Psychological Assessment in the Workplace
Psychological Assessment in the Workplace

A Manager’s Guide

Mark Cook and Barry Cripps

John Wiley & Sons, Ltd
Contents

About the Authors vii
Preface ix

Chapter 1 Assessment in the Workplace 1
Chapter 2 Using Psychometric Tests 25
Chapter 3 Tests of Mental Ability 47
Chapter 4 Personality Tests 67
Chapter 5 Sifting and Screening 99
Chapter 6 References and Ratings 121
Chapter 7 Competence Analysis 139
Chapter 8 Assessment and Development Centres 159
Chapter 9 The Interview 185
Chapter 10 Structured Interviews 205
Chapter 11 Other Special Assessment Methods 223
Chapter 12 Using Assessment to Arrive at a Decision 245
Chapter 13 Workplace Counselling 271
Chapter 14 Performance Appraisal 281
Chapter 15 Training for Testing and Assessment 311
Chapter 16 Professional and Ethical Issues 319
Chapter 17 The Future of Assessment 329

References 341
Index 349
About the Authors

Mark Cook is a Chartered Occupational Psychologist. He completed his first degree and D.Phil. at Oxford University and has been a practitioner in occupational psychology and psychological assessment in the workplace since 1968. He directed the standardization of three important personality tests—California Psychological Inventory, Myers Briggs Type Indicator and FIRO-B—in the UK, and is a former Director of Oxford Psychologists Press, a leading UK test publisher. He has researched extensively on personnel selection and psychological testing, including psychological profiles of managers in formerly communist countries and health care workers, and is the author of Personnel Selection, 4th edn, published by John Wiley & Sons in 2003. Mark Cook teaches psychology of work at Swansea University.

Barry Cripps is a Chartered Occupational Psychologist consulting independently in industry, business, Higher Education and sport. His interests are in organisational learning and development, assessment and coaching. Having served as a main board Director of Training, External Examiner in HR to top ranking Business Schools, and verifier of test competence in the British Psychological Society, Barry Cripps brings his extensive experience of people assessment in the workplace to this book. Barry Cripps’ practice is in Dartington, Devon.
We have both been engaged in psychological assessment in the workplace for some years, in a variety of sectors, and also have extensive experience of using psychological tests. We perceived a need for a book that would describe the best techniques for psychological assessment, in a form accessible to managers.

Our account is intended to be research based and critical. There are some good techniques for psychological assessment available; unfortunately there is also quite a lot that is not really worth using. We aim to sort the wheat from the chaff, for the guidance of line and HR managers. Our style of writing is intended to move away from academic language to explain concepts in a language comprehensible to managers.

In selecting research to describe, we have focused on ‘real’ workers and ‘real’ managers, and have avoided citing research that uses college students to simulate managers or other workers, or uses laboratory simulations of work.

We would like to thank the many people who have helped us prepare this book. First, we would like to thank the many researchers in the selection area who have generously sent us accounts of research in press or in progress. Second, we would like to thank the managers and students on training courses whose questions and comments have helped us write this book. Third, we would like to thank Karen Howard for her help with the illustrations used. Finally, we would like to thank the Psychology team at John Wiley & Sons for their support and help.

Mark Cook
Barry Cripps
CHAPTER 1

Assessment in the Workplace

OVERVIEW

- Nine general headings in what is assessed.
- A preliminary list of ways of assessing people.
- An overview of what is assessed, and how.
- How assessments should themselves be assessed.
- The concepts of consistency (or reliability) and accuracy (or validity) in assessment.
- The correlation and what it means.
- Outlining some reasons why validity of assessment is sometimes low.
- The concepts of fairness and adverse impact in assessment.
- How to identify good work performance.
- How selection is currently practised in the UK, the USA and Europe.

INTRODUCTION

Consider the ideal employee, from the employer’s point of view. Mr/Ms Ideal is never late, never ill, never argues, never refuses to undertake an assigned task, doesn’t join a union, so doesn’t ‘go slow’ or ‘work to rule’ and certainly never goes on strike. Mr/Ms Ideal works 100% of the time, always does every thing exactly right, but also does everything very quickly. Where will the employer find such a paragon? In the showroom of the nearest industrial robot supplier. But suppose we need a human Mr/Ms Ideal. How do we set about finding him/her?

