Personality Assessment: Overview

Adrian Furnham, University College London, London, UK

© 2015 Elsevier Ltd. All rights reserved.
This article is a revision of the previous edition article by D. Cervone, G.V. Caprara, volume 16, pp. 11281–11287, © 2001, Elsevier Ltd.

Abstract

This article looks at different essential methods to assess personality and individual differences. It considers a simple but robust selection model, which includes selecting out and selecting in people at work. It then goes on to consider current trends and recent developments in the area. The main part of this article considers, in more detail, the advantages and disadvantages of the standard methods like interviews, references, biodata, and tests. The final part of the article concerns addressing 12 myths concerning psychometric testing in general.

Introduction

There is a great deal of interest and research into assessing people at work. Assessment can be formal and informal; based on test data or personal intuitive judgments; with more or less serious consequences. Many people invest a large amount of time and energy in assessing others which they believe is a good investment.

Assessing people at work is important for many reasons. The most important is the cost benefit analysis: the benefits of the right decision over the costs of getting it wrong. In other words, there are great financial benefits for hiring a positive and productive person; and many costs, especially financial, to hiring an unhappy, unproductive, and difficult person.

Using good assessment tools has other benefits: it can ‘up-skill’ managers who use various tools and techniques and increase their psychological mindedness. It also has a major benefit for the interviewee as feedback can considerably increase their self-awareness. Further, if a test is used widely in an organization people can often have a useful shared language to discuss issues in terms of psychological profiles and preferences.

The question for assessors is essentially what to assess, who is best suited to do it, when and why. To some extent the ‘what’ can neatly be divided into three areas.

What a person can do? This refers to their ability. It is about their capacity to do various tasks efficiently given that they have the desire to do so. It also refers to their ability to learn new tasks. Assessing what a person can do is more often measured by cognitive ability (intelligence) and skills tests though it may also be useful to assess creativity as well as the ability to lead others.

What a person will do? This refers to a person’s motivation or what they want to do. Motivation refers to a person’s values and drives. Everyone can be persuaded to do things as a function of rewards and punishments but this refers to what a person will do on an everyday basis without strong rewards or punishments trying to shape behavior.

What a person wants to do? This refers to preferences/personality for certain activities over others. It is about what a person likes to do and will do so freely without any form of coercion. It is about their values and personality and motivation which pushes them in one direction or another.

Assessors need to know all three things about a person they are assessing.

The Essential Methods

There are, in essence, mainly five different methods to collect data on people. Of these the first three are most commonly used.

Self-Report

This is essentially what people say about themselves in interviews (both structured and unstructured); in self-report personality and other preference tests; and on their CV (personal statement or application form). These are very common ways of assessing people and they nearly always want and expect an interview when they can answer questions and talk about themselves.

There are however two major problems with self-report. The first is called under various names like dissimulation, faking, or lying. It concerns people giving false information, or embellished information about themselves. This behavior has been broken down by psychologist into two further types of behavior.

Impression Management

This is where the person attempts to create a good impression by leaving out information, adding untrue information (errors of omission and commission) as well as giving answers that are not strictly correct but, they hope, create a good impression in the interviewer’s mind. This is done consciously and is very common. Indeed, it is expected in the answer to some questions but it can be very serious when, for instance, people claim to have qualifications or experiences they have not had, or leave out important information (about their health, criminal past).

Self-Deception

This occurs when a person, in their own view answers honestly, but what they say is untrue because they lack self-awareness. Thus they might honestly believe that they are a ‘good listener’ whereas all the evidence from reliable sources is that
this is not true. This can occur for both good qualities (cognitive and emotional intelligence) as well as weaknesses (impulsivity, depression). People with low self-awareness often self-deceive.

The way personality and other preference tests attempt to deal with this issue is to use Lie Scales in the test. These go under various names (validity, dissimulation) and many exist. They are generally known as measures of response bias.

The second is about self-insight. This is primarily concerned with what people cannot say about themselves even if they wanted to. This is best seen with issue around motivation where people cannot, rather than will not, give honest answers about the extent to which they are motivated by power, recognition, money, or security. Indeed motivation is one of the most difficult topics to accurately assess, and yet, for business people, though as among the most important.

