

More myths and realities of psychometric testing

Selecting the psychometric tool which best suits your organisation and purposes can be a difficult task. In this second article in a two-part series, Dr Paul Englert, Sarah Burke and Teresa MacGregor uncover more myths and realities around testing.

JUDGING THE QUALITY OF A PSYCHOMETRIC TEST, AND choosing the best test for your organisation from the growing number of providers in the market, is a tough call, particularly if you don't have specialist training in the area of psychometrics.

In the last issue of *Employment Today*, we outlined the first seven of 15 myths surrounding the assessment testing industry. In this issue, we take a look at the remaining eight myths and what you need to know in order to make the best purchasing decisions.

Myth 8: People have a 'work personality'.

Reality: A mainstream personality questionnaire is just as valid as a 'work personality' profile. The concept of personality transcends such artificial boundaries as our work or home life. The flexibility to adapt one's behaviour in different situations or not is merely one aspect of personality, and it is not necessary to create a whole new personality questionnaire to cover it. That is done more for plausibility rather than reliability. Meta-analysis is the standard in independent research activity and meta-analysis has found that a test which provides good measures of the 'big five' personality traits does predict performance regardless of the setting. This finding is based on the analysis of patterns in the published research in a whole field, rather than on the research of one entity.

Myth 9: It doesn't matter how a tool is constructed.

Reality: The effectiveness of a tool depends primarily on how well it has been constructed. Factor analysis is generally regarded as one of the most robust statistical processes for ensuring the rigour of a psychometric tool. In particular, it allows us to find dimensions of personality that are distinct from each other and to identify questionnaire items that do not overlap. A fundamental flaw with tests which are not factor analysed is that they tend to be excessively long and claim to measure multiple, independent, dimensions of personality—sometimes as many as 20-30 traits. In reality, there is likely to be a high degree of overlap between the scales inside these assessment tools. We believe tests should be built around broadly accepted psychometric theory. This means that there is a strong reliance on the underlying factor structure that supports the tool.

Myth 10: Research material should only be given to current test users.

Reality: Research material should be made available to everyone to both further the worldwide knowledge base and to allow for

informed consumer decisions. One of the strongest indicators of a test publisher's lack of commitment to psychology is a closed-door policy with regards to accessing technical manuals and/or independent research that supports their tools. If this is made available at all, it tends to be to clients who have already bought into their marketing hype. In this way, technical information is not available for people to make informed decisions, but rather is used to preach to the converted. Moreover, it contravenes the international call for open-access to research. You can read more about this topic at: <<http://www.soros.org/openaccess>>.

Myth 11: Ipsative tests are good for making selection decisions.

Reality: Ipsative tests have been criticised by psychometricians as being inappropriate for use in selection. Using a forced choice option, they require a person to select from a range of traits the one that is most and least like themselves. Non-ipsative tests require a person to respond to a question by indicating their preference on, say, a five- or three-point scale. Put simply, ipsative tests allow the comparison of relative strengths of personality traits within an individual but do not allow comparison between people. The dilemma here, of course, is that selection is all about making comparisons between people. Ipsative tests have numerous other problems including:

- These tests cannot be normed and therefore comparisons against groups of people are nonsense;
- The results cannot be factor analysed and therefore the personality dimensions being measured must be questioned;
- They are subject to input response biases. As a result of these and other issues, the use of ipsative tools for selection has been discouraged by the British Psychological Society. Instead, ipsative tests are of more benefit in the area of individual counselling such as career guidance.

The second problem is that ipsative tests force the spread of a person's scores and while this may make interpretation easier it does not make it any more valid. It is a classic case of people being misled by false logic and simplistic marketing. Indeed, as discussed above, it is the lack of validity with respect to making comparative judgement that is the key problem with ipsative tools. For those seriously interested in exploring the misuse of ipsative personality tools we strongly recommend an article by SF Blinkhorn, et al entitled 'Spuriouser and Spuriouser: The use of ipsative personality tests' which appeared in the *Journal of Occupational Psychology* (61, 153-162).

Myth 12: Psychometric tools should only be interpreted by a psychologist.

