33
Selection Tests

Key concepts and terms

- Ability test
- Criterion score
- Intelligence test
- Normal curve
- Personality test
- Psychometric test
- Trait

- Aptitude test
- Intelligence quotient (IQ)
- Norm
- Personality
- Psychological test
- Selection test

Learning outcomes

On completing this chapter you should be able to define these key concepts. You should also know about:

- The types of tests
- Test validity
- Choosing tests

- The characteristics of a good test
- Interpreting test results
- Using tests in a selection procedure

Introduction

Selection tests are used to provide valid and reliable evidence of levels of abilities, intelligence, personality characteristics, aptitudes and attainments. They typically supplement the information obtained from an interview.

Selection tests can be divided into two broad categories: measures of typical performance such as personality inventories that do not have right or wrong answers, and measures of maximum performance that measure how well people can do things, how much they know and the level of their ability, and ask questions for which there are right or wrong or good or bad answers. The latter category can focus on what people are capable of knowing or doing (ability tests) or what they actually know or can do (aptitude or attainment tests).

In this chapter, a distinction is made between psychological or psychometric tests, which measure or assess intelligence or personality, and aptitude tests, which are occupational or job-related tests that assess the extent to which people can do the work. These are dealt with in the first two sections of this chapter. Before using any type of test it is necessary to be aware of the characteristics of a good test and methods of interpreting test results, and these considerations are examined in the following two sections. The chapter concludes with sections dealing with choosing tests, using them in a selection procedure and guidelines on their use.

Psychological tests

Psychological tests use systematic and standardized procedures to measure differences in individual characteristics such as intelligence and personality. They enable selectors to gain a greater understanding of candidates to help in predicting the extent to which they will be successful in a job. Psychological tests are measuring instruments, which is why they are often referred to as psychometric tests. 'Psychometric' literally means mental measurement. For selection purposes, the main types of tests are those used for measuring intelligence and ability and those concerned with assessing personality characteristics.

Intelligence tests

Intelligence tests measure a range of mental abilities which enable a person to succeed at a variety of intellectual tasks using the faculties of abstract thinking and reasoning. They are concerned with general intelligence (termed 'g' by Spearman, 1927, one of the pioneers of intelligence testing) and are sometimes called 'general mental ability' (GMA) tests. Intelligence tests measure abilities while cognitive tests measure an individual's learning in a specific subject area. The meta-analysis conducted by Schmidt and Hunter (1998) showed that intelligence tests had high predictive validity. In fact, when combined with a structured interview, they had the highest predictive value of all the methods of selection they studied.

Intelligence tests contain questions, problems and tasks. The outcome of a test can be expressed as a score that can be compared with the scores of members of the population as a whole or the population of the whole or part of the organization using the test (norms).

The outcome of an intelligence test may sometimes be recorded as an intelligence quotient (IQ), which is the ratio of an individual's mental age to the individual's actual age as measured by an intelligence test. When the mental and actual age correspond, the IQ is100. Scores above 100 indicate that the individual's level of average is above the norm for his or her age, and vice versa. It is usual now for IQs to be directly computed as an IQ test score. It is assumed that intelligence is distributed normally throughout the population; that is, the frequency distribution of intelligence corresponds with the normal curve shown in Figure 33.1.
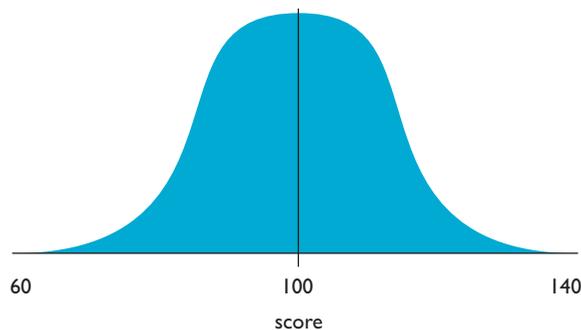

Figure 33.1   A normal curve

The normal curve describes the relationship between a set of observations and measures and the frequency of their occurrence. It indicates that for characteristics such as intelligence that can be measured on a scale, a few people will produce extremely high or low scores and there will be a large proportion of people in the middle. Its most important characteristic is that it is symmetrical – there are an equal number of cases on either side of the mean, the central axis. The normal curve is a way of expressing how scores will typically be distributed; for example, that 60 per cent of the population are likely to get scores between x and y, 20 per cent are likely to get scores below x and 20 per cent are likely to get more than y.

Intelligence tests can be administered to a single individual or to a group. They can also be completed online.