One way psychologists have tried to deal with this issue is projective techniques such as the Thematic Apperception Test, or various sentence completion tasks. For instance, a person may be asked to complete the following sentence: “My greatest regret is...”; “At work I often appear to be...”; “My parents would be most proud of my...” The idea is to content analyze these responses to detect themes. These methods are expensive in terms of time and unreliable meaning too many minor factors (like their mood and where the interview takes place) affect their response. They remain ‘unproven’ for most psychometricians.

Observation Data

This is what other people say about an individual in references and testimonials; 360 ratings (multisource feedback) and appraisal and other performance management data. Most organizations attempt to get reliable reports from other people who know the candidate that they are attempting to assess. Many application processes ask candidates to list people who know them well in some salient setting and may be called upon.

There are also problems with this type of data. The first is the ‘data bank’ of the observer. This essentially means what information the observer has about the candidate. Thus a boss has a different data set than a colleague or a subordinate.

A school teacher or university lecturer will have a different data set than an employer. The question is what they know: the quality, and quantity of data on a person’s ability, motivation, work style, etc.

The second issue is the extent to which they are prepared to tell the truth about an individual. Some organizations refuse staff to write references because of litigation. They can be taken to court for what they did say or did not say. They are told all they can say is that "X worked here from dates A to dates B." People sometimes have a vested interest in ‘talking up’ a person they want to leave, or else they do not disclose important aspect about an individual’s work history such as accusations of inappropriate behavior.

People also choose their specific, favorite referees because they hope that they will be very positive. There seems to be an etiquette with respect to what people write or rate on references. Many know the power of negative information and therefore try strongly to resist providing any negative information. It is therefore rare to get very accurate or useful data on a person’s weaknesses or challenges from references.

Test Performance

This refers to how well people do on tests: power, timed, ability tests. These are of maximum performance. There are also preference, untimed, personality tests which are of typical performance. There are also behavioral tests, which measure actual observed and measured behavior often in groups.

There are thousands of tests of many types to choose from. Most professionals only know about a smaller number. They also are often not clear about why one test maybe psychologically better than another. This is nearly always a question of evidence of the psychometric properties of the test. Tests differ enormously. Some require a one-to-one administration, others can be easily and effectively be run in largish groups. Some are ‘objective’ versus open-ended tests: the former requires the choice of several responses; the latter means one has to generate the response.

Physiological Evidence

This is probably the newest and most disputed of all measures. Thus for some jobs, employees have to go through a medical checkup which they may have to do on a regular (i.e., annual) basis simply to keep their job. This would be true of such jobs as being an airplane pilot. In other jobs, for instance, working in the alcohol industry, it maybe a requirement that people go through a liver function test. Simple blood tests and saliva samples can be used for various diagnoses including drug taking and stress levels. Everyday it seems new simple physical measures are being devised that are claimed to be able to detect such things as whether a person is more likely to get a debilitating mental or physical disease.

Personal History/Biography

This refers to person’s personal history, for instance, where they were born and educated; the family from which they came; and their present family and address. Some information is thought to be very important such as what was the social class of the parents; does the person come from a minority race or religious group; how many brothers and sisters do they have and what is their place in the birth order; what was their schooling like and how successful were they at it.

This area is called biodata. It aims to determine, through empirical methods the biographical markers of success in very particular jobs. It has limitations which will be discussed later.

A Simple Selection Model

Below is the simplest selection model. The aim is to select the good and reject the bad candidates. Through job analysis selectors usually have a list of competencies that they are looking for. However there are two problems with this model. First, the assumption of linearity: this is the idea that more is better. The more you have of a quality (intelligence, creativity, integrity) the better. For most jobs you need an optimal
amount rather than a maximal amount. Second, the failure to select out: that is to actively look for things that you do not want in the person being assessed.

<table>
<thead>
<tr>
<th>Select</th>
<th>Good</th>
<th>Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

In this model, B and C are errors. Everyone is concerned with getting A but nobody with D.

The more important the job and the more the consequences of failure or derailment count, the more important it is to seek assess potential derailers.

**Recent Developments: What Is Hot and What Is Not**

There are a number of new definable products, trends, and developments in the area. These are driven by various different forces.

