

Selection tests

Selection tests are used to provide more valid and reliable evidence of levels of intelligence, personality characteristics, abilities, aptitudes and attainments than can be obtained from an interview. This chapter is mainly concerned with psychological tests of intelligence or personality as defined below, but it also refers to the principal tests of ability etc that can be used.

PSYCHOLOGICAL TESTS: DEFINITION

As defined by Smith and Robertson (1986), a psychological test is:

A carefully chosen, systematic and standardised procedure for evolving a sample of responses from candidates which can be used to assess one or more of their psychological characteristics with those of a representative sample of an appropriate population.

PURPOSE OF PSYCHOLOGICAL TESTS

Psychological tests are measuring instruments, which is why they are often referred to as psychometric tests. Psychometric literally means 'mental measurement'.

The purpose of a psychological test is to provide an objective means of measuring

individual abilities or characteristics. They are used to enable selectors to gain a greater understanding of individuals so that they can predict the extent to which they will be successful in a job.

CHARACTERISTICS OF A GOOD TEST

A good test is one that provides valid data that enable reliable predictions of behaviour to be made, and therefore assist in the process of making objective and reasoned decisions when selecting people for jobs. It will be based on thorough research that has produced standardized criteria that have been derived by using the same measure to test a number of representative people to produce a set of 'norms'. 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 are:

- 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 that 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.

There are five types of validity:

- *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.
- *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.

- *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.
- *Face validity* – the extent to which it is felt that the test ‘looks’ right, ie is measuring what it is supposed to measure.
- *Construct validity* – the extent to which the test measures a particular construct or characteristic. As Edenborough (1994) suggests, 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. This may be difficult and Smith and Robertson (1986) emphasize that 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.

Validity can be expressed as a coefficient of correlation 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 on whether a validity coefficient is big enough was produced by Smith (1984):

over 0.5 excellent
 0.40-0.49 good
 0.30-0.39 acceptable
 less than 0.30 poor

On this basis, only ability tests, biodata and (according to Smith’s figures) personality questionnaires reach acceptable levels of validity.

TYPES OF TEST

The main types of selection test as described below are intelligence, personality, ability, aptitude and attainment tests.

A distinction can be made between psychometric tests and psychometric questionnaires. As explained by Toplis *et al* (1991), a psychometric test such as one on mental ability has correct answers so that the higher the score, the better the performance. Psychometric questionnaires such as personality tests assess habitual performance

and measure personality characteristics, interests, values or behaviour. With questionnaires, a high or low score signifies the extent to which a person has a certain quality and the appropriateness of the replies depends on the particular qualities required in the job to be filled.

Intelligence tests

Tests of intelligence such as Raven's Progressive Matrices measure general intelligence (termed 'g' by Spearman (1927), one of the pioneers of intelligence testing). Intelligence is defined by Toplis *et al* (1991) as 'the capacity for abstract thinking and reasoning'. The difficulty with intelligence tests is that they have to be based on a theory of what constitutes intelligence and then have to derive a series of verbal and non-verbal instruments for measuring the different factors or constituents of intelligence. But intelligence is a highly complex concept and the variety of theories about intelligence and the consequent variations in the test instruments or batteries available make the choice of an intelligence test a difficult one.

For general selection purposes, an intelligence test that can be administered to a group of candidates is the best, especially if it has been properly validated, and it is possible to relate test scores to 'norms' in such a way as to indicate how the individual taking the test compares with the rest of the population, in general or in a specific area.

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. As summarized by McCrae and Costa (1989), this model defines the key personality characteristics. These 'big five', as Roberts (1997) calls them, are:

- *extraversion/introversion* – gregarious, outgoing, assertive, talkative and active (extraversion); or reserved, inward-looking, diffident, quiet, restrained (introversion);
- *emotional stability* – resilient, independent, confident, relaxed; or apprehensive, dependent, under-confident, tense;

- *agreeableness* – courteous, cooperative, likeable, tolerant; or rude, uncooperative, hostile, intolerant;
- *conscientiousness* – hard-working, persevering, careful, reliable; or lazy, dilettante, careless, expedient;
- *openness to experience* – curious, imaginative, willingness to learn, broad-minded; or blinkered, unimaginative, complacent, narrow-minded.

Research cited by Roberts (1997) has indicated that these factors are valid predictors of work performance and that one factor in particular, ‘conscientiousness’, was very effective.

