Talent in Innovation.
Innovation in Talent.

Guidelines for Best Practice in Implementing Assessments

Foreword

Psychometric assessments are used as method to support the decision making process in almost every area of staff development. From selection and recruitment to promotion and succession, research has demonstrated that well-constructed psychometric tests predict job performance better than almost any other single selection method.

Psychometric assessments provide insight into the likely intellectual ability, personality and behaviours of individuals; the way they think, how they relate to others and their working style. The application of tools which are fit for purpose will deliver information towards more informed business decisions. As with any data, it is recommended that managers be guided in their understanding of the output from psychometric assessments, utilising information in conjunction with that from other sources, to validate impressions, align with values and capabilities, and highlight areas for further investigation. In addition to supporting recruitment decisions, insights gained from relevant application of assessment tools can benefit leadership, engagement and development strategies, team enhancement and performance conversations.

The sections that follow present some of our best practice recommendations for appropriate identification and use of assessments. These recommendations are based on our extensive experience in assessment design and implementation, combined with current professional and legal guidelines in this area.[1]

SHL's assessment selection tools and its professional consulting services can help organisations establish hiring and promotion processes with demonstrated validity and the potential for a positive return on investment. Ultimately these enhanced processes can help organisations identify and retain top talent and have a positive impact on business results.

[1] This guide does not constitute legal advice and clients should consult their local counsel for such. To this end a list of resources has been provided in the Appendices.

Contents

What are Psychological Tests?

“A psychological test is a systematic procedure for obtaining samples of behaviour, relevant to cognitive or affective functioning, and for scoring and evaluating those samples according to standards” (Urbina, 2004). As such they powerful tools used by organisations for the selection, development and management of people. Occupational tests have consistently been shown to be better predictors of job success than most other techniques (Dipboye, 2016). Using tests can lead to substantial gains for an organisation in terms of increased output and efficiency, better quality staff, higher morale, more effective performance, lower training costs and reduced turnover.

Specifically, tests can be used:

- To enhance the decision-making process in assessment for selection and promotion
- As an aid to management in areas such as motivation and team building
- To identify development needs, as a basis for employee counselling, as well as
- In organisational areas such as management of change or succession planning

Using Psychological Tests

Psychometric tests provide additional relevant information over and above that obtained from more traditional assessment methods. Relevant tests produce more accurate results than common selection measures such as interviews as they give objective information about a candidate, and have been shown in general to lead to better and fairer employment decisions (Robertson, 2001). This is due in large part to the standardisation of psychometric assessments. Psychological tests are often described as standardised for two reasons, both of which address the need for objectivity in the testing process.

The first has to do with uniformity of procedure in all important aspects of the administration, scoring, and interpretation of tests. The purpose of standardising test procedures is to make all the variables that are under the control of the assessor as uniform as possible, so that everyone who takes the test will be taking it in the same way. To this end there are an array of international, regional, and organisational policies that outline the appropriate way in which to implement psychometric assessments.

The second meaning of standardisation concerns the use of standards for evaluating test results. These standards are most often norms derived from a group of individuals—known as the normative or standardisation sample—in the process of developing the test. The collective performance of the standardisation group or groups, both in terms of averages and variability, is tabulated and becomes the standard against which the performance of other individuals who take the test after it is standardised will be gauged (Domino & Domino, 2006).

Thus, in order to sustain the credibility of psychological tests and psychometric tools, it is necessary that best practice around implementation results in informed policy decisions and outlines the psychometric requirements to maintain the integrity of the tools used. The following notes provide guidance on developing an assessment policy for your organisation and provides guidelines to ensure that tests are used appropriately.

Assessment Policy Management

The use of tests must be properly managed in order to reap all the potential benefits identified above, retain the standing of assessments, and to avoid the ramifications of bad practice. As practitioners of psychological assessments there are certain ethical and legal obligations which are paramount to the successful implementation of assessments.

As with any tool, the potential for misuse is ever present. Bad practice can:

- Reduce the effectiveness of test use
- Cause considerable suffering to individuals
- Damage the company's image with employees, trade unions and clients
- Lead to contravention of the law, and the costs and bad publicity associated with litigation.

A well-thought out policy on test use will help to ensure that the organisation gains maximum benefit from test use and that potential misuse is avoided. It will also demonstrate commitment to good practice. The management of tests should consistently be under review and monitored to ensure the organisational aims, as well as ethical responsibilities, are being met.

The aims of every organisation in using tests are different and policy must be geared to the needs and structures of each organisation. Additionally, regional requirements differ pointedly in some instances. For this reason each organisation must develop its own way of managing test use and formulate a local policy document. This guide serves to indicate the factors that need to be considered in such a policy and provides resources around regional requirements. It also provides examples of what policy statements may look like. Readers are free to use or modify these as appropriate, but the statements do not in themselves constitute a policy.

What should an assessment policy cover?

The test user has a duty to the test-taker to ensure the entire process is as positive as it can be. This will mean choosing the right test, with good psychometric properties, administering it competently, scoring it in a standardised manner and then accurately interpreting those scores. Assessment policies should encompass all of these dimensions.

What should a policy look like?

Policies work best when they are brief and focused. Long manuals that lie unread in a drawer have little effect on practice. However, general statements of principle tend to be vague, easily ignored and provide little guidance for the user.

An effective strategy is to have brief policy statements supplemented with more detailed guidelines. The statements can encompass general principles and are likely to need few changes even in changing circumstances.

The procedural guidelines put the policy into practice, and then can, and should, be amended whenever they cease to meet the aims of the organisation or the policy. The detail with which such procedures are specified will depend on the nature, style and needs of the organisation.

Mission Statement

It is helpful to start with a short general statement of the aims an organisation intends to achieve in using tests. This sets the tone and purpose of the document. It ensures that readers are aware of the positive objectives of the use of tests and the reasons for the existence of the policy. Otherwise, a policy can be seen as restrictive and bureaucratic, rather than as a helpful guide to doing things well.