First, there are changes in the law: legal changes and litigation risks have driven some issues (e.g., integrity testing, diversity training). Second, there are changes in business: as many organizations attempt become more flexible and competitive they become concerned with specific assessment-related issues, e.g., in spotting and managing talent; in strengthening leadership; in managing restructuring, acquisitions, and mergers. Third, there are the ideas of gurus (business writers): popular books highlight various concepts, issues, and methods, some of which are enthusiastically embraced by business people (e.g., emotional intelligence). Fourth, there are the recommendations of consultants/academics: they often have their own agenda, which might or might not be related to the above.

Thus it is possible to see the emergence of various concepts, products, and measures used to assess people for jobs and to develop them. Further just as some of these products begin to emerge others tend to wane. Thus, 360° evaluation/feedback is currently probably at its zenith but it now much less popular than it has been. Equally, outward bound/outdoor training to strengthen teamwork seems on the wane.

Other factors have also played a part in developments in this area. First, labor market shortages: Shortages across Europe have led many companies to rethink their strategy and process to ensure it is fair to all concerned and that they can attract and correctly identify talented individuals. Second, technological developments and two issues are relevant here: The administration of tests via computers as well as recruitment and testing online via the Internet. Third, applicant perceptions which refers to the perceptions of applicants as to the fairness and validity of the assessment process. This is, in part, not only an impression management task for all organizations, but also refers to the effect of assessment methods on job acceptance and subsequent performance. Fourth, construct-driven approaches which refers to being clear about what one is trying to assess and why. That is having a theory and evidence for the factors/constructs that one is trying to assess and show their predictive validity.

There have been very specific developments and various trends are noticeable. There have been many advances in new data tools. For instance, there are now new interactive tests where people respond to carefully prepared video/acted scenes on a computer. These can even be delivered using mobile devices. They can be adapted for very specific company uses. One of the newest, but as yet untested is the use of Avatars in assessment situations. There is now more and more interest in samples of real-life behavior sometimes called ‘thin slides’ whereby people make accurate judgments of real-life behavior perhaps recorded by video. There is an increased interest on the test-takers experience and reactions. Most organizations are aware of the limitations of the traditional interview and are making efforts to introduce well structured, panel interviews. Poor validity response rates and litigation issues have meant the traditional reference is being replaced by structured telephone calls asking specific questions of targeted interviewees (often peers).

There is now great emphasis on the power and usefulness of peer and subordinate ratings for assessment purposes. This has been a major consequence of an interest in 360° multirater feedback work. Organizations are being very sensitive to the possibility of bias (sex, race, language) in ability tests. They are also particularly concerned that they have face validity for candidates. There is always the development of tests to measure new (faddish) constructs, e.g., spiritual intelligence, practical intelligence. Psychometric tests continue to be developed in large numbers and put on the market. Some clearly ‘chime with the spirit of the time’ like tests of emotional intelligence. Biodata, psychomotor tests, and tests of job knowledge are not frequently used. They have however never disappeared and remain the favorite of particular clients in very particular settings.

The concept of using work-samples and probation periods appears to be gaining more attention. Probation periods if strictly enforced give both parties (employer and employee) a chance to revisit their decisions. Assessment Centers remain the basis of many organizations’ preferred methods. They are very expensive and time consuming but recognized as the best. Most are designed very specifically for each company and there particular requirements. There is a great interest in assessing ‘people with potential’ or ‘highfliers’ but people are not sure how to do it. Organizations differ widely to whether they should bring assessment expertise in-house or outsource all or some of it. Much less interest and expertise is devoted to assessment for development purposes. This is done through mentoring and coaching which may or may not involve some formal assessment procedures. Organizations are beginning to realize that they can cost the effectiveness of successful assessment/selection in monetary terms which can have important implications.

**Assessment Methods**

A primary issue in the field of assessment is the wide choice of method to use – ranging from the informal to the highly technical.

**The Interview**

There are many variations on the job interview: How long they last; how many interviewers they have; whether they are
panel/board interviews; and most importantly whether they are structured or not!