Ability tests

Ability tests establish what people are capable of knowing or doing. Although the term can refer primarily to reasoning ability, the British Psychological Society (2007) refers to ability tests as measuring the capacity for:

- verbal reasoning – the ability to comprehend, interpret and draw conclusions from oral or written language;

- numerical reasoning – the ability to comprehend, interpret and draw conclusions from numerical information;

- spatial reasoning – the ability to understand and interpret spatial relations between objects;

- mechanical reasoning – understanding of everyday physical laws such as force and leverage.

Personality tests

Personality tests attempt to assess the personality of candidates in order to make predictions about their likely behaviour in a role. Personality is an all-embracing and imprecise term that refers to the behaviour of individuals and the way it is organized and coordinated when they interact with the environment. There are many different theories of personality and, consequently, many different types of personality tests. These include self-report personality questionnaires and other questionnaires that measure interests, values or work behaviour.

One of the most generally accepted ways of classifying personality is the five-factor model, which defines the following 'big five' key personality characteristics.

The 'big five' personality characteristics

1. Extraversion/introversion – gregarious, outgoing, assertive, talkative and active (extraversion); or reserved, inward-looking, diffident, quiet, restrained (introversion).

2. Emotional stability – resilient, independent, confident, relaxed; or apprehensive, dependent, under-confident, tense.

3. Agreeableness – courteous, cooperative, likeable, tolerant; or rude, uncooperative, hostile, intolerant.

4. Conscientiousness – hard-working, persevering, careful, reliable; or lazy, dilettante, careless, expedient.

5. Openness to experience – curious, imaginative, willingness to learn, broad-minded; or blinkered, unimaginative, complacent, narrow-minded.

As noted by Schmidt and Hunter (1998), integrity and conscientiousness tests have fairly high predictive validity (0.41 and 0.31 respectively).

Self-report personality questionnaires are commonly used. They usually adopt a 'trait' approach, defining a trait as a fairly independent but enduring characteristic of behaviour that all people display but to differing degrees. Trait theorists identify examples of common behaviour, devise scales to measure these, and then obtain ratings on these behaviours by people who know each other well. These observations are analysed statistically, using the factor analysis technique to identify distinct traits and to indicate how associated groups of traits might be grouped loosely into 'personality types'.

'Interest' questionnaires are sometimes used to supplement personality tests. They assess the preferences of respondents for particular types of occupation and are therefore most applicable to vocational guidance but can be helpful when selecting apprentices and trainees.

'Value' questionnaires attempt to assess beliefs about what is 'desirable or good' or what is 'undesirable or bad'. The questionnaires measure the relative prominence of such values as conformity, independence, achievement, decisiveness, orderliness and goal-orientation.

Personality tests can provide interesting supplementary information about candidates that is free from the biased reactions that frequently occur in face-to-face interviews, but they have to be used with great care. The tests should have been developed by a reputable psychologist or test agency on the basis of extensive research and field testing and they must meet the specific needs of the user. Advice should be sought from a member of the British Psychological Society on what tests are likely to be appropriate.

Aptitude tests

Aptitude tests are job-specific tests designed to predict the potential an individual has to perform tasks within a job. They typically take the form of work sample tests, which replicate an important aspect of the actual work the candidate will have to do, such as using a keyboard or carrying out a skilled task such as repair work. Work sample tests can be used only with applicants who are already familiar with the task through experience or training.

Aptitude tests should be properly validated. This will be the case if a test or a 'test battery' (an associated group of tests) has been obtained from a reputable test agency. Alternatively, a special test can be devised by or for the organization to determine the aptitudes required by means of job and skills analysis. The test is then given to employees already working on the job and the results compared with a criterion, usually managers' or team leaders' ratings. If the correlation between test and criterion is sufficiently high, the test is then given to applicants. To validate the test further, a follow-up study of the job performance of the applicants selected by the test is usually carried out. This is a lengthy procedure, but without it no real confidence can be attached to the results of any aptitude test. Properly validated work sample tests have a high level of predictive validity (0.54 according to Schmidt and Hunter, 1998). The operative words are 'properly validated' – many do-it-yourself tests are worse than useless because this has not happened.

Characteristics of a good test

A good test is one that provides valid data which enable reliable predictions of behaviour or performance to be made and therefore assists in the process of making objective and reasoned decisions when selecting people for jobs. It will be based on research that has produced standardized criteria derived by using the same measure to test a number of representative people to produce a set of 'norms' for comparison purposes. The test should be capable of being objectively scored by reference to the normal or average performance of the group.

The characteristics of a good test

- It is a sensitive measuring instrument that discriminates well between subjects.

- It has been standardized on a representative and sizeable sample of the population for which it is intended so that any individual's score can be interpreted in relation to that of others.

