a guide for exploration in interview | YES NO SOMETIMES |
| c) other (please specify) | |

25. When deciding which candidate to appoint from **internal** applicants, are the results from **aptitude** tests:-

- | | |
|---|------------------|
| a) given zero weighting alongside other methods of selection | YES NO SOMETIMES |
| b) given a 10% weighting alongside other methods of selection | YES NO SOMETIMES |
| c) given a 20% weighting alongside other methods of selection | YES NO SOMETIMES |
| d) given a 30% or more weighting alongside other methods of selection | YES NO SOMETIMES |

26. When deciding which candidate to appoint from **external** applicants, are the results from **aptitude** tests:-

- | | |
|---|------------------|
| a) given zero weighting alongside other methods of selection | YES NO SOMETIMES |
| b) given a 10% weighting alongside other methods of selection | YES NO SOMETIMES |
| c) given a 20% weighting alongside other methods of selection | YES NO SOMETIMES |
| d) given a 30% or more weighting alongside other methods of selection | YES NO SOMETIMES |

Personality Tests (eg OPQ, 16PF)

27. When you use **personality** tests for selection, do you:-

- | | |
|--|--------------------------|
| a) always use the same tests | YES NO |
| b) vary tests according to nature of job to be filled | YES NO |
| c) vary tests according to seniority of job to be filled | YES NO |
| d) don't use personality tests for selection | <input type="checkbox"/> |

(If you do not use personality tests for selection, please continue from Question 32)

28. At what stage in the selection process would you use **personality** tests:-

- | | |
|---|------------------|
| a) for initial screening of candidates | YES NO SOMETIMES |
| b) alongside other selection procedures | YES NO SOMETIMES |

29. In **personality** tests, are the candidates' profiles used:-

- | | |
|---|------------------|
| a) to compare with the personal skills and attributes identified by the job description | YES NO SOMETIMES |
| b) as a guide for exploration in interview | YES NO SOMETIMES |
| c) other (please specify) | |

30. When deciding which candidate to appoint from **internal** applicants, are the results from **personality** tests:-

- | | |
|---|------------------|
| a) given zero weighting alongside other methods of selection | YES NO SOMETIMES |
| b) given a 10% weighting alongside other methods of selection | YES NO SOMETIMES |
| c) given a 20% weighting alongside other methods of selection | YES NO SOMETIMES |
| d) given a 30% or more weighting alongside other methods of selection | YES NO SOMETIMES |

31. When deciding which candidate to appoint from **external** applicants, are the results from **personality** tests:-

- | | |
|---|------------------|
| a) given zero weighting alongside other methods of selection | YES NO SOMETIMES |
| b) given a 10% weighting alongside other methods of selection | YES NO SOMETIMES |
| c) given a 20% weighting alongside other methods of selection | YES NO SOMETIMES |
| d) given a 30% or more weighting alongside other methods of selection | YES NO SOMETIMES |

USE OF TESTS FOR OTHER PURPOSES

32. In addition to their use in selection, do you use tests for:-

- | | |
|---|------------------|
| a) assessing potential for promotion | YES NO SOMETIMES |
| b) assessing training needs | YES NO SOMETIMES |
| c) discussing future career development | YES NO SOMETIMES |
| d) restructuring/redeployment | YES NO SOMETIMES |
| e) other (please specify) | |

FEEDBACK

33. Are test results fed back to candidates:-

- | | |
|---------------------|------------------|
| a) by you | YES NO SOMETIMES |
| b) by an HR adviser | YES NO SOMETIMES |

THANK YOU FOR COMPLETING THE QUESTIONNAIRE. IF YOU WISH TO ADD ANY FURTHER INFORMATION, PLEASE DO SO ON THE FOLLOWING PAGE.

APPENDIX D

Questionnaire for Human Resources Advisers

PERSONAL DETAILS

1. Name:
2. Position in company:
3. Location:
4. Length of service in the company:
5. How long have you been in your current position?