Example Call Out Box

We use psychometric tests to enhance the quality and quantity of information available for selection, development and training decisions and as an aid to organisational change. We are committed to the highest standards of practice in the use of all psychometric tests, in order to maximise the benefit of testing to the organisation and the individual, and to promote fairness and equality of opportunity for all.

Overall Responsibility for Testing Standards

Test users are always responsible for the way they personally use tests. However, in some organisations there is a central unit or department that decides policy, oversees practices and provides support for individual users. Often such a unit contains one or more professional IO psychologists. Sometimes such units have an advisory role; at other times they dictate procedures.

In other organisations each test user makes his or her own autonomous decisions regarding the use of tests; small organisations may only have a few trained users; large organisations may have many users in a single personnel department or test use may be decentralised.

Whatever the structure, the responsibilities and accountabilities of each individual must be clearly defined. It is highly recommended that in large organisations with high volume test use, one or more IO psychologists are available to support the design, implementation, validation and review of test procedures either on a consultancy basis or as permanent members of staff.

Example

Each test user must ensure that he / she uses tests to the highest professional standards and only in accordance with the guidelines set out in this policy.

The Central Testing Unit is responsible for ensuring that all test use in this organisation is in accordance with this policy. Any procedure that will involve the use of psychometric testing must be referred to the Unit for approval before implementation. The Unit will provide guidelines for the use of tests that must be followed by all users.

Who should use tests?

Knowledge and experience are required to use psychometric tests effectively. It is recognised throughout the world that psychometric instruments are potentially dangerous in the wrong hands. Indeed, in many countries only qualified psychologists are allowed to use them. In this regard it is necessary to consult with the policy and guidelines of the country or region you work in (see appendices). In the absence of country or region specific guidelines one should refer to the International Test Commission's Guidelines on Test Use.

Example

Only trained test users who hold the relevant qualifications may use and interpret psychometric instruments. Trained test users may delegate test administration to a person trained in this area.

When should tests be used?

Tests may be used for selection, (either for shortlisting or final decisions), placement or promotion decisions, development, team building, counselling, out-placement and organisational development. It is not possible to discuss the considerations for using tests in each context in this short booklet. In each case, the situation must be evaluated to see whether test use would be appropriate to help achieve the desired objectives.

There are some occasions where it is not usually appropriate to use tests. For instance, it is unlikely that test results would be suitable for making redundancy decisions, since direct information on job performance should be available.

However, tests can be valuable in making redeployment decisions or in outplacement counselling. Similarly, an organisation may want to restrict the use of some tests to counselling or development applications.

Tests are best used in decision making in conjunction with other relevant information. In a promotion decision, test results may be integrated with an interview, performance, track record and managers' recommendations to provide the best information about individual suitability. Use of a single test result alone should be avoided whenever possible.

Consideration must be given to where test results fit into a procedure. Tests may be used for shortlisting from a large pool of applicants; to suggest areas to be explored during an interview; or as a final check on the suitability of a chosen candidate.

Example

Tests may be used for selection, development and counselling purposes. Any additional uses should be referred to the Central Testing Unit for approval.

Test Choice

Whenever tests are used, it is vital that there is a match between the skills and characteristics measured and the job or organisational demands. This is particularly important when selection or promotion decisions are based on test results.

Objective job analysis is the best way to determine the skills required for a particular job. These skills are then matched to appropriate tests. In large scale testing procedures, it may be appropriate to perform criterion related validation studies before test implementation.

Trained test users should have the skills to evaluate the quality and relevance to job requirements of an instrument from the information provided by the test publisher in the test manual. Where insufficient information is provided, caution should be exercised in the use of the test.

Some organisations may allow individual users to select the test they want to use. Others may require the involvement of internal or external experts in test choice. A middle way would be the provision of a general list of 'approved' tests from which individual users could choose those appropriate to their needs. This allows an organisation to prevent the use of instruments of uncertain quality. It will not control the relevance of measures - in other words the job analysis stage will still be required.

It is important that whenever tests are chosen there is written documentation of the reasons behind the choice. This may include copies of job analysis reports, job descriptions, person specifications, validation studies, etc. If the relevance of a particular measure is challenged, such evidence supports the test choice, shows the care taken and helps ensure users do not take dangerous shortcuts.

Example

All psychometric tests used must be clearly relevant to the given purpose. Detailed job descriptions and person specifications based on objective job analysis must be prepared prior to the choice of tests for any selection or promotion procedures. All decisions to use tests should be clearly documented with a copy sent to the Central Testing Unit.

Equal Opportunities

Research has shown that well-constructed psychometric tests are the single most effective predictor of job performance. Tests give objective information about a candidate and have been shown in general to lead to better and fairer employment decisions.

Tests of aptitude or ability have sometimes been found to have disparate impact on ethnic, age or gender group; i.e. proportionately fewer members of one ethnic, age or gender group do well in the test in some cases. In that these patterns exist, it is particularly important that appropriate guidelines are followed to avoid improper use of tests and potential contravention of antidiscrimination laws. Considerations of fairness are important in themselves. In addition, there are legal implications of unfair practices in the selection and promotion of employees.

In the absence of validation evidence, there is likely to be a presumption that the group with the lower average performance was being indirectly discriminated against. That is, if the same entry standard were demanded of all applicants, the lower scoring group would find it harder to comply with the requirement.

Positive validation evidence for the test generally justifies the use of the test and rules out the possibility of unfair discrimination. By showing that those who perform poorly on the test also perform poorly on the job, a positive validation result confirms that rejecting low scorers is reasonable.

The greater the degree of adverse impact resulting from the use of a test, the higher the validity of the test should be to justify its use. There remains the possibility that overall validity is masking cases where a test has poorer or no predictive validity for some groups, or that group differences in test scores are not reflected in job performance.

It should be remembered that group differences relate to average performance. Even where there are substantial group differences, there will be members of the lower scoring group who have higher scores than many people from the higher scoring group and vice versa.