Do different interviewers agree? That is what is the inter-rater/observer/judge’s reliability? The answer is about $r = 0.50$ and better as to their reliability and validity in structured interviews. Validity is however much lower. For structured interviews it is about the same for one-to-one versus board interviews ($r = 0.35$) but for unstructured interviews the one-to-one ($r = 0.11$) is about half as good as board interviews ($r = 0.21$). This means that interviewers of the same person do not agree very much on their assessments and that these assessments are not very useful predicting success (or failure) on the job. Only planned, structured interviews offer good data to really be useful in assessment.

Why are interviews so poor at predicting performance? There is a long and growing list of factors that collectively explain this. Interviewers differ in insight, skills, preferences. They differ in motives, attention, and need for justification of their decisions. Interviewees often try hard to manage a positive (not totally realistic) impression by self-promotion and self-enhancement. Interviewees are also increasingly being self-coached on how to behave in interviews. Put simply, interviewees (and interviewers and referees) tell lies of omission and commission. There are many and systematic variations in how interviewers use the rating scale.

There is also range restriction in ratings meaning that raters never use the full scale particularly the two ends. They do not discriminate/differentiate clearly enough between the different candidates. Too often interviewers make up their mind before the interview. Interviewers, like everyone else are susceptible to first impressions and ignoring later data. More importantly we know that reasons to reject (i.e., select out) factors have disproportionate weight compared to select-in factors. Finally many interviewers have their own (wrong, unproved, bizarre) implicit personality theories (i.e., red heads are intelligent, rugby players are good team workers).

Interviews can be improved to increase their reliability and validity by various relatively simple steps. These are very important to make the ever-popular interview a useful assessment method. Select, where possible good, insightful, natural interviewers. Always use more than one interviewer in panel or board interviews and use the same interviewers throughout the process. It is important to train interviewers in the fundamental skills of interviewing and ensure that they take notes to help memory and settle arguments.

**Personal References**

References are usually free response or ratings that an observer reports on another. That is they are observation by an individual known to the candidate. Though cheap and popular they are thought of as unreliable and of poor validity. There are three main reasons given for their poor reliability. First, leniency: Most references are indiscriminately positive. This is for three reasons. Candidates choose themselves those they know are most likely to be very positive; respondents worry negative evaluations may result in a libel suit; respondents have no incentive to take the time or tell the truth. Second, idiosyncrasy: People can and do use idiosyncratic language, examples and criteria to describe and evaluate others. Third, free-form

**References**: Reference writers are often offered no guidelines in what they are required to do. They can offer long/short, descriptive/evaluative, and relevant/irrelevant data.

It is possible to improve the validity of the reference by explaining fully the purpose of the reference, using rating scales or a forces choice format. Using well-designed rating scales also helps a great deal.

The data also suggest that peer ratings are the most useful, valid, and reliable. Referees are best when the employer chooses the referees; those referees are peers who know the candidate well; the referees are asked specific questions and are guaranteed anonymity. Switching to the Internet has been argued to provide quicker, cheaper selection and assessment with wider access for many people. It saves time and travel costs, even leading to the possibility of ‘same day offers.’

Organizations thus now make job offers (recruitment advertisements) on their Web site. Electronic application forms can be used to collect data and do a simple first-filter. Those without certain qualifications may quickly be rejected; software can be written to do a matching task between answers to questions and the ‘ideal’ profile. However, still not everyone has access to the Internet. There are age, educational, ethnicity, income, and gender correlates of those with and without access. This has legal implications. Electronic media do not bypass paper-based problems like faking (social desirability) both impression management as well as omissions and commissions. Also, sifting is fast but many still only look for keywords, thus it may be no more reliable and accurate.

Overall the new assessment technologies (predominantly the Web) have specific goals: improve efficiency, enable new screening tools, reduce costs, standardize the HR system, expand the applicant pool, promote the organizational image, increase applicant convenience.

However there are also unintended effects. The use of the Internet does expand the applicant pool but also increases the number of underqualified and out-of-country applicants. It is easy to be flooded with inappropriate applicants. There is also the loss of personal touch that both assessor and assessee value and respect. There are also concerns about cheating if tests are used. Finally there are still concerns about adverse impact for organizations when certain groups simply do not have access to the technology.

HR technology remains a challenge. The hopes of individuals have been both very high and not all the experiences have been positive. But it seems to be the future. Internet advertising and recruitment seems very cost-effective. Young people expect it.