Think of someone in your organisation whom you consider is a real liability and ask yourself:

- Why is he/she a liability?
- What does he/she do to make you think that?
• How did we come to employ him/her in the first place?
• What selection assessments did we make?

Now let us take a more positive slant. Think of someone who is a real asset, the sort of person you would definitely like more of, and ask yourself:

• Why is he/she an asset?
• What does he/she do to make you think that?
• How did we come to employ him/her in the first place?
• What selection assessments did we make?

Helping you get these decisions right is what the rest of this book is about. There are four main issues:

• Why we assess.
• What to assess.
• How to assess it.
• How to assess the assessment.

WHY WE ASSESS

We can distinguish four main reasons for assessing people in the workplace:

• selection
• promotion
• downsizing
• formal appraisal.

Selection is difficult. We are trying to predict how well someone will work, over a period of perhaps 10 years, on the basis of the information we can collect in a period lasting from 30 minutes to at most 3 days. We are limited by ethics, the law and the natural desire of applicants to present themselves in the best light.

Promotion and downsizing are logically the same as selection: we compare what we need with what people have to offer. The main difference is we have—or should have—much more, and much better, information. If someone has been doing the job for 5 years, we should have a pretty clear idea of what they are like and what they can do. Downsizing can be trickier than promotion for obvious reasons.

These days most organisations in North America and the UK have performance appraisal systems. Employees are formally assessed at regular intervals, and the assessments used either to plan their training and development, or to determine their pay and promotion.
WHAT TO ASSESS

The short answer to this is: ability to do the job. A more detailed answer is provided by a competence analysis, which will list the main competencies that successful employees need (see Chapter 7). We give here a general list of the main headings for assessing staff:

- knowledge
- work skills
- social skills
- organisational fit
- required work behaviours
- physical characteristics
- mental ability
- personality
- interests and values.

Assessing Staff: General Attributes

Knowledge

Every job requires some knowledge. It may be high-level expertise, such as knowledge of current employment law, of human anatomy and physiology, or of aerodynamics and materials science. Or it may be something very limited and basic that can be acquired in a few minutes: where to find the broom and what to do with the rubbish when you have swept it up. Sometimes the knowledge element is implicit: employers tend to assume applicants can read and write, use telephones, give change etc. Knowledge can be acquired by training; so need not necessarily be a selection requirement. Mastery of higher level knowledge may require higher level mental ability.

Work Skills

A skill is the ability to do something quickly and efficiently: bricklaying, heart surgery, driving a vehicle, valuing a property etc. Skills can mostly only be acquired by practice. Reading a book on driving does not enable you to drive, nor does watching someone else drive. Employers sometimes select for skills, and sometimes train for them. Mastery of some skills may require levels of mental or physical ability not everyone has.

Some approaches to selection aim to assess only knowledge and skill(s), because they are very specific, fairly easy to assess and generally uncontroversial. Some psychologists and some managers see limitations in this, because some jobs change so fast that specific knowledge and skills may soon become out of date.
Social Skills

Social skills are important for many jobs, and essential for some. Here is a selection of words and phrases from job adverts: good communicator, gets on well with others, persuasive, able to control or influence people, teamwork, leadership. These skills are also sometimes called social intelligence, or currently emotional intelligence.

Organisational Fit

The term ‘organisational fit’ refers to whether the applicant’s outlook or behaviour matches the organisation’s requirements. These can be explicit: soldiers are expected to obey orders instantly and without question. Fit is often implicit: the applicant does not sound or look ‘right for us’, but there is no written list of requirements, or even a list that selectors can explain to you.

Required Work Behaviours

Examples of such behaviours are punctuality, health (avoiding) absence, theft, disorder, and drink and drug abuse. These are often overlooked by psychologists and competence analysts, but all employers need people who come to work on time, and who do not steal, start fights etc. Most organisations also want people who believe that what the organisation does is useful and important, or who at least behave as if they do.

Physical Characteristics

Some jobs need specific physical abilities: strength, endurance, dexterity. Others have more implicit requirements for height or appearance.