Furthermore, job success does not generally depend on a single ability and tests do not have perfect predictive power. Therefore, on occasion, low scorers on a test will do better on a job than a test score may suggest. For this reason, it is preferable to interpret test scores together with other available information.

Users should take particular care when a candidate is not fluent in the language in which the tests are presented. Actual ability may be confounded with language proficiency and results may be difficult to interpret.

Similar issues arise in the testing of candidates with disabilities. In these cases, it may be necessary to adjust standardised administration procedures to allow for the disability, as typically required by legislation designed to protect against disability discrimination. Changes should not be made arbitrarily as this affects test standardisation.

Test users can call the SHL Helpline for advice. Our Guidelines for Best Practice in Testing People with Disabilities is also available from SHL to help in this process.

Example

The organisation is committed to selection on merit and only measures which are clearly relevant to job demands and free of extraneous bias should be used. All assessments for selection and promotion must be monitored to ensure they do not unfairly exclude or disadvantage any section of the population.

Whenever a disability prevents a suitably qualified individual from undergoing standard selection procedures, appropriate alternative arrangements for assessment must be found. Always contact the test publisher for advice before making any changes to test administration procedures.

Use of Test Scores

Interpretation of results should be accurate and not exceed the limits of the information. Users should beware, for example, of failing to take account of the standard error of measurement in interpreting the difference between any two scores, or of drawing far-fetched inferences from personality data.

Appropriate norms should always be used in interpreting scores. Where no suitable norm groups are provided with a test, further guidance should be sought from the publisher.

Tests should always be interpreted by properly trained individuals in the context of clearly defined criteria. Both quantitative and qualitative interpretation can be used. The former should be restricted to cases where there is sufficient supporting evidence.

The use of fixed cut-offs with personality measures can be particularly misleading, without relevant evidence of validity. Qualitative interpretation may be more appropriate in these cases. Decision rules and their rationale should be properly documented.

Thought should be given to the integration of test results with other relevant information. This is particularly important where many different assessment techniques are used. Our Guidelines for Best Practice in the Use of Assessment and Development Centres discusses the issues that arise here.

However, even where only test results and an interview are combined, decisions need to be made about, for instance, whether good performance in one sphere can compensate for poor performance in the other, or whether some minimum standard needs to be reached on all elements. Similarly, in the use of personality questionnaires, users need to decide whether a criterion referenced, “danger zone” or integrative approach is to be used.

Example

Test scores must be interpreted on the basis of relevant norm groups. Fixed cut-offs may only be imposed where specific evidence of test relevance is available, e.g. job analysis, validation study.

The interpretation of test scores is in accordance with the formula, which will be amended from time to time on the basis of ongoing validation results.

Confidentiality and Storage of Results

Test results, like all personal information, should be stored with due regard to confidentiality. Access should be restricted to those with a need to know and in accordance with what has been agreed with the respondent during administration and feedback. Persons who are untrained should not be allowed access to raw data from tests, but only to clearly described interpretations.

Individuals do change and develop and so psychometric data can become less accurate over time. Test scores should therefore not be kept on file indefinitely. The time period for which scores are valid will differ depending on the nature of the measures and the particular use made of them. Care should be taken with results over 6-12 months old for selection purposes. Little reliance should be put on results over three years old for any purpose.

Data protection legislation varies from country to country, but typically covers the storage of any information about an individual, whether on computer or in another form. Test results should not be stored or used without candidates' permission and they typically have a right to see information stored or used, or a meaningful interpretation of it. As the leading global provider of Talent

Assessment solutions, we take our obligations with respect to ensuring the highest level of protection over our clients' personal data very seriously. Maintaining compliance with applicable data protection legislation continues to be our highest priority. For more information regarding our compliance with, and our programs to support our clients' compliance with the General Data Protection Regulation (GDPR) that came into effect 25 May 2018, please refer to our white paper entitled GDPR – What Do You Need to Know?

Example

Test results should be kept by test users in locked files. A written interpretation of results should be kept in personnel files and provided to relevant individuals.

Results over 12 months old are invalid for selection or promotion decisions. All results are to be destroyed after three years or when the respondent ceases to be employed, whichever is the sooner.

Retesting

An issue that arises where positions regularly become vacant and unsuccessful applicants reapply is whether they should (be allowed to) retake tests. There is no hard and fast rule but an organisation should have a consistent policy. Results can be allowed to stand for up to 12 months. It is not desirable to allow candidates to be retested regularly unless alternate forms of the test are available.

However, it is reasonable to allow an applicant to be retested where there is evidence that he / she might have under-performed the first time, e.g. due to illness. Otherwise, a suitable interval should generally elapse before retesting is allowed.

Example

For selection purposes a candidate's test results are valid for any similar position for 12 months from the date of testing. Candidates may be retested after six months at their request. Candidates may be retested within a shorter period only at the discretion of the relevant personnel manager and should supply details in writing supporting their application.

Monitoring

Use of tests and other psychometric instruments should be continually monitored to ensure continued appropriateness and effectiveness. In small scale applications, this may amount to ensuring that the techniques remain relevant to the job and that up-

to-date test versions and norms are used. Where larger scale use occurs, scores should be monitored at regular intervals to update norms.

Monitoring by ethnic group, age and gender is required to identify any adverse impact. A validation study should be carried out every five years or so, or whenever changes in the job or applicant group are such that initial validity could have been affected.

Monitoring might be the responsibility of local test users, or could be centrally co-ordinated. The performance of a validation study requires detailed knowledge beyond that generally gained in basic level training courses. Advice should be sought from a competent psychologist.

Where substantial disparate impact has been found, the issues listed below should be considered. Many are also relevant when tests have poor validity.

- Has the job changed in some way since the original job analysis?
- Are the skills measured really relevant to the job?
- Is the way the skills are measured, e.g. language used, speed, appropriate to all candidates?
- Would customisation of the context in which skills are measured help?
- Are the tests at the right level for the job?
- Are the tests at the right level for the applicant population?
- Are the tests predicting job performance for all groups?
- Is there another test type which would provide the same information with better validity / without disparate impact?
- Can the selection rule be designed to minimise disparate impact, despite score differences?
- Can the job or training be redesigned so that the entry level required for the relevant skill or ability is lower?
- Can disadvantaged groups be trained to offset differences in test-taking behaviour or to enhance relevant skills?