**Biodata**

This is a method of scoring biographical factors. It makes the assumption that the past predicts the future and that certain life experiences at and before work have predictive validity. Biodata analysis is nearly always very specific. Instruments are devised for, and scored in, organizations and they have to be regularly updated.

It has many obvious advantages. First, objectivity: The same questions are asked of everyone who completes the form, and the answers given are assessed in a consistent way. Second, cost: Although there may be fairly extensive R&D
costs, once criteria are known, biodata are very cheap. Biodata forms can be developed in multiple-choice formats, which are amenable to machine scoring or direct entry to a computer terminal. Third, checkability: It is possible, in theory at any rate, to check up on the answers of the respondents. This fact may reduce faking. Fourth, validity: Nearly all the studies have shown that biodata provides a valid and reliable way of selecting individuals. Questionnaires, are as good as ability tests. Fifth, self-presentation: Most people prefer, and expect, to be asked about their biographies rather than their personalities. For many, biodata is more preferable and acceptable than personality tests.

There are also well-known disadvantages to using this method. First, there is the issue of homogeneity versus heterogeneity: If many biographical items are used in selection, the organization inevitably becomes more homogeneous, which has both advantages and disadvantages. Heterogeneity may occur across divisions (with different criteria) but not within them. Second, there is the problem of cloning the past: Biodata work on the idea that past behavior predicts current performance, but if current criteria are very unstable (say in a rapidly changing market) one is perpetually out of date. Biodata may be best in stable organizations and environments. Third, there is the ever common problem of faking: Biodata has been shown to be fakeable. Checking that a person has not faked can be difficult and extremely expensive. Next, fairness in the law: If biodata items show major biographical correlates such as sex, race, religion, and age, one may want to select or reject particular individuals, which is illegal. Items such as age, sex, and marital status may in fact be challenged by the courts if such items are included in inventories for the purpose of personnel selection. The fifth issue is that possibly minorities cannot easily be identified and treated fairly: In setting up a biodata form, the link between the information on the form and suitability for the job needs to be established. This in itself tends to ensure that any other elements are excluded from the selection process. Biodata forms, especially when computer-scored, are completely blind to incidental items such as a person’s name, which might indicate ethnic background.

Next, there is the critique that biodata do not travel well: The same criteria do not have the same predictive ability across jobs, organizations, countries, or time periods. Because criteria have to be established every time, the development of biodata can be expensive. Seventh, it should be recognized that this method is very time consuming: For example, it might take at least 12 months to obtain reliable, meaningful job-related data on new employees. If an organization does not have a regular intake of new staff, there could be 2 or 3 years delay between sending the draft biodata form to applicants and obtaining a sample of employees large enough to warrant further development work. Last, there is the issue of shrinkage over time. Biodata scoring keys do not appear to hold up indefinitely. There is evidence that the validity of biodata shrinks over time, and periodic revalidation and reweighting may be necessary.

Biodata has never become very popular nor has it ever really disappeared. There are probably three reasons why it remains a relatively little used assessment method. It takes investment of time and expertise to devise a biodata measure. Candidates do not like biodata and may think it is unfair. Organizations worry about litigation. However it is used as part data collection for many Assessment Centers. And of course, simply employing application forms to decide who to, and who not to, interview is of course, represents a use of biodata.

Cognitive and Mental Ability Tests
This refers to achievement aptitude or mental ability tests. Many terms cover the same areas: ability achievement, cognitive ability, mental ability, and intelligence tests. These are distinct from social intelligence tests, creativity, or divergent thinking tests.

Academic research has shown that quite consistently, cognitive ability accurately predicts job performance across all jobs but particularly in complex jobs. Many believe that intelligence is the single best predictor of (senior, managerial) work performance. All recent research points to the predictive power of cognitive ability and hence the importance of using these tests in selection.

The publication of a recent highly controversial book on intelligence - The Bell Curve by Herrnstein et al. – led over 50 of the world’s experts on intelligence to write to the Wall Street Journal on 15 December 1994. Their summary is an excellent and clear statement on what psychologists think about intelligence. However they do also note five points of what they call practical importance. First, IQ is strongly related, probably more so than any other single measurable human trait, to many important educational, occupational, economic, and social outcomes. Its relation to the welfare and performance of individuals is very strong in some arenas in life (education, military training), moderate but robust in others (social competence), and modest but consistent in others (law-abidingness). Whatever IQ tests measure, it is of great practical importance.