Managers are sometimes suspicious of formal competence or job analyses, because they seem to leave important things out. Formal qualification in HRM, they argue, is not the whole picture when selecting a good HR manager. Psychologists like to think that what is left out is ability, personality, interest and values, which psychological tests can assess. We should always be careful that ‘what’s left out’ does not turn out to be ‘people like me’ or ‘white males like me’. There is always the risk that something you cannot articulate very clearly might be something you are not allowed to select for, for example, gender.

Mental Ability

Mental ability divides into general mental ability (or ‘intelligence’), and more specific applied mental skills, for example problem solving, practical judgement, clerical ability, mechanical comprehension etc.
Personality

‘Personality’ is a word that means different things to different people (discussed in more detail in Chapter 4). Psychologists tend to use it to describe from 5 to 30 underlying dispositions, or personality traits, to think, feel and behave in particular ways. An extravert person likes meeting people, feels at ease meeting strangers etc. An anxious person worries a lot, may cope with stress or pressure less well etc. Personality traits are thought to be fairly deep-seated parts of the person, that underlie many aspects of their behaviour, and which might be difficult to alter. The employer will probably find it easier to select someone who is very outgoing to sell insurance, rather than trying to train someone who is presently rather shy.

Interests, Values and Preferences

Someone who thinks drinking alcohol is wrong will normally not seek out work in a bar; someone who wants to help others may find social work or religious ministry more rewarding than selling potato crisps; someone who believes that people should obey all the rules all the time may enjoy being a traffic warden. People often have to take work that does not match their ideals and values, but work that does may prove more rewarding. Applicants have sometimes already assessed themselves in this area, if they have sought career counselling.

Explicit versus Implicit Assessment

Some attributes are very explicitly required, especially knowledge and skill. There is no point applying for a post as a doctor without a medical degree. Other requirements are more implicit, sometimes barely even articulated: personality, interest and values, organisational fit, sometimes aspects of physical appearance.

Specific versus General

Some requirements are very specific, e.g. skill in comptometer operation. Other are much broader, e.g. adaptable, flexible or outgoing. Very specific skills are easy to assess, or can be trained for if lacking. The assessment is accurate, simple and creates few legal problems. Assessing broader requirements like adaptability or calmness tends to involve psychologists and tests, is often much less accurate, and creates more legal and PR issues. Our example illustrates the problem with very specific requirements: who uses comptometer operators today? It is an obsolete skill. In a time of rapid change we may need to assess broader underlying characteristics.
Underlying Attributes versus Surface Skills

This issue is linked to the previous one. We can assess surface knowledge and skills, such as bricklaying or knowledge of employment law, or we can assess characteristics that underlie the ability to learn to lay bricks, such as strength and manual dexterity, or to learn employment law, such as general and verbal mental ability (Figure 1.1).

Recommendation: Ask yourself what assessments of staff you currently make, under each of the nine headings: knowledge, work skills, social skills, organisational fit, required work behaviour, physical characteristics, mental ability, personality, interests and values. Are these assessments explicit or implicit? Do you assess anything that our list has left out?

HOW TO ASSESS

Traditional Selection Methods

Figure 1.2 summarises the traditional approach to selecting staff in the UK and the USA. The advertisement attracts applicants, who return an application form. Applicants with satisfactory references are shortlisted and invited for interview. The employer seeks to attract a large pool of applicants, then pass them through a series of filters, until the number of surviving candidates equals the number of vacancies. The traditional filters are:

- application form
- letter of reference
- interview.

It became apparent early in the twentieth century that these traditional methods are not always very effective, starting a search for something better. Newer methods include:

- psychological tests, of mental ability, personality, interests and values etc.;
- work sample tests, such as typing tests or bricklaying tests;
biodata, or biographical questionnaires, which look for common themes in the background of, e.g. successful salespersons;
structured interviews;
assessment centres, which include group exercises and discussions, as written assessments.

Most ‘new’ methods have actually been around a long time. Work samples can be traced back to the Boston, Massachusetts streetcar system in 1913; the first personality questionnaire dates from 1917; assessment centres were first used during World War II. The only significant major new development in the past 25 years is the structured interview. Perhaps we have already found every possible way of assessing people that exists, or could exist. We will return to this question in Chapter 17.

**Recommendation:** Ask yourself which methods of assessing staff you currently use. Distinguish if necessary between different types and level of job. Do you use any assessment methods that we have not mentioned?

