PSYCHOLOGICAL TESTING Art, science and mysticism

By Tom Sutton

'No matter how helpful... as a tool it may be, a psychological test cannot do its own thinking. This guiding thought is psychological theory ...' *Roy Schafer, Psychoanalytic Interpretation in Rorschach Testing, 1954*

Lawyers are making increasing use of psychologists and their tests without necessarily knowing much about the tests, nor the psychologists using them. This article will attempt to clarify a few relevant issues, mentioning some major tests, while providing a checklist against which the psychologists and their tests can be assessed.

shall not mention every useful or used test, describe tests *per se* nor focus on tests for the very young and the very old, but do not wish to decry their importance by omitting them. Which reminds me of Kevin Walsh's dictum: the absence of evidence is not evidence of absence. Dr Walsh, often considered the 'father of Australian neuropsychology', was pointing out the limits of neuropsychological (and other neurological) testing: another area I shall not focus upon.

My own research (UWA School of Psychiatry and Clinical Neuroscience) is predicated on the notion that some patients with closed head injury do not malinger or have neurotic characteristics, but are unable to get better as quickly as one expects, due to biological differences. I look for various genes and proteins that might identify these people and cause their poor repair mechanisms. Currently, some psychologists label those with poor recovery as being malingering, or having 'functional problems', poor motivation, secondary gain, somatisation, and so forth.

Which reminds me – as a good Freudian, free association is a wonderful thing – of another rule in testing: the more tests one gives, the more likely one of the tests will demonstrate a pathology, even when no pathology exists. This is due to both the probability that as the number of tests given increases, an 'abnormal' result is more likely; and that normal, healthy individuals can produce the odd 'odd' result,¹ much as people can have odd variations in their internal physiology and anatomy without undue ramifications.

Slightly related to this are the notions of *nomothetic* and *idiographic*: the former describes general laws, which in psychological testing essentially boil down to group data trends based upon research; while 'idiographic' refers to detailed descriptions of the individual case.

For the clinician and the lawyer, the question about any >>

individual's test results is important: do we interpret their score as part of the group data, or are they one of those in the sample who do not participate in the general trend, maybe are even an 'outlier' (the real rebels in health research)? This important decision is why the lawyer needs to insist their psychologist has sufficient breadth of clinical experience to spot the individual who may not be 'part of the herd'. Breadth of clinical experience is not simply being old.

In medicolegal cases, the psychologist often chooses the nomothetic approach to interpreting test results, as this is an evidential-based method of deciding what the client's scores mean. In clinical work, idiographic methods dominate. Sometimes medicolegal psychologists use the nomothetic method to protect themselves, as they can cite the research without having to put themselves on the line, or to make up for a lack of experience in the test and/or the type of condition presented.

My advice, if you want the best appraisal of a client, is to use a psychologist who is not afraid to adopt idiographic approaches to medicolegal cases: but they need not cite their longevity as a psychologist as evidence for their opinion, but clinical cases they have seen and verified, alongside any other research evidence available.

For example, 30 years ago I saw a 43-year-old nursing sister admitted to the psychiatric ward for deep depression: her test results, based on the normative data, were equivocal and unusual, but consistent with depression, and the staff decided to treat her depression with aggressive medication. She was later confirmed to have a rare measles-type virus invading her brain and died with a global dementia. The history, her presentation and test data remain in my memory due to my inability to make a correct diagnosis. At the same time, a 14-year-old girl was sent to psychiatry as she was not performing well at boarding school, and family issues were present. Again, the data were equivocal and highly atypical of an 'organic' condition, but due to her age we did not opt straight away for a psychiatric approach. She was confirmed as having subacute sclerosing panencephalitis, a progressive brain disease. Both cases – their history, the way the person presents and their test results – remain as guide posts in my memory, ready to be called to mind should anything similar arrive

Schafer's observation refers to the variety of theories embedded in the construction of a test, the theoretical orientation of the psychologist and the perceptual framework of the tester. Most tests of intelligence assess function and the concept - intelligence - means no more than the test score. These tests are useful 'tools' and good for predicting, say, educational performance. But many purpose-built neuropsychological tests use theoretical terms (such as 'word fluency'), which assume a brain mechanism or system independent of the test measure, and assume the test score measures something other than simply the test score. These are two different kinds of animal, and inferences from both should not be confused.² Describing intellectual functional ability is not the same as ascribing brain damage, and the psychologist needs to abide by the theoretical terms built into the test construction: which often they do not.