- It is reliable in the sense that it always measures the same thing. A test aimed at measuring a particular characteristic, such as intelligence, should measure the same characteristic when applied to different people at the same or a different time or to the same person at different times.

- It is valid in the sense that it measures the characteristic the test is intended to measure. Thus, an intelligence test should measure intelligence (however defined) and not simply verbal facility. A test meant to predict success in a job or in passing examinations should produce reasonably convincing (statistically significant) predictions.

Types of validity

There are five types of validity:

1. Predictive validity – the extent to which the test correctly predicts future behaviour. To establish predictive validity it is necessary to conduct extensive research over a period of time. It is also necessary to have accurate measures of performance so that the prediction can be compared with actual behaviour.

2. Concurrent validity – the extent to which a test score differentiates individuals in relation to a criterion or standard of performance external to the test. This means comparing the test scores of high and low performances as indicated by the criteria and establishing the degree to which the test indicates who should fit into the high or low performance groups.

3. Content validity – the extent to which the test is clearly related to the characteristics of the job or role for which it is being used as a measuring instrument.

4. Face validity – the extent to which the test 'looks' or 'feels' right in the sense that it is measuring what it is supposed to measure.

5. Construct validity – the extent to which the test measures a particular construct or characteristic. Construct validity is in effect concerned with looking at the test itself. If it is meant to measure numerical reasoning, is that what it measures?

Measuring validity

A criterion-related approach is used to assess validity. This means selecting criteria against which the validity of the test can be measured. These criteria must reflect 'true' performance at work as accurately as possible. A single criterion is inadequate. Multiple criteria should be used. The extent to which criteria can be contaminated by other factors should also be considered and it should be remembered that criteria are dynamic – they will change over time.

Test validity can be expressed as a predictive validity co-efficient in which 1.0 would equal perfect correlation between test results and subsequent behaviour, while 0.0 would equal no relationship between the test and performance. The following rule of thumb guide was produced by Smith (1984) on whether a validity coefficient is big enough:

over 0.50, excellent
0.40–0.49, good
0.30–0.39, acceptable
less than 0.30, poor.

On the basis of the research conducted by Schmidt and Hunter (1998), only work sample tests and intelligence tests with coefficients of 0.54 and 0.51 respectively are excellent.

Interpreting test results

Test results can be interpreted by the use of norms or through criterion scores.

Norms

An individual's score in a test is not meaningful on its own. It needs to be compared with the scores achieved by the population on whom the test was standardized – the norm or reference group. A normative score is read from a norms table and might, for example, indicate that someone has performed the test at a level equivalent to the top 30 per cent of the relevant population.

Criterion scores

Norms simply tell us how someone has performed a test relative to other people. A more powerful approach is to use the relationship between test scores and an indication of what the test is designed to measure, such as job success. This is described as a criterion measure. For example, when the test is validated it might be established that for scores of less than 10 on a test, 50 per cent of people would fail in the job, while the failure rate may be 35 per cent for those who score between 10 and 15 and 20 per cent for those scoring more than 15. The score achieved by the individual would therefore enable a prediction to be made of the likelihood of success.

Choosing tests

It is essential to choose tests that meet the four criteria of sensitivity, standardization, reliability and validity. It is very difficult to achieve the standards required if an organization tries to develop its own test batteries, unless it employs a qualified psychologist or obtains professional advice from a member of the British Psychological Society. This organization, with the support of the reputable test suppliers, exercises rigorous control over who can use what tests and the standard of training required and given. Particular care should be taken when selecting personality tests – there are a lot of charlatans about.

Do-it-yourself tests are always suspect unless they have been properly validated and realistic norms have been established. They should not be used.

The use of tests in a selection procedure

While validated intelligence and personality tests can produce useful data, they should not be relied on entirely. It is best to combine them in a selection procedure with structured interviews.

Tests are often used as part of a selection procedure for occupations where a large number of recruits are required, and where it is not possible to rely entirely on examination results or information about previous experience as the basis for predicting future performance. In these circumstances it is economical to develop and administer the tests, and a sufficient number of cases can be built up for the essential validation exercise. Tests usually form part of an assessment centre procedure.

Intelligence tests are particularly helpful in situations where intelligence is a key factor, but there is no other reliable method of measuring it. It may, incidentally, be as important to use an intelligence test to keep out applicants who are too intelligent for the job as to use one to guarantee a minimal level of intelligence.

Aptitude tests are most useful for jobs where specific and measurable skills are required, such as word-processing and skilled repair work. Personality tests can complement structured interviews and intelligence and aptitude tests. Some organizations use them for jobs such as selling where they believe that 'personality' is important, and where it is not too difficult to obtain quantifiable criteria for validation purposes. They may be used to assess integrity and conscientiousness where these characteristics are deemed to be important.