**What x How**

We have a list of things we want to assess, and another list of ways to assess them. Different methods are used for different aspects of applicants. Table 1.1 tries to indicate which methods are used for assessing what aspects. An X means that assessment is a main way of assessing that aspect: for example, we assess social skills by interview or assessment, and mental ability by test. We cannot assess social skills...
Table 1.1  How to assess main employee characteristics

<table>
<thead>
<tr>
<th>What</th>
<th>Q</th>
<th>AF</th>
<th>R</th>
<th>I</th>
<th>SI</th>
<th>AT</th>
<th>PI</th>
<th>B</th>
<th>AVI</th>
<th>WS</th>
<th>AC</th>
<th>PR</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>U</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>Work skills</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(x)</td>
</tr>
<tr>
<td>Social skills</td>
<td>X</td>
<td>X</td>
<td></td>
<td>(x)</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organisational fit</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Required work behaviours</td>
<td>X</td>
<td>(x)</td>
<td>(x)</td>
<td>U</td>
<td>X</td>
<td>U</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HR records</td>
</tr>
<tr>
<td>Physical ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>(x)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Physical test</td>
</tr>
<tr>
<td>Mental ability</td>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
<td>X</td>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personality</td>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
<td></td>
<td>X</td>
<td>(x)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interests and values</td>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:  Q, Qualifications; AF, Application Form; R, Reference; I, Interview; SI, Structured Interview; AT, Ability Test; PI, Personality Inventory; B, Biodata; AVI, Attitudes and Values Inventory; WS, Work Sample; AC, Assessment Centre; PR, Peer Rating.

from the application form, nor mental ability by worksample test. An (x) in Table 1.1 means the aspect can be assessed by that method, but in more incidental way: personality questionnaires may provide some information about social skill. A U in Table 1.1 indicates that the assessment may assess, or try to assess, the underlying basis of that aspect. Personality tests might pick out the sort of people who will be absent a lot, or drink too much or otherwise be deficient in required work behaviour.

One thing that leaps out from Table 1.1 is the shortage of Xs for some aspects of applicants. There are not very many core ways of assessing social skills, or required work behaviours, or interests and values.

Recommendation: Draw up your own version of Table 1.1, indicating how you assess, and what you assess.

HOW TO ASSESS THE ASSESSMENT

Assessment methods need to be assessed against six main criteria. An assessment should be:

- **reliable**, meaning it gives a consistent account of the person being assessed;
- **valid**, meaning it selects good applicants and rejects bad ones;
- **fair**, meaning it complies with equal opportunities legislation;
- **acceptable**, to applicants as well as to the organisation;
- **cost-effective**, meaning the assessment saves the organisation more than it costs to use;
- **easy to use**, meaning it fits into the selection process, and allows the selectors to appear organised and prepared to the applicants.
Selection methods do not automatically possess all these qualities; we need research to tell us which possess what. Few assessment methods meet all six criteria so our choice of assessment is always a compromise. We will return to this issue in a later chapter, after reviewing assessment methods in turn.

**RELIABILITY: DOES THE ASSESSMENT GIVE CONSISTENT RESULTS?**

Reliability means consistency. Physical measurements, e.g. the dimensions of a piece of furniture, are usually so reliable their consistency is taken for granted. Selection assessments often are not so consistent. At their worst they may be so inconsistent that they convey little or no information. The Rorschach test asks people what they see in inkbloths, and tries to interpret this as evidence of personality, especially conflicts and defences. The Rorschach is not often used in the UK because it is very unreliable, in two ways:

- what a person sees in the blot on Monday may be quite different from what they see on Wednesday;
- what one expert says about what the person’s Rorschach means may be quite different to what a second expert says.

A test that gives inconsistent results from day to day, and that you cannot agree about is not very useful. Unfortunately there are a lot of assessments about that do fail these two simple tests.

There are several types of reliability in general use:

- **Re-test** reliability is calculated by comparing two sets of information—Rorschach interpretations, or interview ratings, or IQs—obtained from the same set of people, on two separate occasions, typically a month or so apart. If the test is assessing some enduring aspect of the person, as selection assessments are intended to, the two sets of information should be fairly similar.
- **Inter-rater** reliability is calculated by comparing two sets of ratings given by two assessors to a set of applicants they have both interviewed, or observed in a group discussion, or written a reference for. If the assessors do not agree, one at least of them must be wrong, but which?

