som
Supporting occupational health and wellbeing professionals

# Psychometric Assessments in Occupational Health

## What they are and how to use them

January 2025

som.org.uk

ISBN 978-1-0369-1445-5

# Authors

This guide was produced by members of the SOM Occupational, Health and Work Psychology (OHWP) Special Interest Group:

**Dr Kevin Teoh, Professor Almuth McDowall, Professor Gail Kinman, Professor Nancy Doyle, Professor Richard Kwiatkowski** and **Dr Rainer Kurz.**

The authors are grateful to the following individuals for their support in reviewing the guide:

**Anna Harrington, Professor Anne Harriss, Dr Anil Adisesh, Dr Debbie Rees-Adams, Professor Ewan MacDonald, Professor Jo Yarker, Dr John Gration, Professor Neil Greenberg, Nick Pahl and Dr Steve Malleson.**

This guide supports occupational health practitioners and associated professionals in using psychometric assessments in their practice. The SOM journal Occupational Medicine has published a wide range of questionnaire reviews of relevance to people working in OH and associated fields. See https://academic.oup.com/occmed/pages/questionnaires

# Summary

## Why use psychometrics in occupational health?

Psychometric assessments provide an objective and standardised measure of individual differences, including personality traits, behavioural tendencies and cognitive abilities. In occupational health, psychometrics can be used in the assessment of physical and mental health outcomes and symptoms, risk factors and working conditions.

The key aspects of psychometric assessments are that they are:

1. **Reliable:** producing consistent and accurate results across different situations and contexts
2. **Valid:** accurately measuring what they say they are measuring
3. **Minimise bias:** as much as possible
4. **Standardised:** administered and scored in a consistent way.

This guide provides examples of psychometric assessments across the following areas:

- Physical health assessments (e.g. musculoskeletal, pain)
- Mental health assessments (e.g. symptoms of depression, anxiety)
- Health behaviour assessments (e.g. fatigue, alcohol use)
- Work-related stress/psychosocial risk assessments (e.g. the HSE Management Standards)
- Neurodiversity (i.e. the need for specialised assessments)
- Developmental and cognitive functions (e.g. attention, memory)
- General functioning (e.g. work ability)

## Tips for using psychometrics

Here are some key tips for using psychometric assessments effectively and fairly:

- **Choose the right tool:** Ensure it is well aligned with your aims and check for reliability and validity.
- **Clarify the purpose:** Clearly define the objective of the assessment and avoid placing too much emphasis on the results alone.
- **Minimise bias:** Ensure fairness by considering cultural differences and administering the assessment in a consistent, standardised way.
- **Give clear instructions:** Explain the purpose of the assessment and how the results will be used.
- **Interpret carefully:** Consider the context and avoid overgeneralising from the results.
- **Ensure confidentiality:** Protect privacy and obtain informed consent from participants before conducting the assessment.
- **Stay updated:** Use up-to-date tools and stay informed about best practices.

Using unreliable and invalid psychometrics can pose significant risks, such as inaccurate results and harm to individuals or organisations. You may find it useful to consult a relevant registered professional (e.g. a psychologist or physician) for expert guidance. Correct use of psychometrics can ensure the appropriate information is collected to support decision-making and problem identification.

There are things you can do to help improve decision-making, understanding and practical implementation of psychometric assessments. Consider the following actions:

- **Be clear about what the assessment is being used for,** and what 'good' looks like. A thorough analysis of the context ensures the process is well-grounded, useful and defensible. Assessments for a pre-recruitment process, essential annual medical check-ups, and clinical settings are very different situations. For example, there is little evidence supporting mental health screening before, during or after employment.
- **Consider the available evidence** surrounding the validity and reliability of the measure to ensure it is accurate and dependable.
- Work with people who are on the **Register of Qualifications in Test Use** (RQTU) or who are familiar with the tests.
- **Review The British Psychological Society's Registered Test Database.** This helps users find appropriate tests by providing detailed reviews and registration information.
- **Question developers and test publishers.** When commissioning or purchasing a psychometric test, ask providers how and why the assessment is suitable for your client group and organisation and the intended purpose.
- **Adhere to assessment guidance.** Follow the guidelines to ensure the assessment is administered and scored consistently, maintaining standardisation.
- **Evaluate the process.** Assess whether the test achieved its goals, considering not only user and administrator feedback but also whether it led to the intended improvements, such as enhanced health outcomes.

### Training and competence

The British Psychological Society offers a Qualification in Test Use. However, the required training and competence levels vary for different measures. It is therefore important to recognise your professional responsibilities and work within the limits of your knowledge and skills.