Second, a high IQ is an advantage in life because virtually all activities require some reasoning and decision-making. Conversely, a low IQ is often a disadvantage, especially in disorganized environments. However, a high IQ no more guarantees success than a low IQ guarantees failure yet the odds for success in our society greatly favor individuals with higher IQs. Third, the practical advantages of having a higher IQ increase as life settings become more complex (novel, ambiguous, changing, unpredictable, or multifaceted). For example, a high IQ is generally necessary to perform well in highly complex professional jobs; it is a considerable advantage in moderately complex jobs (crafts, clerical, and police work); but it provides less advantage in settings that require only routine decision-making or simple problem solving (unskilled work).

Fourth, differences in intelligence are certainly not the only factor affecting performance in education, training, and highly complex jobs, but intelligence is often the most important. When individuals have already been selected for high (or low) intelligence – and so do not differ much in IQ – as in graduate school (or special education), other influences on performance loom larger in comparison. Fifth, certain personality traits, special talents, aptitudes, physical capabilities, experience, and the like are important (sometimes essential) for successful performance in many jobs, but they have narrower (or unknown) applicability or ‘transferability’ across tasks and settings compared with general intelligence. Some scholars choose to refer to these other human traits as ‘intelligences.’
The recent research evidence is very clear. The single best predictor of success in complex, changing managerial jobs is intelligence. Brighter people learn faster; they have a greater store of knowledge; and they tend to be intellectually more self-confident. They analyze problems more efficiently and are less threatened by change. One can, of course, be too bright for the job if it is unskilled or unchallenging. However, all mental ability tests are the focus of much controversy.

Mental ability tests predict work performance fairly well – as well as any other method available. The relationship between mental ability and work performance is continuous and linear – the brighter someone is, the more likely their work performance is to be good. Research on mental ability and team (rather than individual) performance suggests the existence of more complex relationships. Low scorers may hold the whole team back. Research on why mental ability tests predict work performance is less well developed, but suggests that mental ability leads to improved job knowledge, which in turn leads to better work performance. Mental ability tests create a large adverse impact on some ethnic minorities in the United States. There is a dearth of information on this issue elsewhere. Attempts to solve the adverse impact problem in the US include trying to identify features of mental ability testing that will reduce adverse impact. Attempts to deal with the adverse impact problem in the US include score banding, which defines a range of scores as equivalent, thus allowing selection within the band to be based on achieving diversity.

Personality Tests

The bottom line in this complex and very hotly debated area is this. The validity of personality tests in predicting future work performance depends on the test, the criteria, and the population sample. The tests are at their most useful/accurate when a good, well-validated test is used and the criteria are clear, well measured and theoretically related to the trait is measured.

As a result of some hundred years of research and disagreement researchers have (for the most part) agreed on what is the underlying structure (i.e., personality, periodic table) of traits. The Big Five factor model is widely accepted, but may not give the best prediction of work behavior, though other models exist. Recent studies have shown that two of these are relevant to all jobs, namely trait neuroticism and conscientiousness. Successful managers have a clear profile: they tend to be low on neuroticism (i.e., very stable), low on agreeableness, average on openness, high on extraversion, and very high on conscientiousness. Different jobs require different profiles but there are none where high neuroticism and low conscientiousness are an advantage.

Personality is a vaguer and more diffuse concept than ability, and there are many models of personality. Personality is most conveniently assessed by inventory or questionnaire. It can also be assessed by interview, Assessment Center, references, or peer ratings. Personality inventories are compiled using a variety of methods which generally overlap to some extent, including internal analysis by statistics, and external analysis comparing groups of people with known characteristics.

Personality inventories have limited value in predicting how well a person can do a job. Personality inventories may be more successful in predicting effort and drive. Personality inventories may also be more successful in predicting avoidance of deviant or problematic behavior at work. Personality tests are self-reports and not tests, and can be faked. Faking appears to affect the outcome of selection. Personality tests are also widely used to screen out people with problem characteristics, such as violent tendencies, deviant sexual impulses, or dishonesty. Screening tests for employee honesty are surprisingly successful. Gender differences are often found with personality tests, but ethnicity differences do not seem to be a problem. Personality inventories have also been criticized on the grounds that some of their questions are unduly intrusive.