TRAINING AND EXPERIENCE IN TESTING

6. Have you received training in psychometric testing for:-

- | | |
|--|--------|
| a) test administration and scoring | YES NO |
| b) test interpretation: ability and aptitude | YES NO |
| c) test interpretation: personality | YES NO |

7. When was your most recent training course?

8. How regular is your involvement in psychometric testing?

WEEKLY FORTNIGHTLY MONTHLY LESS OFTEN THAN MONTHLY

9. How confident are you about your knowledge of psychometric testing?

VERY CONFIDENT FAIRLY CONFIDENT NOT VERY CONFIDENT

USE OF JOB DESCRIPTION

- | | |
|--|------------------|
| 10. Prior to selection for a particular job, is the job description revised? | YES NO SOMETIMES |
| 11. Does it identify the main tasks and activities of the job? | YES NO N/A |
| 12. Does it identify the personal skills and attributes of the person to fill the job? | YES NO N/A |

USE OF TESTS IN SELECTION

13. For selection purposes, do you use:-

- | | |
|--|------------------|
| a) ability tests (eg tests of verbal or numerical reasoning) | YES NO SOMETIMES |
| b) aptitude tests (eg clerical checking tests) | YES NO SOMETIMES |
| c) personality tests (eg OPQ, 16PF) | YES NO SOMETIMES |

Ability Tests (eg tests of verbal or numerical reasoning)

14. When you use **ability** tests for selection, do you:-

- | | |
|--|--------------------------|
| a) always use the same tests | YES NO |
| b) vary the tests depending on seniority of job to be filled | YES NO |
| c) vary the tests depending on tasks/activities of job | YES NO |
| d) don't use ability tests for selection | <input type="checkbox"/> |

(If you do not use ability tests for selection, please continue from Question 20)

15. At what stage in the selection process would you use **ability** tests:-

- | | |
|---|------------------|
| a) for initial screening of candidates | YES NO SOMETIMES |
| b) alongside other selection procedures | YES NO SOMETIMES |

16. In **ability** tests, are minimum cut-off scores used? YES NO SOMETIMES

17. Is the **ability** test score used:-

- | | |
|---|------------------|
| a) to rank candidates for suitability for the job | YES NO SOMETIMES |
| b) as a guide for exploration in interview | YES NO SOMETIMES |
| c) other (please specify) | |

.....

18. When deciding which candidate to appoint from **internal** applicants, are the results from **ability** tests:-

- | | |
|---|------------------|
| a) given zero weighting alongside other methods of selection | YES NO SOMETIMES |
| b) given a 10% weighting alongside other methods of selection | YES NO SOMETIMES |
| c) given a 20% weighting alongside other methods of selection | YES NO SOMETIMES |
| d) given a 30% or more weighting alongside other methods of selection | YES NO SOMETIMES |

19. When deciding which candidate to appoint from **external** applicants, are the results from **ability** tests:-

- | | |
|---|------------------|
| a) given zero weighting alongside other methods of selection | YES NO SOMETIMES |
| b) given a 10% weighting alongside other methods of selection | YES NO SOMETIMES |
| c) given a 20% weighting alongside other methods of selection | YES NO SOMETIMES |
| d) given a 30% or more weighting alongside other methods of selection | YES NO SOMETIMES |

Aptitude Tests (eg clerical checking tests)

20. When you use **aptitude** tests for selection, do you:-

- | | |
|--|--------------------------|
| a) always use the same tests | YES NO |
| b) vary the tests depending on seniority of job to be filled | YES NO |
| c) vary the tests depending on tasks/activities of job | YES NO |
| d) don't use aptitude tests for selection | <input type="checkbox"/> |

(If you do not use aptitude tests for selection, please continue from Question 26)

21. At what stage in the selection process would you use **aptitude** tests:-

- | | |
|---|------------------|
| a) for initial screening of candidates | YES NO SOMETIMES |
| b) alongside other selection procedures | YES NO SOMETIMES |

22. In **aptitude** tests, are minimum cut-off scores used? YES NO SOMETIMES

23. Is the **aptitude** test score used:-

- | | |
|---|------------------|
| a) to rank candidates for suitability for the job | YES NO SOMETIMES |
| b) as a guide for exploration in interview | YES NO SOMETIMES |
| c) other (please specify) | |