Each psychometric measure should provide guidance on its use, administration process and scoring procedure. It should also specify the competence and qualifications needed for the occupational health practitioner.

## Determining the quality and relevance of a psychometric assessment

Two key aspects to evaluate when assessing the quality and relevance of a psychometric measure are its validity and reliability.

- **Reliability** is the foundation of any psychometric assessment and refers to its consistency. This includes how reliable a measure is over time, how it performs across different administrators, and the consistency of the items within it. Assessments should provide measures of internal consistency and test-retest reliability.
- **Validity** broadly questions whether the test measures what it claims to measure. This can be assessed in various ways: does the test correlate with other relevant factors? Does the test appear credible? Does the test accurately measure what it says it does? What evidence supports these claims?

## Bias

Factors such as honesty, stigma, fear, privacy or desire for a particular outcome can affect how a person may respond to a test. Instructions and tests may be more easily understood by people with specific educational backgrounds or language skills. To promote fairness and equity, it is important everyone has adequate time and opportunities to practise.

Seeking specialist advice on this issue is highly recommended, as legal requirements also need to be considered. For example, the AGCAS guide for psychometric tests for candidates with disabilities provides additional practical considerations when assessing employees with disabilities.

# Psychometric Assessments in Occupational Health: What they are and how to use them

## Why use psychometric assessments?

This guidance is written to support occupational health practitioners in commissioning, understanding and interpreting psychometric assessments in their practice.

Using a valid and reputable assessment means the results will be accurate and helpful for the practitioner and those they are working with. Robust psychometric assessments save time, can be administered to more than one person at once, and can effectively compare the changes within an individual against expected outcomes.

Anyone can develop and sell a psychometric assessment in the United Kingdom. Therefore, it is crucial to understand the difference between a reliable, valid, evidence-informed assessment and one that is unlikely to provide accurate results.

## What are psychometric assessments?

Psychometric assessments – also known as tests, measures, tools, assessments, instruments or questionnaires – provide an objective and standardised way to measure individual differences, for example internal characteristics such as preference or external behaviours.

Psychometric assessments should be rigorously developed according to best practice standards and administered, scored and interpreted in a standardised way. This ensures the results are reliable and valid.

Psychometric assessments are different from simple self-assessments, which lack scientific rigour; for example, a newspaper quiz that asks about mood, 'personality' or preferences.

Psychometrics have four defining features:

1. **They must be reliable.** This means they should provide consistent and accurate results across different situations and contexts. If something is not reliable, it is impossible to know what is being measured. Imagine a tape measure made of elastic – the reading could not be trusted.
2. **They must be valid,** so they must measure what they claim to measure. The content of the psychometric assessment must align with the specific characteristic being evaluated. Imagine measuring personality and then making conclusions about someone's running speed – that would clearly be absurd. Validity also depends on reliability; if an instrument is inaccurate – such as diagnosing a person with dyslexia one day and not the next – it will lead to confusion.
3. **They must be as free from bias as possible.** A good assessment should provide accurate results for everyone, regardless of age, gender identity or background. While this can be challenging due to differences in experiences, languages and cultural backgrounds, well-developed tests aim to minimise bias. Some tests, such as the Short Form Health Survey (SF-36), which measures eight different aspects of health (e.g. physical functioning, mental health, pain and emotions), address this issue by using different norm tables based on characteristics such as sex, age and occupation.
4. **Psychometrics should be standardised.** This means there must be a clear protocol for administering the assessment so that everyone has a fair chance and a consistent experience. Standardisation also involves establishing a clear benchmark for comparison, either by using a reference group of individuals who have taken the assessment, or by setting a common threshold all participants should meet.

## Four things to know about psychometric assessments

1. Any measurement of human characteristics will never be 100% accurate and should be viewed as a 'best estimation'. One advantage of psychometric assessments is that they calculate and account for margin of error. This contrasts with methods such as interviews, where the accuracy and bias of the assessor's judgements are usually unknown – despite most assessors believing themselves to be accurate and unbiased!
2. It is essential to select a good assessment tool – one that is both reliable and valid. To do this, individuals must either be appropriately trained or collaborate with someone who is trained and has access to that test (most good tests restrict access).
3. It is highly advisable to incorporate psychometrics into a process as one piece of the jigsaw only, rather than relying on any single assessment to form a final opinion on any particular matter. Since individuals possess a wide range of characteristics, some are better measured through methods such as direct questioning or observation. This approach provides a more comprehensive understanding of the situation.
4. What is being assessed should be clear at the start. For example, conducting a job analysis (i.e. a thorough review of the job, its tasks and responsibilities) will help determine the tools most appropriate for a workplace health assessment to assess the capability and fitness for undertaking a particular role. This makes the process more context-specific and defensible.[1]