**Recommendation:** Ask yourself what information you have about the reliability of the assessments you presently make of staff.

**CORRELATION**

Selection research uses correlation very extensively. If you are familiar with correlation, you can skip this section. Correlation is the extent to which two characteristics
go together. In selection, we are usually looking at the correlation between the assessment and work performance, but let’s first look at a more concrete example.

Height and weight are correlated: tall people usually weigh more than short people, and heavy people are usually taller than light people. However, height and weight are not perfectly correlated—there are plenty of short fat and tall thin exceptions to the rule. Figure 1.3 shows height plotted against weight, for 100 people in their twenties.

The correlation summarises how closely two measures like height and weight go together. A perfect one-to-one correlation gives a value of +1.00. Height and weight correlate about 0.75, meaning a strong relationship, but not perfect.

Zero Correlation

If two measures are completely unrelated, the correlation is zero—0.00. For example, we know that extraversion and conscientiousness are not correlated. Knowing how extravert someone is will tell us nothing about how conscientious they are, whereas knowing someone’s weight gives us a generally fairly good idea of their likely height.

Weak Correlations

In selection, we are looking at the correlation between selection assessment and work performance. It would be nice if we could find selection methods that correlated 0.75 with work performance, like the results in Figure 1.3. Unfortunately, most selection
methods do not work that well, and we very rarely find a link as close as that between height and weight. Figure 1.4 shows the sort of results that are more typical. Overall there is some relationship between, e.g. interview rating and work performance, but there are a lot of cases of good performers being rejected and poor performers being accepted. Figure 1.4 shows a correlation of 0.30, which is fairly typical of what you find when comparing selection rating with work performance ratings. Note how many good applicants are being rejected, and how many poor applicants are being accepted.

**VALIDITY: IS THE ASSESSMENT ACCURATE?**

Validity means accuracy. A valid selection method accepts good applicants and rejects poor ones. This is clearly the most important thing to check for in any assessment. There is little point using an assessment method that is completely inaccurate. Unfortunately most ways of assessing staff are less than perfectly accurate, and many are so inaccurate they convey little or no useful information.

**Validation Research**

The basic building block of selection research is the *validation study*. A validation study collects two sets of information, *predictor* and *criterion*.
• the predictor is the assessment: interview, test, reference etc.;
• the criterion is some index of work performance.

Let us take the example of the interview. Interviewer ratings form the predictor, while the criterion is a performance appraisal rating, after one year on the job. Table 1.2 shows the information we need to collect. We have 100 people to study in this example.

We then calculate a correlation between predictor (interview rating) and criterion (appraisal rating), which is the validity coefficient. In our example, this gives 0.32, meaning there is some link between interview rating and performance appraisal—in other words, that the interview is working, and is a valid selection method.

Note that the interviewer’s opinion of the candidates has to be quantified, typically as a rating. You can correlate interview ratings, but not interview impressions, such as ‘good chap’ or ‘I didn’t think much of him’.

For some types of assessment, hundreds, even thousands, of validation studies have been carried out. Work psychologists summarise these to find how well each assessment does on average, and whether it does better or worse in particular circumstances (for example, with particular sorts of work or particular types of people).

Basic validation research is fairly easy to follow: the bigger the correlation, the better. Two linked aspects of validation—construct validity and incremental validity—are a little more complicated, but very important for any one planning a selection programme. Both concern what an assessment is actually assessing, and how it relates to other assessments.

**Recommendation:** Ask yourself what information you have about the validity of the assessments you presently make of staff.

### Table 1.2

The information we need to calculate the validity or accuracy of a selection assessment

<table>
<thead>
<tr>
<th>Employee number</th>
<th>Name</th>
<th>Interview rating</th>
<th>Performance appraisal rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>JS</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>MM</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>BK</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>SM</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>AB</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

**Construct Validity**

When a new selection system is devised, people sometimes ask themselves: ‘What is this assessing?’ ‘What sort of person will get a good mark from it?’ One answer to
this question should always be ‘People who will do the job well’. But it is worth going a bit deeper and trying to get some picture of what particular aspects of the applicant the test is assessing: abilities, personality, social background, specific skills etc. The technical name for this is construct validity. There are several reasons why it is important to explore construct validity:

- If your new test is mostly assessing personality, and you already use a personality test, you may well find that your new test is not adding much to the personality test. Your new test may not be called a ‘personality test’; it may be labelled ‘emotional intelligence’ or ‘sales aptitude’.
- If your 2-day assessment centre measures the same as a 30-minute ability test, it would be much cheaper to use the 30-minute test.
- If applicants complain about your selection methods, and you have to defend them in court or tribunal, you may want to be able to say exactly what your test assesses, and what it does not. You may be made to look very silly if you cannot!
- If you understand what you are doing, you can often devise ways of doing it better.
- If your new test turns out to be mostly assessing intellectual ability, you will be alerted to the possibility of adverse impact on certain groups.

Construct validity is usually assessed by comparing one selection method, e.g. interview ratings, with other methods, e.g. psychological tests, by calculating correlations. This is fairly easy to do if you have an ongoing assessment programme, and you keep good records. Construct validity tells you what a particular method is actually assessing (which is not necessarily that same as what it is intended to assess). For example, the traditional interview turns out to be assessing mental ability to a surprising extent.

**Recommendation:** Ask yourself what information you have about what the assessments you presently make of staff are actually measuring.

### Incremental Validity

Suppose you are using an interview and a mental ability test to select office workers. Someone sells you an expensive computer simulation of key office skills. The people who sell the simulation produce lots of evidence that it predicts good office work very well. After a year or so, management start to complain that the expected improvement in the quality of office workers has not materialised. Did the people who sell you the test mislead you about its validity? No: the simulation does predict good work, but so do the tests you were already using. The new simulation covers exactly the same ground as your existing interview and mental ability test, so it does not improve your overall selection procedure at all. The technical term for this is incremental validity.

This problem always arises when sets of selection tests are used. You can never assume that the validities of all the tests in the set can be simply added, to give overall
validity. Many selection tests cover the same ground, so adding new tests often does not improve overall validity. If you use two equally valid tests, there are three possible outcomes:

- both tests together do twice as well as either alone;
- both tests together do a little better than either alone;
- the second test adds nothing at all.

All three outcomes are possible, but a little better is probably the most likely. The fact that tests have different names or claim to be measuring differing aspects of employee effectiveness does not mean they will not turn out to be duplicating each other’s efforts. This problem is now being analysed and researched more and more, so advice can be given on combinations of tests that are likely to improve accuracy, and combinations of tests that, while effective in their own right, will not build on each at all.

**REASONS FOR POOR VALIDITY**

Why do selection assessments sometimes fail to give good results?

- *Low reliability*. An assessment that does not succeed in measuring much cannot predict much either. Low reliability can sometimes be improved by training, in e.g. interviewing, or by being careful to use the assessment correctly.

- *Imperfect measures of work performance*. We calculate a correlation between the selection assessment and some index of work performance. The most frequent index is performance appraisal rating. Managers who do these will know why they can sometimes be poor reflections of true performance. Some organisations even today lack clear ideas about good and poor work performance, so cannot define it accurately. If we have defective work performance measures, we cannot achieve valid selection. If we do not know what we are looking for, we cannot know if we have found it. For this reason, the organisation should make sure it has an effective performance appraisal system, or some other way of assessing people’s work. If necessary, the organisation should devote some time to clarifying what it exists to achieve, and how it could be certain it has been achieved, and who by.

- *Selection is not the whole story*. Employees’ work performance is influenced by many other factors, besides those characteristics we can assess during selection. Employees’ performance is affected by how they are managed, by the organisation’s culture, by the employment market and by the economy as whole.

- *Impression management or faking good*. Selectors expect applicants to present themselves well. One sometimes hears interviewers complain that an applicant ‘didn’t make much of an effort’ (to describe his/her achievement and potential at sufficient length and with sufficient enthusiasm). Presenting oneself well unfortunately often shades into exaggeration, and from there into lying. Many selection methods rely on information the applicant provides, and often have no way of
verifying it. We will come across this limitation quite frequently in our more
detailed review of selection assessments.

- **Quality of information.** Table 1.3 is Table 1.1 again, but with some indication of
  the nature of the information each assessment provides. This divides into:

  - **self:** information provided by the applicant, on the application form, in the
    interview, and when answering questionnaires.
  - **reported:** information provided by other people about the applicant, principally
    through the reference, but also in ratings of him/her.
  - **recorded:** the applicant has obtained a qualification, which indicates he/she has
    knowledge or a skill.
  - **demonstrated:** the applicant performs a task or demonstrates a skill, in a work
    sample, or on a test with right and wrong answers.