Self-report personality questionnaires are the ones most 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.

Specific work behaviour questionnaires cover behaviours such as leadership or selling.

Personality questionnaires were shown to have the low validity coefficient of 0.15 on the basis of research conducted by Schmitt *et al* (1984). But as Saville and Sik (1992) point out, this was based on a rag-bag of tests, many developed for clinical use and some using ‘projective’ techniques such as the Rorschach inkblots test, the interpretation of which relies on a clinician’s judgement and is therefore quite out of place in a modern selection procedure. Smith’s (1988) studies based on modern self-report questionnaires revealed an average validity coefficient of 0.39, which is reasonably high.

A vigorous attack was launched on personality tests by Blinkorn and Johnson (1990). They commented: ‘We see precious little evidence of personality tests predicting job performance.’ But Fletcher (1991) responded: ‘Like any other selection

procedure, they (psychometric tests) can be used well or badly. But it would be foolish to dismiss all the evidence of the value of personality assessment in selection on the basis of some misuse. Certainly the majority of applied psychologists feel the balance of the evidence supports the use of personality inventories.' 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.

Ability tests

Ability tests measure job-related characteristics such as number, verbal, perceptual or mechanical ability.

Aptitude tests

Aptitude tests are job-specific tests that are designed to predict the potential an individual has to perform tasks within a job. They can cover such areas as clerical aptitude, numerical aptitude, mechanical aptitude and dexterity.

Aptitude tests should be properly validated. The usual procedure is to determine the aptitudes required by means of job and skills analysis. A standard test or a test battery is then obtained from a test agency. Alternatively, a special test is devised by or for the organization. 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. Many do-it-yourself tests are worse than useless because they have not been properly validated.

Attainment tests

Attainment tests measure abilities or skills that have already been acquired by training or experience. A typing test is the most typical example. It is easy to find out how many words a minute a typist can type and compare that with the standard required for the job.

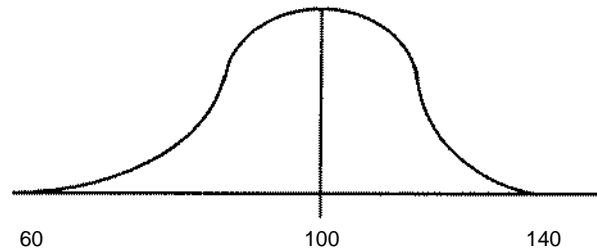


Figure 29.1 A normal curve

INTERPRETING TEST RESULTS

The two main methods of interpreting test results are the use of norms and the normal curve.

Norms

Tests can be interpreted in terms of how an individual's results compare with the scores achieved by a group on whom the task was standardized – the norm or reference group. A normative score is read from a norms table. The most common scale indicates the proportion of the reference who scored less than the individual. Thus if someone scored at the 70th percentile in a test, that person's score would be better than 65 per cent of the reference group.

The normal curve

The normal curve describes the relationship between a set of observations and measures and the frequency of their occurrence. It indicates, as illustrated in Figure 29.1, that on many things 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.

The most important characteristic of the normal curve 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 , 15 per cent are likely to get scores below x and 15 per cent are likely to get more than y .

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. Generally speaking, it is best to avoid using them.

THE USE OF TESTS IN A SELECTION PROCEDURE

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 and attainment tests are most useful for jobs where specific and measurable skills are required, such as typing or computer programming. Personality tests are potentially of greatest value in jobs such as selling where 'personality' is important, and where it is not too difficult to obtain quantifiable criteria for validation purposes.

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, aptitude and personality 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 six criteria for the use of psychological tests produced by the IPD (1997a) are:

1. Everyone responsible for the application of tests including evaluation, interpretation and feedback should be trained at least to the level of competence recommended by the British Psychological Society.
2. Potential test users should satisfy themselves that it is appropriate to use tests at all before incorporating tests into their decision-making processes.
3. Users must satisfy themselves that any tests they decide to use actively measure factors which are directly relevant to the employment situation.
4. Users must satisfy themselves that all tests they use should have been rigorously developed and that claims about their reliability, validity and effectiveness are supported by statistical evidence (*The Data Protection Act 1998 is relevant here. If candidates are selected on the basis of a test they have the right to know the rationale for the selection decision.*)
5. Care must be taken to provide equality of opportunity among all individuals required to take tests.
6. The results of single tests should not be used as the sole basis for decision-making. This is particularly relevant with regard to personality tests.