Example

Test monitoring forms should be completed for each exercise involving test use and sent to the Central Testing Unit for processing. Whenever more than 100 people are employed in a job category for which tests are used in selection, a validation study must be performed within three years from the commencement of the use of tests.

Access to Materials

This element of a policy should cover who can buy materials, where they are stored and who has access. The security of materials is paramount. Free circulation leads to over familiarity and devalues psychometric instruments. Responsible test publishers only supply materials to trained users, who, in turn, must ensure untrained users do not gain access to them.

Best practices governing the storage of test materials for online tests are different than those governing fixed or paper / pencil tests. Policies should still exist that govern who can buy materials, where they are stored and who has access, but the availability of the tests and results online may expand access to other users inside the organisation that would not normally have had access to test materials in the past.

For most online tests today, scoring keys or other sensitive scoring information are no longer available in a distributable form. Therefore, organisations are unlikely to have scoring keys on site that must be protected and kept under lock and key. If scoring keys are available, all existing best practice should apply to the protection of these materials.

If online tests are fixed tests, appropriate due diligence and policy should be developed internally to ensure that test items and content are not inappropriately copied and distributed for unauthorised use. Whenever possible, randomised tests should be used to manage the potential risk of fixed test items being inappropriately copied or disclosed.

Most reputable test providers have updated their guidance to allow for the access and use of online tests and materials by minimally trained users, as long as the tests are designed for online use and have been selected and implemented under the advice of a qualified test user.

Because of the potential for wider use and access to tests and test materials from online distribution, it is critical that organisations select tests that are developed and designed for online use. This includes selecting tests that provide results or outputs that would be appropriate for minimally trained users to use safely.

Access to computer-based testing should be controlled by restricted passwords or other appropriate means.

> **Example**
>
> All access to the online assessment system should be controlled by suitably trained system administrators each with their own username and password to the system. Usernames and passwords should be kept confidential at all times, changed regularly and may not be shared by multiple users.

Copyright

Test materials are extremely vulnerable to copyright infringement. In most countries, the reproduction of test materials by any means (including computer installations) without the permission of the author is a criminal offence, whether or not the reproduced materials are to be sold. Illegal copying of materials leads to lack of standardisation and poor control of materials, and gives respondents a bad impression. Ultimately, the resulting loss of income will contribute to less new test development, poorer updating services, or higher prices.

The responsibility for obtaining the publisher's permission to install a particular test on a generic computer "shell system" rests with the user, not the supplier, of such a system. The user will be in breach of copyright if permission from the publisher has not been obtained.

All SHL materials, including profile charts and software supplied on computer installations, are subject to copyright. SHL has in the past actively pursued potential breaches of copyright where these are discovered and will continue to do so in the future.

> **Example**
>
> Under no circumstances should any test materials be photocopied or installed on computer without the test publisher's express permission.

Policy Management

Thought should be given to the implementation procedure for the policy, to ensure that test users are aware of the requirements and are committed to working within them. Communication is essential in this process. Test users should understand the function of the policy in preserving high standards.

Commitment from the top is essential in implementing a policy successfully. Managers should ensure that they promote a supportive atmosphere towards the policy and actively implement it themselves.

A review procedure should form part of the policy to ensure it remains relevant to the needs of the organisation. This should specify when and how the workings of the policy will be examined and how changes are to be made.

Choosing the Right Test

Use assessment tools in a purposeful manner. The first thing to consider when implementing assessments is whether they are fit for purpose. It is critical to have a clear understanding of what needs to be measured and for what purpose.

Assessment instruments, like other tools, can be extremely helpful when used properly, but counterproductive when used inappropriately. Often inappropriate use stems from not having a clear understanding of what you want to measure and why you want to measure it. Having a clear understanding of the purpose of your assessment system is important in selecting the appropriate assessment tools to meet that purpose.

Job Analysis

Assessment strategies should be developed with a clear understanding of the knowledge, skills, abilities, characteristics, or personal traits you want to measure. Whether to use a test or not, what test to use, or whether to construct a new assessment or not, should be based on the results of an objective job analysis or pre-existing job description. A job analysis will permit the employer to identify and document the primary work activities, competencies, knowledge, skills, abilities, and other characteristics (KSAOs) required for successful job performance, which can then be used to identify appropriate tests (Brannick & Levine, 2002).

Conducting a job analysis may include asking job experts (e.g., incumbent employees or supervisors of those target employees) to describe the requirements of a job through focus groups or interviews that may include direct observations and / or data collection through surveys or related methods. Job experts from various facets of the job should be selected (e.g., from various shifts, locations, incumbent demographics, etc.). Data should be analysed to determine competencies required for successful performance on the job and to further identify which of those competencies new candidates need to bring to the job (compared to those on which they will be trained). Additionally, important work behaviours should be identified from analysis of survey data.

Job analysis is critical for two important reasons:

1. It assists test users in choosing a relevant assessment or combination of assessments to identify candidates who are most likely to be successful
2. It is a critical part of validating and defending a selection system if a legal challenge is made

In large-scale testing procedures, it may be appropriate to perform criterion-related validation studies before test implementation.

It is also essential to know what each assessment tool you are considering using is designed to measure. The best source for this information is generally the assessment manual. When insufficient information is provided, caution should be exercised in the use of the test.

It is important that whenever tests are chosen there is written documentation of the reasons behind the choice. This may include copies of job analysis reports, job descriptions, person specifications, validation studies, etc. If the relevance of a particular measure is challenged, such evidence supports the test choice, shows the care taken, and helps ensure users do not take shortcuts within the process.