## Different types of psychometric assessments

Psychometric assessments can be broadly categorised into the following types:

- **Cognitive tests,** which measure an individual's ability and capacity to learn, reason and process information.
- **Health states and behaviour assessments,** which can be used as an aid for clinical assessments, for making adjustments to the working environment or processes, or for evaluating an individual's level of functioning.
- **Personality questionnaires**, which assess an individual's typical behaviour, emotions, preferences, patterns or work styles. Contemporary personality questionnaires usually cover the 'Big 5' personality factors: Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism.
- **Motivational and/or personal values surveys**, which measure an individual's drive, energy and engagement and/or their enduring beliefs.
- **Situational judgement and simulation tests,** which evaluate an individual's judgement, practical knowledge and responses to specific scenarios at work.

Psychometric measures can be used in several areas, with examples provided below. It is important to remember that mentioning these examples **does not constitute an endorsement of the measure,** as other measures may be better suited to a particular context.

As with all psychometric assessments, it is crucial that the assessments used are both valid and reliable, relevant to the assessment context, and administered by an appropriately qualified practitioner. Diagnosing specific health conditions requires additional expertise and the use of more than one measure, alongside other possible clinical tests. Furthermore, the purpose of a particular measure may be limited; for example, it might indicate a potential issue but not identify the underlying cause or associated risk factors.

1. https://www.peoplemanagement.co.uk/article/1745644/ensure-recruitment-process-legally-defensible

## Physical health assessments

Occupational health practitioners may use measures to assess for risk for musculoskeletal issues (e.g. the Musculoskeletal Online Assessment Tool or the Assessment of Repetitive Tasks tool), symptoms related to musculoskeletal issues (e.g. Office Work Screen or the Versus Arthritis Musculoskeletal Health Questionnaire (MSK-HQ), and specific aspects of managing musculoskeletal issues (such as the Workstyle measure for upper limb pain and functional limitations).

Occupational health may also use psychometric tools to evaluate other aspects of physiological functioning. For example, the MRC Dyspnoea Scale assesses breathlessness during daily activities, which can be helpful for assessing workers with respiratory issues. Tools for measuring noise-induced hearing loss and occupational dermatology are also available. Additionally, assessing pain levels may be needed as part of return-to-work assessments or of routine surveillance of workers with certain conditions (e.g. post-accident recovery, cancer or rheumatoid arthritis). Measures for pain assessment include the Brief Pain Inventory (BPI) and the Pain Disability Index (PDI).

## Mental health assessments

With the growing emphasis on supporting mental health in the workplace, it is essential first to determine which aspect of mental health is being assessed. This guide does not cover the use of psychometric assessments for mental health screening before, during or after employment. For depression, commonly used tools among occupational health practitioners are the Patient Health Questionnaire (PHQ-9 or PHQ-4) and the Beck Depression Inventory (BDI). For anxiety, the Generalised Anxiety Disorder (GAD-7) is frequently used. For both depression and anxiety, measures focus on evaluating the presentation of symptoms over time to gauge the extent of ill health.

Burnout is now officially recognised as a factor influencing health status in an occupational setting (but not a mental health disorder) by the International Classification of Disease (ICD-11). It can be assessed with the Burnout Assessment Tool (BAT) or the Maslach Burnout Inventory (MBI), which have good psychometric properties. In some cases, a measure of general psychological distress may be needed. The Kessler Psychological Distress Scale (K10) and the General Health Questionnaire (GHQ) have been used to identify individuals who are likely to have or be at risk of developing psychiatric disorders. Depending on the context, other more specific measures of mental health (e.g. trauma) may be needed.

Linked to mental health is the concept of wellbeing. The World Health Organization-Five Well-Being Index (WHO-5) and the Warwick-Edinburgh Mental Wellbeing Scales (WEMWBS) both provide concise and robust measures of subjective wellbeing. However, despite the popularity of wellbeing as a concept – since there are many ways to define and measure it (along with related concepts such as happiness and engagement) – it is crucial to ensure the correct measure is selected for the purpose.