That is, strictly speaking we can sometimes say a test score describes a function and is predictive of future performance, and sometimes that a test score indicates brain damage or dysfunction to a system within the brain, but we cannot use the same test to say both. But often we do.

There is a great deal of research science and empirical methodology behind psychological instruments, as with any other scientific instrument. Indeed, unlike single session unstructured subjective interview assessments, we know exactly the *errors of measurement* in any one instrument, along with its *reliability* (the capacity to achieve the same result on separate occasions) and *validity* (the capacity to measure what it is supposed to measure). *Normative data* mean that the comparison sample pool is wider and more objective than our own clinical experience.³

So we have five good rules to begin with:

- 1. No 'Tester' is theory free: the data are not neutral when interpreted.
- 2. Psychological tests do not find every disorder or dysfunction.
- Psychological tests can find disorder when none is present.
- 4. Psychologists need clinical experience.
- 5. Psychological tests are subject to rigorous public criteria, while subjective judgements based on interview are not.

LAWYER'S DILEMMA: WHICH SPECIALIST PSYCHOLOGIST TO CHOOSE?

Psychological tests do not differentiate between settings and psychological specialities. The WAIS (Wechsler Adult Intelligence Scale) measures intelligence. It does this regardless of which professional it is used by. This goes for all psychological tests, including those assessing personality and emotional states, vocational abilities or pain-coping strategies.

Patients do not differentiate between the psychological specialities or the tests required. An injured patient may have suffered a combination of traumatic brain injury, chronic pain from other injuries, emotional changes, while having preexisting personality issues, and needing vocational guidance for future work directions.

Psychologists do differentiate and their training reflects this. The profession, of all the health professions, has managed the unique feat of both not having a health practitioner training undergraduate degree and specialising before they acquire generalist training. They gain legal registration under state, territory and now federal health portfolios with these poor qualifications. Your neuropsychologist or clinical psychologist will have a two-year practitioner training degree which includes time spent on research: that is, they will have spent an extremely limited amount of time on acquiring clinical experience with tests and with patients.

Psychologists doing psychological testing should have experience of all major tests, in all clinical circumstances, if they are to properly assess a patient (your client), who has not conveniently demarcated themselves into special areas of existence.

The specialist psychologist (neuropsychological and so on) is more likely than not to have little experience in the administration and interpretation of tests outside their area of training: but the patient requires differential diagnoses and an overall assessment of all components of their state, be it brain injury, emotional, pre-morbid personality and pain experience.

By definition, rare events are rare. One needs to give tests many times before a rare case is seen. But it is these that stick in one's clinical mind. Extensive experience in giving many tests to many patients in many contexts is a prerequisite for an expert to become an expert.

If you are employing someone to provide a *specialist opinion*, then my suggested checklist to ensure the psychologist, regardless of title, position, paper credits and professorial status, has some relevant skills, knowledge and experience is as follows:

CHOOSING YOUR PSYCHOLOGIST: CHECKLIST

- 1. Ensure that they are legally registered in their state or territory (or under the new federal scheme).
- 2. Ensure that they have worked *full time*, in a hospital where they have tested and treated patients from a psychiatric facility, from general and specialist wards, where patients with pain, brain injury and general medical conditions required both test assessment and treatment. The amount of time spent in the clinical facility should be sufficient for broad testing and treatment experience skills to have been acquired.
- 3. Ensure that they have been appropriately trained or exposed to personality and neurocognitive theory-based treatment approaches beyond 'CBT' (which has no theory of personality dynamic underpinning it). Check to see if they belong, or have belonged, to suitable specialist associations – for example, the Australian Society for Hypnosis, and so on.
- 4. If they are a neuropsychologist or forensic psychologist, or some other 'specialist', then further university post-graduate training and specialised experience is acquired after the generalist training in points 2 and 3 above, with further experience gained by working in relevant facilities to the speciality area for example, prisons for a forensic psychologist.
- 5. Ensure that they possess the latest acceptable psychological tests used in assessment, covering the range of conditions they are likely to see in a normal generalist clinical practice, including alternative batteries of the major cognitive systems (intelligence, memory, executive functions), tests of emotional functioning and psychiatric disorder, vocational aptitude, and pain management.

ADVERSARIAL PSYCHOLOGISTS: PART A

Because the medicolegal system often has psychologists from 'both sides' assessing the one individual, it is important that they can deploy alternative methods of assessment to avoid bias and practice effects. For example, they should hold more than one set of measures of intelligence, memory or executive functions. Many psychologists in private practice do not bother to purchase many tests, being rather fiscally 'tight': point 5 above should always be queried.

Some very brief examples of alternative tests that can be held in the test library:

ADULT INTELLIGENCE:

WAIS IV: Wechsler Adult Intelligence Scale. Ages 16 – 90+ Woodcock Johnson III: Ages 2 – 90+

MEMORY BATTERIES:

WMS IV: Wechsler Memory Scale. Ages 16 – 90.11 WRAML2: Wide Range Assessment of Memory and Learning. Ages 5 –90

MTOA: Memory Test for Older Adults. Ages 55+ LAMB: Learning and Memory Battery. Ages 20–80 Doors and People: Ages 18 – 80 There are many alternative tests and batteries: ensure

your psychologist has a decent test library with alternative tests of major cognitive processes.

Another way of dealing with practice effects from too many psychologists testing the same client is by using a Reliable Change Index (RCI): a method for determining if changes in tests scores over time are reliable. There are technical problems with the RCI concept, but it is the best we have.

- If two psychologists have tested the same patient:
- Ensure your psychologist has alternative tests in their library; or
- that they calculate RCI scores wherever possible.

ADVERSARIAL PSYCHOLOGISTS: PART B

The adversarial legal context may produce a different set of motives, anxieties and defensive adaptations in the client than does the clinical arena. As a clinical problem, patient and psychologist share a common goal and focus on the nature of the disorder causing suffering; but as a medicolegal problem, psychologists have become more interested in the nature of the litigant not the disorder.

As a result, the legal context is driving the creation of a sophisticated range of assessments based upon theory and research for detecting malingered memory and thinking, called SVT: Symptom Validity Testing. These measure the degree of 'effort' placed into the performance of a cognitive test, and are often used as markers of 'malingering'. In the cognitive domain, the test must:

- be sensitive to malingering;
- be insensitive to variables that affect memory such as age, education and neurological impairments;
- have face validity as a test of memory or reasoning; and

• have perceived difficulty greater than actual difficulty.⁴ All test instruments have empirically determined levels of *sensitivity* and *specificity*. Sensitivity is a true positive and refers to the ability of the test to truly identify the diagnostic category in question; while specificity is a true negative. We really need high sensitivity. If we have 100 malingerers, the sensitivity is the number out of this 100 we can identify; and if we have 100 non-malingerers, the specificity is the number we can identify from this group.

An old SVT (and one that should now not be used), the Rey 15 item, has very good specificity at 97.5 per cent, (truly identifying 97.5 out of every 100 non-malingerers) but very poor sensitivity, identifying, on average (in the published >> research), 4.9 per cent of malingerers (4.9 out of 100 true malingerers).