Customisation

In some instances, an assessor will decide to develop a custom assessment to more accurately assess for a particular role or function. A customised test is one which is constructed especially for a particular use. Customisation enables the production of a test measuring a skill and in a context which are both tailored to organisational needs. Customised tests have the following advantages and disadvantages:

Advantages:

- The design of the tests is likely to be well suited to the intended use. The skill tested, as well as the level and content of the test, have the best relationship to the job analysis. Validity is therefore enhanced.
- In designing the tests, account can be taken of the probable knowledge and background of likely applicants where this information is available.

- The tests are likely to seem most fair to candidates, in that they reflect the content of the job.
- The security of the test material is under the control of the commissioning organisation. Candidates will not have seen the tests before and therefore cannot benefit from specific practice effects.
- Although the initial outlay is greater, where large numbers of people are tested, customised tests can be more economical in the longer term than an off the-shelf test.

Disadvantages:

- Customised tests require development time – often six months or more – before being available for use.
- They require relevant trial groups of several hundreds to be available.
- Good test publishers ensure that the off-the-shelf tests they distribute are monitored and updated. The organisation would be responsible for commissioning this work with customised tests.
- Should the job or applicant population for which the tests were designed change, the tests may become obsolete.
- External norm groups will not be available so no comparisons can be made with other groups. Internal norms have to be developed.

Making Sure the Assessment is Psychometrically Sound

Tests should be psychometrically sound. A full description of what this entails is beyond these guidelines but may be found in textbooks on psychometrics (see attached a list of referrals). The relevant information and statistics for judging the tests should appear in the test manual. These should include:

Test Content

Two aspects of test content should be considered. The most important is the actual skill or attribute being measured by the test, e.g. verbal comprehension, manual dexterity. All the tests used should measure skills or attributes identified as necessary to do the job. These facets are termed the inherent requirements of the job. If skills not required within the role (not inherent to the job) are assessed and used as a means to filter out candidates there are severe legal implications that arise in most areas of the world. Additionally, and possibly more importantly, it results in unethical discriminatory practices.

By law, the employer may be required to show that tests used correspond to a real need, are appropriate with a view to achieving the objective, and are necessary to that end.

The second aspect of test content is the context in which the skill is measured. This should, as far as possible, reflect the type of content found in the job. For example, a typing test should require the typing of material similar to that required on the job.

However, care must be taken not to include material requiring knowledge specific to the organisation, which would put external applicants at an unfair disadvantage.

Test content that is of a more general nature should be equally accessible to all applicant groups - men and women, people from ethnic minority groups or older and younger people. For instance, in the typing test, if job relevant text is too technical for an external applicant to deal with before training, more general text should be used. This should be of content matter equally familiar to all groups.

Test Level

The level of difficulty at which the skill is measured should be appropriate to the job. A test which is too easy will not differentiate between applicants with good and poor potential. One that is too difficult could lead to greater adverse impact. The level of the test should also be appropriate for the likely applicant pool. If the general level of applicants is below the standard required for the job, employers should consider what they can do to attract better applicants.

Details of the Developmental Process

Test manuals should include information around the developmental process of an assessment. This should outline the relevant research that was conducted around the construct, methods of trialling and verifying concept models, statistical analysis and investigation, standardisation procedures, etc. It is important that the assessor has the requisite knowledge to assess the merit of this information, and make appropriate decisions about the utility of a test as a result.

Do not judge a test solely by how widely it is used; a test can become outdated, and companies sometimes use inappropriate or even bad poorly-constructed tests.

Reliability and Validation Considerations

Reliability

Reliability is concerned with how accurate or precise a test score is. When a test is administered, the outcome is an observed score on the quality measured by the test. However, all measurement procedures are subject to some degree of error. In order to know how much weight to place on the observed score, you need to know how accurate the test is as a measuring device. Measures of test reliability allow us to estimate that accuracy (British Psychological Society, 2007).

Before we use a personality questionnaire we must ensure that it gives an accurate and consistent measurement each time it is used.

How do I know a reliable questionnaire from an unreliable one?

Good psychometric tests should have been tested for reliability by the test publisher. There are several types of reliability, the one used most commonly for personality questionnaires is the internal consistency alpha coefficient frequently represented as 'α', or sometimes 'r' in the technical manuals. For ability tests reliability should always be higher than 0.7. The optimum range of coefficients for personality dimensions lies in the range 0.6 to 0.8, i.e. not too high or too low, to allow for the breadth of the personality dimension.

Where can this information be found?

This value should be in the test documentation, or available from the test publisher and should always be checked before using the test. All SHL tests are rigorously tested and the reliability coefficients obtained explicitly stated in the technical manual for each test battery.

Validity

On the other hand, validity is concerned with what the test score actually measures. Validity offers the meaningful statistical information that demonstrates whether a test score is actually representative of the quality or qualities that it purports to measure (British Psychological Society, 2007).

Test performance must be related to job performance for the test to be effective. Wherever possible test users should be carrying out a validation study to establish that the tests correlate with job performance or other relevant criteria such as labour turnover. Such a study shows whether higher scorers on the test tend to be more successful on the job. Validation is imperative if a test is to be used for purposes other than those for which it was designed.

Methods of Assessment Validation

It is important to use methods to validate assessments that are supported by current legal and professional standards, including content validation, criterion-related validation, and / or validity transportability. The feasibility of each approach may depend on the assessment in question, the validity evidence that has already been accumulated for that assessment, and the nature of the job or setting in which the assessment may be used.

Content Validity

A content-related validity strategy focuses on demonstrating that the content of the assessment is relevant to the work requirements on the target job. For example, a word processing skills assessment can be validated for an administrative assistant job by showing that the operation of word processing software, as measured by the assessment, is an important work requirement of the job. The focus of this validation strategy is on demonstrating the correspondence between the tasks or competencies required by the assessment and the tasks or competencies performed on the job. SHL typically establishes this link between

assessment content and job requirements based on carefully documented judgments made by subject matter experts (e.g., incumbent employees and immediate supervisors), both in the original design of the assessment and through evaluation by job experts when examining the relevance of an assessment for a particular job.