Assessments that demonstrate a skill are clearly better evidence than ones where
the applicant claims to possess it. If someone says he/she can use a computer, we
may doubt him/her, but if they succeed in doing specified tasks with it, there is
no room for doubt.

We noted previously that some of our nine main headings of things to assess
were short of good means of assessment. Table 1.3 adds a further problem.
There is a far greater shortage of assessments that clearly demonstrate that
quality, as opposed to relying on someone’s word for it—usually the applicant’s.
Only knowledge, skill and mental and physical ability can be assessed by
demonstration.

### Table 1.3  How to assess main employee characteristics, indicating the nature of the
information obtained

<table>
<thead>
<tr>
<th>What</th>
<th>How</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>Q, AF, R, I, SI, AT, B, AVI, WS, AC, PR, Other</td>
</tr>
<tr>
<td>Organisational fit</td>
<td>rep, (self), (self), (self), self, self, dem, (rep)</td>
</tr>
<tr>
<td>Required work behaviours</td>
<td>rep, self, self, self, self, self, rep, HR records: rec</td>
</tr>
<tr>
<td>Physical ability</td>
<td>(rep), (self), (self), dem, (rep) Physical test: dem</td>
</tr>
<tr>
<td>Mental ability</td>
<td>(rep), (self), (self), dem, (dem), (rep)</td>
</tr>
<tr>
<td>Personality</td>
<td>(rep), (self), (self), self, (self), (rep)</td>
</tr>
<tr>
<td>Interests and values</td>
<td>(rep), (self), (self), (self), self, (rep)</td>
</tr>
</tbody>
</table>

Notes:  Q, Qualifications; AF, Application Form; R, Reference; I, Interview; SI, Structured Interview; AT, Ability Test; PI, Personality Inventory; B, Biodata; AVI, Attitudes and Values Inventory; WS, Work Sample; AC, Assessment Centre; PR, Peer Rating; rec, recorded; rep, reported by others; self, reported by self; dem, demonstrated; S/D, reported by self and/or demonstrated.
FAIRNESS: WILL THE ASSESSMENT CREATE ADVERSE IMPACT?

Fairness in selection means conforming with equal opportunities laws and codes of conduct. All developed countries have these. The USA led the way with the Civil Rights Act (CRA) of 1964, which prohibited discrimination in employment on grounds of race, colour, religion, national origin or gender. Later laws in the USA brought in age and disability. Similar laws followed in the UK: the Race Relations Act (1976), the Sex Discrimination Act (1975) and the Disability Discrimination Act (1995).

Figure 1.5 shows how fair employment laws work in the USA; other developed countries, including the UK, have followed the same general model and adopted many of the key concepts.

- An assessment that excludes too many of a protected minority, such as non-white persons, is said to create adverse impact.
- The employer can remove the adverse impact by quota hiring to ‘get the numbers right’, i.e. equal number of men and women, correct proportion of ethnic minorities etc.
- Or else the employer can argue that the adverse impact is justified because the selection test is job related, meaning accurate or valid.

![Figure 1.5 Stages in deciding whether a test is legally ‘fair’](image-url)
The employer who succeeds in proving the test job related faces one last hurdle—proving that there is no alternative test that is equally valid but does not create adverse impact.

Note that adverse impact is not what the layperson thinks of as ‘discrimination’. Adverse impact does not mean turning away minorities or women, to keep the job open for white males, or otherwise deliberately treating women or minorities differently. This is clearly unacceptable, and not likely to be done by any major employer these days. Adverse impact means that the selection method results in more men or majority persons getting through than women or minority persons. Adverse impact means an employer can be proved guilty of discrimination, by setting standards that make no reference to race or gender, and that may seem well-established, ‘commonsense’ practice.