The SVTs that one should use have sensitivities in the high 90s, meaning that they truly identify over 90 out of 100 malingerers, and have been tested on many different patient groups (dementia, depression, brain damage, young children,

elderly), as well as those faking and exaggerating cognitive dysfunction.

There are numerous 'embedded' methods of gauging effort: these use a standard test of reasoning, memory, etc, but have internal checks to ensure that the responses 'make sense', are consistent, and have been validated on compliant and noncompliant patient groups.

The major tests of psychiatric functioning – MMPI (Minnesota Multiphasic Personality Inventory) and PAI (Personality Assessment Inventory) also have validity checks. Most checklists of specific moods, PTSD and pain do not have validity checks.

Recommendations:

- 1. Ensure that the SVTs used for assessing memory and reasoning 'effort' are instruments designed and researched for that use: any by Green, Tombaugh and Fredericks are acceptable.
- 2. Emotional assessments should use a major test such as the MMPI or PAI, which have numerous validation indices.

Psychological testing, both SVT and validity scales on the major personality instruments, is the most scientific way of assessing whether a patient is producing optimal effort on cognitive testing, and in reporting their emotional symptoms without bias. The empirical nature of the test has been repeatedly shown to be far superior to that of a subjective interview when making these very important judgements. Remember:

- 1. Malingering is not an all-or-none phenomenon, it is likely to be a continuum.
- 2. The existence of malingering does not preclude the existence of bona fide symptoms.
- 3. Malingering is not a personality trait.
- 4. Empirically researched tests are best at assessing effort and symptom reporting.

TWO CURRENT ISSUES IN PSYCHOLOGICAL TESTING

1. Should one use the latest version of a test?

Older versions have far more research attached to them and hence are more validated, more widely normed on different samples and better understood. New versions may have better psychometric properties, more up-to-date norms and current item content (a picture of a mobile phone as opposed to a bakelite dial model), but lack research on

Rule: Check the date of the latest version of the test being used by the psychologist, and the latest version of the test itself. relevant patient samples. The general rule is: 'how old is the old test and how recent is the new version?' Current use of the WAISIII and WMSIII, despite the recent release of the WAISIV and WMSIV, is arguably permissible. However, using the WAISR (published in 1981) and WMSR (published in 1987) a full ten years

into the 21st century when the WAISIII and WMSIII were published in 1997, is not. The normative groups for both the older instruments are 20 to 25 years old, and the 'Flynn Effect' means IQs are inflated: the effect describes a finding suggesting that IQs on tests increase with each generation. Rule: Check the date of the latest version of the test being used by the psychologist, and the latest version of the test itself.

2. Are computers useful in testing?

At work, I am not connected to the internet: never get an email, and cannot google – life is good. But computers are very useful for some activities and are increasingly employed by testing companies and psychologists. One major use is the computation of many scores against many parameters. This saves much time and is a very good use of the computer.

Another use is for the computer to 'write' the report by describing the computed results in the form of a narrative. There are some psychologists who either rely upon the narrative report from the computer as the interpretation of the results or worse, copy and paste that narrative as part of their report.

The first issue is that computers suffer obsessivecompulsive personality disorder (OCD), which is a pattern of preoccupation with orderliness, perfectionism and control. This involves a degree of rigidity due to having high internal rules and following these in an inflexible or unyielding manner. The computer program brings in an interpretation if a score hits or exceeds a predetermined number, say T = 70, without regard to whether the person may have scored 69 or 71.

And the computer does not know the history of the patient. Psychologists often interpret the scores on the PAI and MMPI from patients with chronic pain as if they were either part of the normal population or a psychiatric sample, rather than a patient with pain. Extensive research shows that pain patients are more depressed than normal people and have, with good reason, more pre-occupation with their health; and should not be seen as hypochondriacal, or diagnosed with somatisation because of these 'psychiatrically' elevated concerns. Patients with pain and injury need to be compared with other chronic-pain patient samples. These norms are available for the MMPI and PAI and should be used by your testing psychologist.