## Health behaviour assessments

Assessing health and lifestyle behaviours is a common aspect of many occupational health roles. This may be as part of general health assessments or specifically related to performance in safety-critical roles. For example, the Modified Fatigue Impact Scale (MFIS-5) and the Fatigue Assessment Scale (FAS) can evaluate the impact of work or health conditions (e.g. Long Covid, fibromyalgia or mental health) on fatigue levels. The Insomnia Severity Index (ISI) assesses for insomnia and sleep problems, while the Sleep Quality Scale (SQS) is commonly used to assess problems with initiating and maintaining sleep, difficulty waking and sleep satisfaction.

More generally, the General Practice Physical Activity Questionnaire (GPPAQ) assesses physical activity, while alcohol use and abuse is often evaluated via the Alcohol Use Disorders Identification Test (AUDIT) or the CAGE questionnaire. These assessments provide valuable information to help identify potential behaviour and lifestyle changes, determine additional support needs, and make referrals to specialist treatments or support services.

### Work-related stress/Psychosocial risk assessments

Substantial evidence links the design, organisation and management of work (i.e. the psychosocial environment) to worker health, particularly concerning mental and musculoskeletal health outcomes. The HSE Management Standards Stress Indicator Tool offers a valuable and widely used assessment of six key aspects of the working environment: demands, control, support, change, role and relationships. Other tools, such as the Copenhagen Psychosocial Questionnaire (COPSOQ), offer a broader range of work factors for consideration. These assessments help identify aspects of the working environment where workers may need support or adjustments, or where a better-suited environment may be necessary. Importantly, at the organisational level, these tools can form the basis of a risk assessment, to guide the necessary changes for creating a healthier workplace.

### Neurodiversity

The assessment of neurodivergent conditions, such as autism and attention deficit hyperactivity disorder (ADHD), is a challenging and complex area due to the limitations of many screening tools for providing differential diagnoses. This can lead to inappropriate diagnoses that are difficult to retract. Any assessments require specialised training and qualifications.

An alternative approach may involve conducting a thorough assessment of an individual's background history, which can help identify any developmental or cognitive issues (e.g. memory or attention). This assessment can then guide a referral to a specialist for further testing. Evaluating background history provides the opportunity to identify potential support and adjustment needs in the context of current performance, role demands and environmental factors.

### Developmental and cognitive functions

Occupational health practitioners may need to determine a worker's fitness for work or monitor their developmental and cognitive functioning if there are concerns. This might be necessary after a traumatic brain injury, or a stroke, possible cognitive decline in older workers, or for assessing neurodiversity functioning. The results of these assessments could lead to referrals to relevant specialists or inform workplace adjustments. Also known as ability tests, such measures can be used in coaching contexts to support individuals' development.

The Wechsler Adult Intelligence Scale, the Montreal Cognitive Assessment (MoCA), the Differential Aptitude Tests (DAT), Raven's Standard Progressive Matrices (SPM) and the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) collectively evaluate multiple cognitive functions (e.g. memory, attention, verbal fluency and visuospatial ability). Other assessments focus on specific areas, such as general cognition (e.g. the Mini Mental State Examination (MMSE)) or dementia (e.g. the Mini-Cog, Addenbrooke's Cognitive Examination-Revised (ACE-R)).

### General functioning

Several assessments enable occupational health practitioners to evaluate the physical functioning of workers, such as the Work Disability Functional Assessment Battery (WD-FAB) and the EQ-5D-5L. Additionally, the Work Ability Index (WAI) assesses a worker's current ability in relation to their overall health status, job demands and available resources compared to their lifetime capabilities.

### Training and competence

The British Psychological Society offers a Qualification in Test Use for professionals who use psychometric or psychological tests as part of their role. This qualification covers the fundamentals of psychometric testing and scoring, with registered members agreeing to adhere to the Code of Good Practice for Psychological Testing and to maintain their competence in testing.

Each psychometric measure should provide guidance on its use and administration process, as well as the scoring procedure. It should also specify the level of competence and qualification needed for the practitioner. There is, however, substantial variation in the level of training required for different measures, such as:

- Being on the Register of Qualifications in Test Use (RQTU), for example being trained in several instruments)
- Being a relevant practitioner or chartered psychologist (e.g. the Wechsler Adult Intelligence Scale) or clinically qualified (e.g. the State–Trait Anxiety Inventory (STAI) Self Evaluation Questionnaire)
- A relevant educational qualification (e.g. the Beck Depression Inventory (BDI), the Chronic Pain Coping Inventory (CPCI))
- Undergoing specialised training in the measure being used (e.g. Mini-Cog).

Certain measures have been designed to be used without any additional formal training (e.g. the PHQ-9, GAD-7 and GPPAQ). However, it is important to adhere to the administration instructions, recognise the professional responsibility to work within the limits of one's knowledge and skills, and keep in mind the limitations of self-report measures.