- Height and strength tests for police, fire brigade and armed forces create adverse impact, because they exclude more women.
- In Britain some employers sift out any applicants who have been unemployed for more than six months, in the belief they will have lost the habit of working. The (UK) Commission for Racial Equality (CRE) argues this creates adverse impact on ethnic minorities, who are more likely to be unemployed.
- The important Griggs case in the USA ruled that high school diplomas and ability tests created adverse impact, because fewer African American applicants had diplomas or reached the pass mark set on the ability test.
- The Green v. Missouri Pacific Railroad case in the USA showed that excluding applicants with criminal records created adverse impact because more non-whites than whites had criminal records.

Adverse impact is very easy to prove; all you need is a breakdown of the workforce by gender and ethnicity. If there are too few women or minorities, there has been adverse impact. Table 1.4 shows the composition of the UK House of Commons after the 2002 election. There are clearly far ‘too few’ women and minority MPs. In the USA, the Uniform Guidelines introduced the four-fifths rule. If the number selected divided by the number of applicants for a protected minority is less than four-fifths of the highest ratio for any group, adverse impact applies. Table 1.4 shows how this would apply to the House of Commons: there are still ‘too few’ women and

<table>
<thead>
<tr>
<th>Table 1.4 Composition of the British House of Commons in 2002, actual and expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Minority</td>
</tr>
</tbody>
</table>

Note: ‘Expected’ composition is based on the assumption that MPs are selected regardless of race and gender, and that 6% of the population belong to an ethnic minority.
minority MPs. (In fact the law does not apply to the House of Commons, because MPs are not classed as employees.)

Adverse impact assesses the effect of selection, not the intentions of the people who use it. Adverse impact is a very serious matter for employers. It creates a presumption of discrimination, which the employer must disprove, possibly in court. This will cost a lot of time and money and may create damaging publicity. Selection methods that do not create adverse impact are therefore highly desirable, but unfortunately not always easy to find.

If the employer is convinced that the assessment is necessary, even though it creates adverse impact, they must prove it is job related or valid. For example, a police force that thought all officers needed to be at least six feet tall would have to prove that shorter men and women were unable to do the job as well. Some American police forces tried to do this, without success. They were unable to demonstrate any relationship between height and any indicator of effective performance in police work.

The risk of legal challenge, and possible poor publicity, means that the HR department often has to act defensively and may make our account of assessment sound a little negative or pessimistic in places.

US law covers age discrimination, but British law does not—yet. The British government plans to comply with European Directives about age discrimination and introduce legislation by 2007. Britain has recently—December 2003—added religious belief and sexual orientation to the list of legally protected minorities.

Acceptability: What Will Applicants Think of the Assessment?

Selection methods should be acceptable to applicants. In times of high unemployment, employers may feel they can ignore applicants’ reactions. In times of labour shortage, unpopular methods could drive applicants away. Applicants’ views of selection methods may influence their decision on whether to accept a job offer. A recent Belgian survey even reports that applicants who do not like an organisation’s assessment methods may stop buying its products (Stinglhamber et al., 1999).

Surveys indicate that some assessment methods are more popular with applicants than others: they like interviews, work samples and assessment centres, but do not like biodata, peer assessment or personality tests. Applicants see more job-related approaches as fairer: simulations, interviews and more concrete ability tests (vocabulary, maths). They see personality tests, biodata and abstract ability tests (letter sets etc.) as less fair because they seem less job related. Preferences in the USA and France are generally similar, but personality tests and graphology are more

Recommendation: Keep careful records of the gender and ethnicity of all the people you assess at work, and of everyone who applies for employment with you. Check these periodically to ensure that you are not selecting out too many of any group at any stage of the assessment process.

Acceptability: What Will Applicants Think of the Assessment?

Selection methods should be acceptable to applicants. In times of high unemployment, employers may feel they can ignore applicants’ reactions. In times of labour shortage, unpopular methods could drive applicants away. Applicants’ views of selection methods may influence their decision on whether to accept a job offer. A recent Belgian survey even reports that applicants who do not like an organisation’s assessment methods may stop buying its products (Stinglhamber et al., 1999).

Surveys indicate that some assessment methods are more popular with applicants than others: they like interviews, work samples and assessment centres, but do not like biodata, peer assessment or personality tests. Applicants see more job-related approaches as fairer: simulations, interviews and more concrete ability tests (vocabulary, maths). They see personality tests, biodata and abstract ability tests (letter sets etc.) as less fair because they seem less job related. Preferences in the USA and France are generally similar, but personality tests and graphology are more