Reality: Psychometric tools can be interpreted by anyone who has had the relevant training. Psychometric tools are built to be interpreted in a standardised way. This is why computer-based reporting is not only valid but may be less prone to error than human generated reports. A computer report will always provide a similar reading based on a person's scores and is not affected by any potential errors such as political bias or stereotyping.

Computer programmes are not, however, sophisticated enough to make all the required integrations between individual scales but provide a fair first draft of a report. Human input is required in order to make some of the more subtle scale interactions. The irony is that psychometric tools are easier to interpret than an employment interview. Moreover, just because someone is a psychologist it does not mean that they have had the training in psychometric tools required to utilise them to their potential.

Myth 13: If tests are objective anyone can interpret them and therefore training is unnecessary.

Reality: You need to be trained to make psychometric tools really useful. The rationale for training or ensuring certain skills is summarised as follows:

1. **Ethical:** Psychometric tools are used on real people to make decisions that effect their lives. People who complete these tests provide a lot of information about themselves and they are entitled to receive feedback on their results. Skills are needed to do this constructively.
2. **Standardisation:** As discussed earlier, the usefulness of psychometric tools is that they are administered in a standardised manner so that valid comparisons can be made. This requires some education and training.
3. **Legal:** Selection can be a litigious activity. It is therefore vital that best practice protocol is rigorously followed in the administration and interpretation of psychological tests as well as the delivery of candidate feedback.
4. **Utility:** Training teaches people how to maximise the usefulness of test data. In particular, training teaches people how to relate the information from each tool to competency models, selection decisions and people development.
5. **Psychological and HR guidelines:** Given points 1-4, it is understandable why the New Zealand Psychological Society and the Human Resources Institute of New Zealand stipulate that tests should only be made available to trained users. While training is of value, it should not be training simply for training's sake, nor a mechanism for extracting excessive revenue. It is a matter of providing sufficient training to use the tools provided properly, without overkill.

Myth 14: The size of a norm group is often promoted as the most important norm criteria.

Reality: The relevance and distribution is often the most important norm criteria. When a person is compared to a norm group, the size of that group has little relevance to the comparisons that can be made. Instead, it is the relevance of a norm group to your own organisation and the distribution (or spread) of scores inside that norm that is of critical importance. If a norm group is not reflective of the people you are testing and does not differentiate people well, there is no benefit in knowing that a test publisher has collected norm groups of several thousand people. If a norm group is well built and is found to be reflective of the population of interest, a norm grouping of 200 people may be sufficient!

One of the key issues with norm grouping is that we must be comparing like with like. Thus, the people inside a norm group must have sat the test in the same conditions and come from like ethnic and educational backgrounds. Only if the conditions of internet testing are sufficiently controlled do they achieve sufficient standardisation to allow the building of robust norm categories. Many internet tools do not. Beyond these criteria, a test publisher should also be able to confirm the gender mix of participants and what roles people were tested for.

All of these factors will help define the suitability of one particular comparison group over another. As a final note, the most relevant norm comparison group will often be a representative cross section of in-house staff who perform each of the role(s) under review.

Myth 15: There needs to be an additional charge for reporting.

Reality: You need only be charged once for testing. Test producers have looked at various means of extracting additional money from client organisations. This includes everything from having an organisation pay each time a candidate sits a test to every time a report is generated or both. In reality, once test data is input into a scoring system, no additional time commitment is required for a report to be automatically generated. If clients do not wish to incur the often significant reporting charges, they will opt to rely solely on verbal feedback to a candidate or recruiting manager (using just a profile chart) or they may prepare short-form, less comprehensive reports by hand for distribution.

In conclusion: Clearly not all tests are equal. Those who use assessments as part of their selection process must therefore be educated on the myths and reality of testing to be discerning purchasers. While far from a panacea for selection difficulties, if used correctly psychometric testing can be a powerful addition to any selection process, providing valuable data on the high and low performers. Too often however testing is sold by a sleight of hand and on faulty logic. Only by being aware of the myths behind the testing industry can consumers differentiate testing providers and identify those providers that meet their needs. **ct**

Paul Englert and Sarah Burke are directors and Teresa MacGregor is the South Island manager for Opra Consulting Group (www.opra.co.nz).