## Case study 1: psychometrics in coaching

Evan had been referred to workplace coaching because co-workers had complained that Evan was often rude and erratic at work. The coach, a qualified occupational psychologist, conducted a thorough assessment of personality and cognitive ability as requested by the commissioning client, a university hospital. The psychologist took a lot of time on this assessment, and asked a series of open questions about how Evan did their best work and what they found challenging. This process elicited that Evan found certain aspects of their current job challenging, including tasks which require independent self-organisation or having to work with spreadsheets.

The psychometric assessments corroborated Evan's preference not to work with detail and their difficulty in working with numbers, particularly when under stress. The psychologist wrote Evan a personal report and a shorter (agreed) report to be shared with their line manager. Next, the coach worked with Evan to develop better time management and self-organisation strategies, as Evan said a key goal was to get ahead of their workload so that they had more time to digest information which they found complex.

Evan developed their own self-management system as a result, using simple project management software and to-do lists. Having become more aware of their preferences, Evan then negotiated with their line manager to delegate some of the tasks working with spreadsheets to a colleague who excelled at this, which then left them more time for clinical duties – and less stressed. This reshaping of their work role has helped everyone. Evan is performing well in their role and is no longer seen as rude.

# How to evaluate psychometrics – including questions to ask

To effectively assess the quality of a psychometric assessment, it is important to understand its reliability and validity, as well as how to evaluate these aspects. It should be noted that there is no one single method for determining this. This section can help in evaluating psychometrics to provide guidance on the sorts of questions that need to be asked.

## Reliability

Reliability is the foundation of any psychometric assessment as it refers to how consistent a specific measure is. When evaluating reliability, some important factors are:

- **Is it reliable over time?** If the characteristic being measured (e.g. personality, memory) remains stable over time, an individual should obtain similar scores when measured repeatedly. For example, a standardised ruler provides a reliable measure of a person's height, as it consistently yields the same result for the same person regardless of the ruler used, the timing of the measurement, or the person doing the measuring. In contrast, using rubber bands to measure height would result in varying outcomes depending on the type of rubber bands, how tightly they are pulled, and the method of measurement.
- **Is it independent of the person administering it?** A reliable assessment should produce similar results regardless of who administers it. Different assessors using the same tool with the same individual should obtain comparable results. For example, if a psychometric tool is used to assess a specific health risk to a worker (e.g. the risk of clinical depression), different assessors should arrive at the same conclusion when using the same tool.
- **Does it have consistent items?** The items (i.e. questions) within a psychometric assessment should consistently relate to the same construct. For example, an individual who is safety conscious should score highly across most items related to safety behaviours in the workplace, while someone who is less safety conscious should score consistently lower on these same items.

## How to assess reliability

Understanding how reliability is assessed can enable practitioners to evaluate a psychometric tool or review information in the test manual. Reliability across items, known as **internal consistency**, is typically represented by a single number from a Cronbach's alpha (α) or an Omega (ω) statistical test. This value ranges from zero (0) to one (1), with a higher value indicating better consistency. A score of at least 0.70 is generally considered indicative of good internal consistency. In a measure with good internal consistency, an individual with a healthy lifestyle would score consistently high on most items related to health behaviours (e.g. exercise, fruit and vegetable intake, and alcohol intake). This information should be available in the test manual or guide.

Reliability across time is known as **test-retest reliability**. This is typically assessed through correlation analysis, which evaluates the strength of the relationship between results at two different time points. For example, a verbal reasoning test might be administered to a group of people at Time A and then again three weeks later at Time B. If the test is reliable, the results at Times A and B will be very similar, as an individual's

underlying verbal ability should not change significantly over a short period. Here, the correlation coefficient (known as r) ranges from zero (0) to one (1), with a higher value indicating a stronger relationship. In most cases, a strong correlation coefficient of above 0.70 is expected. However, if the measure involves variables that might naturally fluctuate over time (e.g. health behaviours such as exercise), a slightly lower score might still be acceptable.