Criterion-Related Validity

Another indicator of validity is the degree to which assessment scores are related to an important outcome for the job or organisation – typically some measure of employee performance on the job or an organisational outcome such as job tenure. A criterion-related validity strategy investigates whether there is a significant statistical relationship between assessment scores and "criterion" measures such as job performance, training performance, and / or job tenure. If a significant relationship is observed, candidates with more favourable scores will be expected to demonstrate better performance on the job. Feasibility of this validation approach depends on a number of factors (e.g., sufficient numbers of job incumbents or candidates who can participate, availability of appropriate performance measures, etc.), and it is often used when a validity transportability study (described below) is not possible.

Validity Transportability

In many instances, it is possible to use criterion-related validity evidence accumulated from other assessments using a validity transportability approach. If an assessment has accumulated validity evidence across numerous studies, that supports their relationship with relevant metrics for a particular job or job family, it is possible to use validity transportability. A validity transportability strategy focuses on examining the similarity between jobs and the jobs for which criterion-related validity evidence has already been established for a particular assessment.

Validity transportability is appropriate when:

- substantial evidence of criterion-related validity has been established in other organisations, and
- there is substantial similarity between the job in question and the jobs that were included in the original criterion-related validity studies

When developing a new selection system for a target job, job analysis information gathered for that target job is compared to job analysis information gathered in prior validity studies. When job analysis evidence suggests that aspects of the target job are "substantially the same" as aspects of jobs found in archives, evidence of criterion-related validity obtained from these prior jobs can be "transported" for predicting how the assessment will work in the target job. A critical step in establishing jobs are "substantially the same" is demonstration of how important competencies and / or work behaviours of the jobs in the original validation studies are substantially similar to those important to performance for the job in question. If a clear relationship can be established between the jobs, then validity evidence (e.g., evidence that the assessment predicts performance) can be presumed to exist for the job in question.

The research evidence shows that, where a type of test has been shown to be valid for a particular job, it is likely to be valid for all similar jobs. This finding of validity generalisation means that where a test is to be used to select small numbers of people (say less than around 50), evidence of validity from similar job / test combinations in other places may be relied on.

When larger numbers of employees are to be selected, specific validation studies can be carried out. The performance of a robust validity study requires in-depth knowledge of practical issues and statistical procedures. Expert advice / assistance should be sought.

Validation studies should occur before tests are introduced as selection measures. There are two basic study designs.

In a concurrent study, job incumbents are tested and their performance evaluated at the same time. The relation between current performance on the test and in the job can then be examined.

Positive Aspects	Negative Aspects
This type of study is generally quite quick	Experience on the job sometimes affects test scores, making the results of the study less reliable

A **predictive study** avoids this problem by testing applicants before they start work. However, sometime must then elapse before performance ratings can be made. For this reason a predictive study will not yield results as quickly as a concurrent study.

Ideally, the test scores should not be used in making selection decisions until the validation is complete. In practice, however, the scores can be used conservatively while awaiting results, where not using them is likely to be detrimental to selection procedures.

In both these cases the test results should be monitored for any differences in scores by ethnic group, age or sex. Even where a test is not validated before use, administration to a pilot sample should be undertaken if a test is to be used with large numbers of applicants.

Test manuals and norms should also be consulted to see whether group differences are common and what has been done to ensure the test is fair. Although differences do not necessarily mean the test is biased, where they are found further investigation should follow.

Preparation of Candidates and Administration of Tests

While job analysis and validation are critical components to an effective and defensible selection system, consistent and appropriate implementation of the assessment process is also important. Consistency improves defensibility, and without a consistent and fair application of a selection program, return on investment can be compromised, and legal challenges are more likely to occur and less likely to be successfully defended.

The challenge in recent years however, has been that a vast majority of assessments occur in an online context and as a result are largely unsupervised. This poses certain challenges to maintaining consistency across testing conditions, ensuring that candidates are not getting assistance in completing assessments, and ensuring the security of the data being captured.

The following outlines the process for administering an assessment in both a supervised context and an unsupervised online setting. While particular focus is given to testing in an online unsupervised format, it is important not to discount the value of supervised assessments. Additionally, while this guide outlines the assessment process for unsupervised online assessments it is not an exhaustive guide of the challenges and opportunities associated with this method of assessment. For more detail on unsupervised online assessments, please refer to the Best Practice Guide on Guidelines for Best Practice in Unsupervised Online Assessment.

Before the Assessment

Whenever tests or questionnaires are used it is important that respondents are given clear information about the nature of the instruments and the reason for using them. In the instance of online assessments, this can done through a web site, web pages that precede and follow the assessment, or through information provided by email prior to an assessment.

The key points that need to be included in this communication supporting an assessment are outlined below:

Describe what the objectives of the assessment are
This should be an explanation and not a justification or defence, and should cover the relevance of the assessment to the job, role or position.
Explain why these assessments have been selected
Reiterate the relevance of the assessments to the job, role or position.
Explain the process
Communicate what will happen once the candidate has registered, completed the assessment, and what will happen once their results have been received. Also explain what the ‘authentication policy’ is. While verification is about the validity of scores, authentication is about validating the identity of the candidate. Many organisations have now adopted authentication as an essential part of their selection processes, for example requiring the candidate to provide some form of identification, preferably a legal document containing photographic ID, such as a passport or driver’s licence.
Explain the rules for the assessment
This should include confidentiality of assessment results and reports, when and in what form feedback or decisions will be communicated to them, and what they are signing up to, including what the consequences will be of fraudulent test taking or cheating.
Include an honesty agreement
A key step in ensuring that candidates understand the rules for the assessment is requiring them to agree to a set of conditions before they proceed with the assessment. This process can be described as signing up to an ‘honesty agreement’. The conditions should include: • The expectation that the candidate will take the assessment honestly and without the help of others

- The expectation that the candidate will respect the confidentiality of the assessment and not share its contents with others
- The expectation that the candidate will respect copyright, trademarks and other legal rights over the content of the assessment
- That the candidate accepts that verification will be undertaken and that they may be required to sit a verification assessment at some point in the process

Also, explain to the candidate where he or she can get further information on the assessments and on where he or she can try out practice assessments. Tell the candidate what to do and who to notify if they experience problems undertaking the assessment, and what the process to follow is should they feel they should be granted an accommodation.