Tests should be administered only by people who have been trained in what the tests are measuring, how they should be used, and how they should be interpreted.

It is essential to evaluate all tests by comparing the results at the interview stage with later achievements. To be statistically significant, these evaluations should be carried out over a reasonable period of time and cover as large a number of candidates as possible.

In some situations a battery of tests may be used, including various types of intelligence, personality and aptitude tests. These may be a standard battery supplied by a test agency, or a custom-built battery may be developed. The biggest pitfall to avoid is adding extra tests just for the sake of it, without ensuring that they make a proper contribution to the success of the predictions for which the battery is being used.

The CIPD (2007c) has noted that online testing is growing in popularity (25 per cent of respondents to their survey made some use of them). Online tests are most used for recruiting graduates and when high volumes of applicants have to be dealt with.

Good practice in psychological testing

The British Psychological Society and The International Test Commission have respectively produced guidelines on the general use of psychological tests and on the use of online testing.

General use of tests

People who use psychological tests are expected by the British Psychological Society (2007) to:

1. Take steps to ensure that they are able to meet all the standards of competence defined by the British Psychological Society for the relevant Certificate(s) of Competence, and to endeavour, where possible, to develop and enhance their competence as test users.

2. Monitor the limits of their competence in psychometric testing and not to offer services that lie outside their competence nor encourage or cause others to do so.

3. Use tests only in conjunction with other assessment methods and only when their use can be supported by the available technical information.

4. Administer, score and interpret tests in accordance with the instructions provided by the test distributor and to the standards defined by the British Psychological Society.

5. Store test materials securely and to ensure that no unqualified person has access to them.

6. Ensure test results are stored securely, are not accessible to unauthorized or unqualified persons and are not used for any purposes other than those agreed with the test taker.

7. Obtain the informed consent of potential test takers, making sure that they understand why the tests will be used, what will be done with their results and who will be provided with access to them.

8. Ensure that all test takers are well informed and well prepared for the test session, and that all have had access to practice or familiarization materials where appropriate.

9. Give due consideration to factors such as gender, ethnicity, age, disability and special needs, educational background and level of ability in using and interpreting the results of tests.

10. Provide the test taker and other authorized persons with feedback about the results in a form that makes clear the implications of the results, is clear and in a style appropriate to their level of understanding.

Guidelines on the use of online tests

The key points made in the guidelines on the use of online tests produced by The International Test Commission (2005) are:

1. Only use websites supplied by test publishers who offer validated psychometric tests.

2. Tests alone may not provide a complete assessment of an individual as other confirmatory or ancillary information is not included.

3. Provide test-takers with clear instructions on how to take the test.

4. Provide relevant feedback to test-takers.

5. When individuals take an unsupervised test, procedures should be used in the form of a confirmatory test to check whether the test-taker's original responses are consistent with the responses to the confirmatory test.

Selection tests – key learning points

The types of tests

Intelligence, ability, personality and aptitude.

The characteristics of a good test

- It is a sensitive measuring instrument.

Selection tests – key learning points (continued)

- It has been standardized on a representative and sizeable sample of the population for which it is intended.

- It is reliable in the sense that it always measures the same thing.

- It is valid in the sense that it measures the characteristic the test is intended to measure.

Test validity

The five types of validity are: predictive, concurrent, content, face and construct.

Interpreting test results

Test results can be interpreted by the use of norms or through criterion scores.

Choosing tests

It is essential to choose tests that meet the four criteria of sensitivity, standardization, reliability and validity.

Using tests in a selection procedure

While validated intelligence and personality tests can produce useful data they should not be relied on entirely. It is best to combine them in a selection procedure with structured interviews.

Questions

1. What does the term 'validity' mean when applied to selection tests? How can it be measured?

2. What are the advantages and disadvantages of personality tests as a method of selection?

3. From a colleague: 'I have just come back from a spell in our French associated company where they swear by graphology as a method of selection. Is there anything in it for us?'

References

British Psychological Society (2007) *Psychological Testing: A user's guide*, Psychological Testing Centre, Leicester

CIPD (2007c) *Psychological Testing*, CIPD Fact Sheet, www.cipd.co.uk

International Test Commission (2005) *International Guidelines on Computer-based and Internet Delivered Testing*, British Psychological Society, Leicester

Schmidt, F L and Hunter, J E (1998) The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings, *Psychological Bulletin*, **124** (2), pp 262–74

Smith, M (1984) *Survey Item Blank*, MCB Publications, Bradford

Spearman, C (1927) *The Abilities of Man*, Macmillan, New York