## Validity

Broadly speaking, validity asks the question: are we measuring what we intend to measure? There are different ways to assess validity, and each examines different aspects of how well the measurement meets its intended purpose.

| Type of validity | What this means | Example | Why this is important |
|---|---|---|---|
| Construct validity | Does the test measure what it says it does? | Different assessments measuring anxiety levels should have very similar scores if the same person completed them. | Assesses whether a test is related to other known tests that measure the same or a similar construct so you know you are measuring what you think you are measuring. |
| Criterion-based validity | Do assessment results relate to other information, such as job performance or health data? | Is there a relationship between how stressful someone reports their working environment to be and the likelihood of them reporting low job satisfaction? Is performance on a test related to subsequent job performance? | Powerful indicator to justify use where the criterion data are available. |
| Face validity | Does the assessment make sense and appear credible to people who are taking and using it? | Some consider wearable technology (e.g. smart watches) to assess health irrelevant or inappropriate to the workplace context. | Important to gain support and buy-in from people taking the test and the organisations using them. |
| Faith validity | Do people trust the assessment and its results? | Several factors influence how much a stakeholder (e.g. a candidate worker or an organisation) trusts an assessment result, including its appearance, novelty, familiarity, experience, popularity and its marketing. These factors can impact perceptions of the assessment independent of the/any scientific evidence supporting the test. | Important to gain support and buy-in from individuals and organisations.<br><br>Likely to occur when the test is used, supported and produced by well-known organisations or individuals. |
| Consequential validity | What are the wider effects of using the assessment (intended and unintended)? What might be the impact? | Research indicates that some ethnic minorities may perform less well than majority populations on certain cognitive tests. This means that relying too heavily on these tests for selection purposes could lead to a less diverse workforce and potential unfair discrimination. | Often neglected and can be difficult to establish.<br><br>Vital for assessing issues of diversity, equality and inclusion in the workplace. |

## How do test constructors establish validity?

As highlighted above, construct validity can be established by examining how well a particular test correlates with a similar test. For example, if we are interested in measuring generalised anxiety, it is important to check how well this test relates to established anxiety tests. Here, a correlation coefficient (r) is used to quantify this relationship, with a strong correlation coefficient (ideally above 0.55, according to the test review guidelines from the European Federation of Psychologists' Associations (EFPA) guidelines[2]) being desirable. This information should ideally be included in the test manual or guide; if it is not, you can request it.

**Criterion-based validity** determines how well what is being measured is linked with relevant outcomes. For example, if a test is supposed to measure factors contributing to better wellbeing, there should be evidence supporting this. If an assessment of exercise habits is given alongside measurements of body mass index and cardiovascular fitness, the results from both the exercise habit assessment and the fitness measurements would be used to calculate a correlation coefficient (r) or a regression coefficient (B) value. According to the EFPA test review guidelines, the observed validity coefficient should be higher than 0.2. Therefore, if a test measuring exercise habits shows a correlation of 0.1 with fitness measures, this indicates a weak relationship and suggests that a different measure of exercise habits be considered.

**Face** and **faith validity** focus on the perceived validity of an assessment. This can be evaluated by gathering feedback from users, managers and organisations who have used an assessment via methods such as focus groups or questionnaires. For example, while general cognitive ability tests such as verbal, spatial and numerical ability are good predictors of work performance, individuals may question their relevance, asking: "Why am I being asked to solve online puzzles under time pressure when this is nothing like what I do in the real world?" Balancing the organisation's need for predictive accuracy with individuals' perceptions of the test's relevance is crucial. Candidates may decline job offers if they view the assessment as irrelevant and unfair. Case studies and marketing materials can enhance the face or faith validity of an assessment, but it is important to remember that perceived validity does not guarantee actual validity. Therefore, the criteria for assessing reliability and validity shown above should still be established. Additionally, it should be recognised that for particular groups these tests may not be highly predictive and should be used with caution and that they require professional judgement and understanding of diversity.

Establishing **consequential validity** can be challenging, as it requires evaluating the broader impact of an assessment by considering feedback from individuals and across the organisation (or even beyond). It could also entail analysing employee surveys or staff statistics related to working experience, diversity and inclusion, and observing any changes in these outcomes due to the use of a particular test. It is important to consider external implications, such as whether the use of a particular test influences perceptions of a specific issue (e.g. fairness, performance or the organisation's reputation).

## Standardisation

Any psychometric assessment should be administered and evaluated in a standardised manner. If the test manual specifies a time limit of 20 minutes, then this should be strictly adhered to, unless the publisher advises an extension for specific individuals or groups (for example, people with particular neurodiverse conditions). Additionally, if there are standard instructions, preparation activities, or practise questions, they must be provided to all participants in the same way.

Standardisation also involves establishing a clear point of comparison. It is meaningless to learn that someone scores '20' on a test without understanding how this score compares with other people. This is why most tests use standardised scores, which require practise and training to calculate and interpret.

2. EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests, version 4.2.6. (2013).

## Case study 2: online health platforms

The HR director of a large manufacturing company returned from a conference where they were presented with a new platform that can measure health behaviours of employees (e.g. alcohol consumption and exercise) using an app and a website. This platform is growing in popularity, with several rival companies currently using it. It is easy to use and can provide the organisation with live updates on how employees are feeling.

Further exploration of the product showed that the platform, or its underlying principles, had not been evaluated in the peer-reviewed scientific literature. The platform developer emphasised their long-standing experience and expertise in the field of wellbeing but was unable to provide data around the validity and reliability of the platform. After reviewing alternative providers, the organisation decided to commission the service of another platform, which was grounded in a relevant theoretical and research base with published reports evaluating the instrument being used.

Returning to the example of exercise habits, comparing scores among a diverse group of neighbours – some running marathons, others gardening, a few swimming, and some doing no exercise at all – will provide different insights to comparing scores within a local netball team, where all members regularly engage in exercise. Therefore, it is crucial to define the point of comparison. In some cases, this 'norm referenced' approach may not be suitable – for example, if the goal is to determine if an individual can lift a weight of 20 kg comfortably, rather than lifting a weight sufficient to reach the Olympic final, a criterion-referenced approach would be more appropriate.

### Bias

Assessments can be biased for various reasons. As an example, individuals with certain educational backgrounds or language skills might better understand instructions and test items. Items themselves can be biased. Measures of depression and anxiety might include items like "I get a sort of frightened feeling, like 'butterflies' in the stomach" and "I feel blue", which may be interpreted differently across cultural and language groups. This can lead to significant variations in scores and potential bias.

It is important to ensure a test minimises adverse impact, where one group of people systematically scores lower than another group. An effective approach is to give everyone equal time and opportunity to practise. It is advisable to seek specialist advice on this issue, as there are legal requirements regarding unfair discrimination. For further guidance, the AGCAS guide for psychometric tests for candidates with disabilities provides practical considerations for assessing employees with disabilities.

Understanding the context of why the assessment is being undertaken is important to address how a person may respond to a test. Reasons include honesty (e.g. "I want the job", "I don't want to be assigned to that role"); stigma and fears (e.g. "I want to ask for help, but I believe that if I respond in a certain way, then I will be treated in an unfair manner"); privacy (e.g. "I don't want my employer to know about my problems"); or desire for a particular outcome (e.g. "I want this employment tribunal claim to be successful"). Helping people understand why they are completing a particular test, how the information will be used, and how it is administered can influence the extent to which people feel safe or open to disclose certain information. It is also important that any report, whether written for an employer, legal purposes or otherwise, clearly states the limitations of the tests used and is upfront about the possible explanations for the results found.

### Where to find reliability, validity and standardisation scores

Statistics for a particular test:

- Should be available from the test publisher – in the technical manual, in publicity material or websites relating to the test, or upon request for more specialist information (such as specific norm groups).
- Might be available in published research studies that have used the assessment.
- Reports or reviews into a particular test (e.g. The British Psychological Society's Psychological Testing Centre database of Registered Tests – see *Occupational Medicine*'s Questionnaire Review section below.

### The risks of using unreliable or invalid psychometrics

Using unreliable or invalid psychometrics means measuring something other than what is intended, which leads to inaccurate information for decision-making. Risks include:

**a)** Misunderstanding individuals or inaccurately estimating health risk, which, in turn, will influence how they are supported or treated.

**b)** Using inaccurate information to inform a coaching or other development process and having unintended consequences, such as denting people's self-confidence, failing to identify and address actual development needs, or disrupting workplace dynamics.

**c)** Failing to select individuals who are well suited to a role or, conversely, selecting people who are not appropriate for the role.

**d)** Making an assessment process illegal, for example by discriminating against certain groups of individuals

**e)** Spending resources that could be better used in other ways.

**f)** Producing a report that is simply inaccurate or untrue.

**g)** Labelling someone with characteristics that are not psychologically valid, e.g. "They scored highly on the laziness quotient."

**h)** Excluding a group of people based on spurious results (particularly worrying if this group disproportionately includes individuals with protected characteristics) can lead to unfair discrimination and potentially violate legal and ethical standards.

### Tips for using psychometrics

Using psychometric assessments requires people to have relevant knowledge. You may find it useful to consult a registered professional (e.g. a psychologist or a physician) who has an appropriate understanding of psychometrics.

Several measures can improve decision-making, understanding and practical implementation of psychometric assessments. While these steps do not guarantee quality, they should be considered alongside the tips outlined at the start of this guide to provide additional checks to bolster assurance.