24. When deciding which candidate to appoint from **internal** applicants, are the results from **aptitude** tests:-

- | | |
|---|------------------|
| a) given zero weighting alongside other methods of selection | YES NO SOMETIMES |
| b) given a 10% weighting alongside other methods of selection | YES NO SOMETIMES |
| c) given a 20% weighting alongside other methods of selection | YES NO SOMETIMES |
| d) given a 30% or more weighting alongside other methods of selection | YES NO SOMETIMES |

25. When deciding which candidate to appoint from **external** applicants, are the results from **aptitude** tests:-

- | | |
|---|------------------|
| a) given zero weighting alongside other methods of selection | YES NO SOMETIMES |
| b) given a 10% weighting alongside other methods of selection | YES NO SOMETIMES |
| c) given a 20% weighting alongside other methods of selection | YES NO SOMETIMES |
| d) given a 30% or more weighting alongside other methods of selection | YES NO SOMETIMES |

Personality Tests (eg OPQ, 16PF)

26. When you use **personality** tests for selection, do you:-

- | | |
|--|--------------------------|
| a) always use the same tests | YES NO |
| b) vary tests according to nature of job to be filled | YES NO |
| c) vary tests according to seniority of job to be filled | YES NO |
| d) don't use personality tests for selection | <input type="checkbox"/> |

(If you do not use **personality** tests for selection, please continue from Question 31)

27. At what stage in the selection process would you use **personality** tests:-

- | | |
|---|------------------|
| a) for initial screening of candidates | YES NO SOMETIMES |
| b) alongside other selection procedures | YES NO SOMETIMES |

28. In **personality** tests, are the candidates' profiles used:-

- | | |
|---|------------------|
| a) to compare with the personal skills and attributes identified by the job description | YES NO SOMETIMES |
| b) as a guide for exploration in interview | YES NO SOMETIMES |
| c) other (please specify) | |

29. When deciding which candidate to appoint from **internal** applicants, are the results from **personality** tests:-

- | | |
|---|------------------|
| a) given zero weighting alongside other methods of selection | YES NO SOMETIMES |
| b) given a 10% weighting alongside other methods of selection | YES NO SOMETIMES |
| c) given a 20% weighting alongside other methods of selection | YES NO SOMETIMES |
| d) given a 30% or more weighting alongside other methods of selection | YES NO SOMETIMES |

30. When deciding which candidate to appoint from **external** applicants, are the results from **personality** tests:-

- | | |
|---|------------------|
| a) given zero weighting alongside other methods of selection | YES NO SOMETIMES |
| b) given a 10% weighting alongside other methods of selection | YES NO SOMETIMES |
| c) given a 20% weighting alongside other methods of selection | YES NO SOMETIMES |
| d) given a 30% or more weighting alongside other methods of selection | YES NO SOMETIMES |

USE OF TESTS FOR OTHER PURPOSES

31. In addition to their use in selection, do you use tests for:-

- | | |
|---|------------------|
| a) assessing potential for promotion | YES NO SOMETIMES |
| b) assessing training needs | YES NO SOMETIMES |
| c) discussing future career development | YES NO SOMETIMES |
| d) restructuring/redeployment | YES NO SOMETIMES |
| e) other (please specify) | |

FEEDBACK

- | | |
|--|------------------|
| 32. Do you feed back test results to candidates? | YES NO SOMETIMES |
|--|------------------|

MONITORING TEST USE

- | | |
|---|------------------|
| 33. Is the relationship between tests and work/job performance monitored? | YES NO SOMETIMES |
| 34. Are jobs monitored for changes in tasks and activities which may affect test suitability? | YES NO SOMETIMES |

35. Are the test results monitored - eg for discrimination? YES NO SOMETIMES

**THANK YOU FOR COMPLETING THE QUESTIONNAIRE. IF YOU WISH
TO ADD ANY FURTHER INFORMATION, PLEASE DO SO BELOW.**