Among the main advantages of personality testing are first, that tests provide numeric information, which means individuals can more easily be compared on the same criteria. In interviews, different questions are asked of different candidates, and the answers are often forgotten or not properly analyzed. Next, tests give explicit and specific results on temperament and ability rather than the ambiguous coded platitudes that are often found in references. A percentage (provided of course that it is valid) makes for much clearer thinking about personal characteristics than terms such as ‘satisfactory,’ ‘sufficient,’ or ‘highflier.’ Third, it could be argued that tests are fair because they eliminate corruption, favoritism, old-boy, mason – or Oxbridge – networks from self-perpetuating. That is, if a person does not have the ability, or has a ‘dangerous’ profile, they will not be chosen, irrespective of their ‘other assets.’ Next, it is often asserted that tests are comprehensive in that they cover all the basic dimensions of personality and ability from which other behavior patterns derive. Finally psychometricians assert that tests are scientific in that they are empirically based on theoretical foundations; that is, they are reliable, valid, and able to discriminate the good from the mediocre and the average from the bad.

Many, often related, objections are made to psychological testing. Among the most common and sensible objections include the idea that many of these tests can be faked; that is, people like to put themselves in a good light and receive a good score so that they may be accepted, but this in a way reflects their real personality (some tests have lie scores to attempt to overcome this). Tests are unreliable in that all sorts of temporary factors, such as test anxiety, boredom, weariness, a headache, period pains, all lead people to give different answers on different occasions (although this is partly true, it is only a minor factor).

Most importantly, some people claim that most, if not all tests are invalid; they do not measure what they say they are measuring and those scores do not predict behavior over time. For many tests, this is indeed the Achilles heel and they are often lamentably short of robust proof of their validity. It has also been asserted that while tests might be able to measure all sorts of dimensions of behavior, but not the crucial ones to the organization such as trustworthiness and likelihood of absenteeism. Buying personality tests is like having a set menu, and what many managers want is an a la carte menu where they can select only factors or dimensions that they want.

People have to be sufficiently literate or articulate to do these tests, not to mention sufficiently familiar with North American jargon. Many organizations therefore believe their workforce
could not complete them properly, would take up too much time or would cause needless embarrassment. Others claim that there are not good norms (normative population data), at least for the population they want to test, and comparing them to White US students can be dangerously misleading. One old assertion is that tests are unfair and biased toward WASPS. Also, males tend to do better or get a more attractive profile and therefore get selected – which although ostensibly meritocratic, can conflict with antidiscriminatory legislation. It is also true that the interpretation of the tests takes skill, insight, and experience, and this is either too expensive or not available. In the wrong hands they are dangerous, because profiles are either given inaccurate or too literal interpretations. Finally freedom of information legislation may mean that candidates would be able to see, and hence, challenge either the scores themselves or their interpretation of the decisions made on them. The less objective recorded data the better. Finally, as tests become known, people could buy copies and practice, so that they know the correct or most desirable answers. This happens extensively with ability and personality testing and, when it does happen, results may have more to do with preparation and practice than actually reflecting true ability.

All assessment methods have advantages and disadvantages. Personality tests can greatly facilitate assessment. But the user needs to have some knowledge of test construction and theory both to choose tests wisely but how and when to implement them in the process. We do know that the two of the ‘Big Five’ factors seem to be important for all jobs – neuroticism which is a select-out factor because it is correlated negatively with success at work, and conscientiousness which is a select-in factor because it predicts work success. Test users would do wisely then to select tests which measure both these important traits.

But if tests do not predict behavior at work very well, the question is why not. First, there may be problems with the theoretical formulation in the first place, in that the personality theory or system has been poorly or inaccurately conceived. Quite simply, the theory upon which the test is based is wrong. Next, the measurement of the personality trait is poor for a variety of reasons, for example, it measures a single trait which is actually multidimensional. Also the measure has poor reliability and validity.

See also: Intellectual Functioning, Assessment of; Intelligence: Assessments of; Intelligence: Central Conceptions and Psychometric Models; Personality Assessment, Faking and; Personality Assessment; Psychometrics.

Bibliography