- **Clarify the purpose of the assessment.** Ensure you understand what the assessment is intended to measure and how it will be used. Conducting a thorough job analysis can provide context and make the process more defensible. Useful questions to consider include:

    **a)** What is the purpose of this assessment? What are the limitations of the tests used, and how might the concerns from participants affect how they respond?

b) How confident am I about what I want to assess? For example, have I clearly identified the key characteristics needed for a particular job or role?

c) Do I have access to trained professionals in psychometrics or know where to find expert advice?

d) How can I effectively align the qualities required for a job or role with the psychometric assessments?

e) What is the budget for this assessment process?

f) What is the intended use of the psychometrics? Is it for job selection, development, health assessment or another purpose?

g) Do I understand the potential risks and benefits associated with using these psychometrics?

- **Consider the available evidence**

a) Psychometric tests featured in academic studies are generally more transparent regarding their validity and reliability. Academic journals typically use a peer review process, ensuring published articles have been reviewed and approved by subject matter experts and editors. Relevant data may also be available through technical reports or from the test publisher.

b) Access to academic databases can be challenging and costly. However, using Google Scholar to search for a specific test can provide a quick and inexpensive overview of relevant literature, including free access to brief summaries (i.e. abstracts).

c) Consult test reviews published by reputable organisations, such as The British Psychological Society (BPS). The BPS is the representative body for psychologists and psychology in the United Kingdom and provides authoritative reviews and insights into psychometric tests.

- **Work with people who are on the Register of Qualifications in Test Use (RQTU)**

a) The RQTU is the official record of all test users who have been awarded qualifications in occupational, educational and forensic test use by The British Psychological Society. Register members agree to abide by the Code of Good Practice for Psychological Testing and to maintain their competence in testing.

b) The RQTU is a searchable listing that appears in the public domain and can be used to check an individual's credentials.

- **Review The British Psychological Society** Registered Test Database**.**

This test review and registration process helps users identify an appropriate test suitable for their needs. Over 160 tests have been reviewed by the Psychological Testing Centre test review editorial team, who are all BPS chartered psychologists and experts in the field of testing and test use.

a) A summary review of individual tests is available free of charge. These cover the fields of work and occupations, clinical, neuropsychology, and education and training.

b) The full report can be purchased but is free for members of The British Psychological Society and the RQTU. This report contains more detailed information about the reliability, validity and use of a particular test.

c) A test registration does not automatically ensure quality, but publishers are more likely to submit their tests to be reviewed if they are confident in their validity and reliability.

- **Question developers and test publishers**

When commissioning or tendering for psychometric tests, question providers on how and why the assessment will be suitable for your organisation and the intended use. Request details on the validity and reliability of the measures, as well as the development process. Find out if the assessment has been used in peer-reviewed studies published in academic journals and studies or is listed in the BPS Registered Test Database, and if it has not, ask why. If the provider cannot answer these questions directly or easily, it may indicate a lack of quality.

- **Follow the assessment guidance**

To ensure standardisation, it is crucial that the assessment is administered and scored consistently according to the instructions. This includes making any necessary adaptations or adjustments for specific contexts (e.g. virtual vs in-person administration, individual or group settings, or different demographic groups). Following these guidelines helps maintain the fairness, validity and reliability of the assessment. Make sure the assessments are used in appropriate contexts and comply with any existing restrictions. For example, some assessments are not suitable for recruitment or diagnosis or have copyright or commercial restrictions.

- **Be mindful of the Barnum effect and the Guru effect**

It can be easy to be sucked into using psychometric tests which appear to be good – but could be a result of the Barnum effect, which occurs when individuals believe that general personality descriptions apply specifically to themselves. This explains why generic horoscopes seem personalised. Similarly, be wary of the Guru effect, where people use assessments simply because they are popular, or endorsed by well-known organisations or individuals, without critically evaluating their actual effectiveness or suitability.

- **Evaluate the process**

Assess whether the test achieved its intended goals by examining not just the user and administrator experiences, but also the overall impact. For example, if the test aimed to enhance health, teamwork, productivity or recruitment, did this actually occur? Consult with your test provider or psychologist on best practices for conducting a thorough evaluation.

## Conclusion

Psychometric assessments can provide reliable, valid and useful information when used correctly, offering significant benefits in the assessment process. It is therefore important to handle them with due professionalism, expert training and advice as appropriate to avoid serious ethical and legal issues.

Supporting occupational health and wellbeing professionals



**2 St Andrews Place, London NW1 4LB**

Charity Commission No: 1184142
Scottish Charity No: SC049793
VAT No: 927 0030 57