SHL offers a link to practice tests online (www.shldirect.com) so that candidates can familiarise themselves with what is required and prepare accordingly. The use of practice tests is particularly recommended for candidates who are likely to be unfamiliar with formal testing procedures. The offer of a practice test session is highly desirable, and may be a particularly useful way of supporting internal candidates applying for promotion.

Additionally, it is often useful to engage in practice tests at the beginning of a session – in fact most assessment have three or four practice items at the beginning of an assessment built into them. Practice items before an assessment can reduce the bias that may arise from differential 'test sophistication', helping some people but not others. Completing practice items can also reduce nervousness by allowing a candidate to gain confidence in their ability to answer the test questions.

In instances of supervised assessment, in addition to information around the assessment and the skills being tested, it is important to highlight any logistics surrounding the testing session so that candidates have adequate time to make travel arrangements and so that they know what to expect when they arrive.

This helps to reduce anxiety and allows the candidate to prepare constructively for the session. It also allows candidates with disabilities to recognise if they will need any special arrangements.

During the Test Session

In an unsupervised but controlled mode of administration, the onus is on the candidate to ensure that they set up a quiet area where they will not be disturbed for the duration of the assessments. As mentioned above, these instructions should be highlighted to a candidate in various communications prior to the testing session. An example of these instructions may be to remind candidates to turn off mobile phones or other devices that might interrupt them and to avoid working in a busy or noisy environment.

Of further importance is that candidates are provided with options to leave the process and return later. Before the test begins, candidates should have the option to delay starting the real test until they are in a comfortable environment.

Candidates should also be made aware when an assessment allows for use of a calculator and paper for rough calculations, and they should be reminded of this allowance prior to the initiation of the assessment.

Finally, it is vital that candidates have access to resources for assistance should something go wrong. They need to be made aware of what to do if there is a technology problem or what to do if for example a disability may prevent successful completion of the test.

Maintaining control of a testing session in a supervised context may be more straightforward as control resides with the test user, but entails more from the administrator.

The administration instructions are extremely important and must always be strictly followed. Only suitably qualified persons should administer tests. Abuse of procedures described in the test manual can lead to bias and possible unlawful discrimination.

Make sure all candidates know why and how test scores are going to be used and who will have access to the results. Test administrators should promote a serious but sympathetic atmosphere. It is important to remember that the testing session will be an extremely important event for the candidate, even if it is a routine one for the administrator. Instructions should be clear and not rushed. Administrators should ensure candidates know what they have to do before each test begins. Special care should be taken with people whose first language is not the language of the assessment or the assessor, to ensure that they have understood the

administration instructions properly. For instance, some tests which are fair for speakers of English as a primary language may present problems for people with a lesser command of English.

Tests requiring reading skills, when these are not an integral part of the job, are particularly likely to be unfair. Where possible, such candidates should be tested in their native language and given an additional test of their command of the business language, if necessary.

An encouraging attitude on the part of the test administrator is always desirable, but it is particularly important to establish rapport with individuals who might lack confidence or who feel anxious about testing.

There should be an opportunity for candidates to ask general questions before the formal testing procedure starts. The procedure should not be rushed so that candidates are not discouraged from asking questions if they have any concerns or do not understand exactly what they have to do. The administrator should also check that no candidate is experiencing difficulty with a test due to a disability, e.g. holding the booklet unusually close to read.

After the Test Session

Arrangements should be made to provide candidates with feedback on their results as soon after testing as possible. Personality and motivation questionnaire feedback is critical and will often enhance the interpreter's own understanding of test results.

Feedback does not need to be lengthy, indeed with a large number of applicants (which is increasingly more likely when organisations are making use of online unsupervised assessments) this might be very time consuming. A face to face interview is preferred, but computer generated reports supported by telephone feedback may be appropriate in some circumstances.

Feedback should be given by qualified users and should be accurate and open. Profile charts may be shown to respondents, but they should not be given copies to take away. A short narrative summary may be provided if desired. This is particularly useful where testing is for counselling and development purposes.

Throughout the process, testers should be honest and open with candidates about why the tests are being used and what will happen to the results. Candidates should always be offered feedback of their results.

Scoring in a Standardised Manner and Interpreting Assessment Results

Tests should always be interpreted by properly trained individuals in the context of clearly defined criteria. Both quantitative and qualitative interpretation can be used. The former should be restricted to cases where there is sufficient supporting evidence. Interpretation of results should be accurate and not exceed the limits of the information. Users should beware, for instance, of failing to take account of the standard error of measurement in interpreting the differences between any two scores, or of drawing far-fetched inferences from personality data.

Choice of Norm Groups

An individual's test score is generally evaluated in relation to the performance of a comparison or 'norm' group. The norm group should be as representative of the applicant group as possible.

In large-scale testing, the previous applicant pool can form a suitable comparison group. Where numbers are too small, an appropriate norm group should be chosen from the test manual. This should be, as far as possible, a group applying for a similar type of position with a comparable composition with respect to educational background, work experience, age, ethnic and gender mix.

Where a perfect fit is not available, the norm group chosen should have similar average scores to those of the applicant group as a whole.

Selection Strategies

There are two basic approaches to selection on the basis of the test scores of a group of applicants.

The 'top-down' approach selects people from the highest scorer downwards until sufficient people have been selected. This approach maximises the benefit that can be gained from using the test, its 'utility', and is likely to produce the highest calibre work force. Where the selection ratio is low, that is where there are many applicants for few jobs, this method quickly reduces the number of candidates to be considered.

However, it tends to increase adverse impact where there are score differences between groups, which is this method selects fewer applicants from the lower scoring group. It also reduces the opportunity to select on other skills and competencies.

The 'cut-off' approach selects people who score above a designated cut-off. The cut-off score is generally chosen to represent the ability level consistent with a reasonable chance of successful performance on the job; it clearly depends both on the level of the test being used and the skill requirements of the job.

This method is particularly appropriate where there is a high selection ratio (few applicants for many jobs) and the test is used to screen out applicants who are not up to standard. The lower the cut-off is set, the lower the utility of the test for the organisation, but the less disparate impact the test will have if there are score differences between ethnic or gender groups.

The choice between these approaches and the actual cut-off score used, depends, in each individual case, on:

- The selection ratio
- When the test is used in the selection process
- The level of the job and / or training
- The calibre of the applicant pool
- Whether there is any question of disparate impact

The greater the demands of the job and the less training available, the higher cut-off scores would need to be. Where less skills are required and more training is available, or where group score differences have been observed, lower cut-off scores are more appropriate.

Often a more qualitative strategy is used, whereby the individual's relative performance on the test compared to the norm is evaluated together with other information.

Measures such as personality questionnaires, practical exercises and interviews, as well as details about previous work experience and biodata, should provide additional information about a candidate's strengths and weaknesses.

Additional Resources

Psychological Organisations

International Test Commission - www.intestcom.org

European Federation of Psychologists Associations - www.efpa.eu

European Association of Work and Organizational Psychology - www.eawop.org

Psychology Organisations by Country

Country	Psychology Organisation	Website
Australia	Australian Psychological Society	www.psychology.org.au
Belgium	Belgian Association for Psychological Science	www.baps.be
Canada	Canadian Psychological Association The Canadian Society for Industrial and Organizational Psychology	www.cpa.ca www.csiop-scpio.ca
China	Chinese Psychological Association	www.cpsbeijing.org
Czech Republic	Czech-Moravian Psychological Society / Českomoravská Psychologická Společnost	www.cmps.ecn.cz
Denmark	Danish Psychological Association	www.dp.dk
Finland	Finnish Psychological Society / Suomen Psykologinen Seura	www.psykologia.fi
France	Federation of French Psychologists	www.psychologues-psychologie.net
Germany	Federation of German Psychologists' Associations / Föderation Deutscher Psychologenvereinigungen	www.psychologie.de
Hungary	Hungarian Psychological Association / Magyar Pszichológiai Társaság	www.mpt.hu
India	National Academy of Psychology	www.naopindia.org
Indonesia	Indonesian Psychology Association / Himpunan Psikologi Indonesia	www.himpsi.or.id
Italy	Italian Network of Psychologists Associations / Associazione Unitaria Psicologi Italiani	www.inpa-europsy.it
Japan	Japanese Psychological Association	www.psych.or.jp
Netherlands	Netherlands Institute of Psychologists / Nederlands Instituut van Psychologen (NIP)	www.psynip.nl
New Zealand	New Zealand Psychological Society	www.psychology.org.nz
Norway	Norwegian Psychological Association	www.psykologforeningen.no
Philippines	Psychological Association of the Philippines	www.pap.org.ph

Poland	Polish Psychological Association / Polskie Towarzystwo Psychologiczne Polish Association of Organizational Psychology / Polskie Stowarzyszenie Psychologii Organizacji	www.ptp.org.pl www.pspo.org
Portugal	Order of Portuguese Psychologists/Ordem dos Psicólogos Portugueses	www.ordemdospsicologos.pt
Romania	Asociatia Psihologilor din Romania / Romanian Psychologists' Association Romanian Association for Industrial Organizational Psychology/Asociatia de Psihohogie Industriala si Organizationala	www.apsipro.ro www.apio.ro
Russia	Russian Psychological Society	www.psyrus.ru
Singapore	Singapore Psychological Society	www.singaporepsychologicalsociety.org
South Africa	Health Professions Council of South Africa Society for Industrial Organizational Psychology in South Africa	www.hpcsa.co.za www.siopsa.org.za
Spain	Spanish Psychological Association / Colegios Oficiales de Psicologos	www.cop.es
Sweden	Swedish Psychological Association / Sveriges Psykologförbund	www.psykologforbundet.se
Switzerland	Federation of Swiss Psychologists / Föderation der Schweizer Psychologinnen und Psychologen / la Fédération Suisse des Psychologues / Federazione Svizzera delle Psicologhe e degli Psicologi	www.psychologie.ch
Turkey	Turkish Psychological Association / Türk Psikologlar Dernegi	www.psikolog.org.tr
UAE	Emirates Psychological Association	
United Kingdom	The British Psychological Society	www.bps.org.uk
United States	American Psychological Society Society of Industrial and Organizational Psychology	www.apa.org www.siop.org

Psychometric Textbooks

An Introduction to Psychological Assessment and Psychometrics - Keith Coaley

Using Psychometrics: A Practical Guide to Testing and Assessment - Robert Edenborough

Psychometrics: An Introduction - R. Michael Furr and Verne R. Bacharach

Psychological Testing: An Introduction - George Domino and Marla L. Domino

References

Brannick, M, T, & Levine, E, L., (2002). *Job analysis: Methods, research and applications for human resource management in the new millennium*.

British Psychological Society (2007). *Psychological Testing: A User's Guide*.

Dipboye, R, L., (2016). *Exploring Industrial and Organizational Psychology: Work and Organizational Behavior*. Sattvic Publishing.

Domino, G, & Domino, M, L., (2006). *Psychological testing: An introduction* (2nd Ed.,). Cambridge University Press.

Robertson, I, A., (2001). Personnel Selection. *Journal of Occupational and Organisational Psychology*, 441-472.

Urbina, S., (2004). *Essentials of Psychological Assessment*. John Wiley & Sons, Inc.

This practice guide tries to answer some common questions people have in using tests fairly. However, it is difficult to cover all the issues in brief. If you would like more information or you have any queries on this topic, please visit the SHL website or contact our Global Customer Service Centre.

Whilst SHL has made every effort to ensure that these guidelines reflect best practice, SHL does not accept liability for any loss of whatsoever nature suffered by any party as a result of its reliance on these guidelines.