Computers are very good at measuring reaction time, which is how fast you can respond to a stimulus. This has a very long history in general experimental psychology and as a measure of information-processing speed in pathology. Information-processing speed underpins many higher-level cognitive processes such as working memory; is a major cause in the decline of mental abilities in normal ageing; is a primary symptom in traumatic brain injury and certain neurological conditions such as multiple sclerosis, where it will also influence cognitive functions such as memory, verbal comprehension, etc.

A typical test is the CTIP (Computerised Test of Information Processing), which measures three reaction time tasks: simple (pure speed of information processing), choice (added complexity with a decision component) and semantic (most complex with a conceptual component to the decision process).

Remember:

- 1. Computerised reports and interpretations suffer from OCD.
- 2. Computers do not know the patient's history.
- 3. Computers are excellent for computing complex scores.
- 4. Some individual tests are suitable for computerisation.
- 5. Computers are excellent for measuring reaction time and speed of information processing.

SHORT NOTES ON PERSONALITY TESTS

The two major tests used in clinical and forensic diagnostic assessments are the MMPI-2 (Minnesota Multiphasic Personality Inventory 2001) and PAI: Personality Assessment Inventory (1990). They are both self-answer questionnaires (576 and 344 questions respectively), can be computerscored and interpreted. They sample a wide range of disorders and both have a large number of validity scales.

Validity scales assess potential limitations to the accuracy of the information provided by the respondent, and are superior to a clinician's (psychologist or psychiatrist) subjective judgement as to exaggeration or defensiveness in a patient's verbal symptom reporting.

The PAI has been gradually supplanting the MMPI, despite the latter's rich research history. The PAI is shorter, based on modern nosology, feels cleaner than the MMPI, less clunky, and needs less tinkering with to 'make sense of' when interpreting the results. But it requires correct interpretation. The lawyer should ensure that the psychologist using it has received expert training in the PAI's use, has used it on a large number and variety of patients, including non-medicolegal treatment settings, and will not simply refer to either the computer interpretations or technical manual descriptions.

Despite the word 'personality' in both instruments, neither is attached to a theory of personality. There are many 'personality' tests, only some of which actually are of personality, or underpinned by a theory of personality.

The major tests that relate to a theory of personality, as the layman understands it, are the so-called 'projective tests' such as the RIM (Rorschach Inkblot Method), where psychodynamic models are referred to. The distinction between 'objective' and 'projective' tests is odd: patients subjectively decide how to answer the questions on the MMPI and PAI ('objective' tests), but they cannot do so on 'projective' methods, where they produce behaviour and do not 'project'. Both types of tests have 'objective' marking and coding criteria (though it takes much great training, experience and brain power to do so on the 'projective' tasks). The latter is why most psychologists find the literature from academic psychologists on the lack of scientific credibility of the Rorschach so appealing: it saves them an awful lot of time in training and scoring the tests if they stick to self-answer questionnaires.

MMPI and NEO (a Five Factor Theory Personality test) authors, Butcher⁵ and Costa⁶, have no issues in appearing alongside their Rorschach colleagues, Weiner⁷ and Erdberg⁸, in training sessions assessing the same patient. The artificial controversy over 'objective' and projective testing is limited to academic university circles. Disregard it.

The bottom line in personality testing

If a major test of personality function or clinical diagnostic assessment is used, check that the psychologist:

- 1. has been properly trained in its use by an expert on that test and not just read the manual;
- 2. has given it to a very large variety of patients with different conditions, in different contexts: especially non-medicolegal settings with a clinical treatment emphasis; and
- 3. has a theory of personality underpinning their conclusions.

If they have no experience with the Rorschach, regardless of what they think of it, then they have not really been exposed to clinical pathology assessment.

Notes: 1 In the 1970s, I gave the Rey Complex Figure - a complex geometric design - to Army Reserve recruits alongside their normal military selection battery. Despite a clear medical history and passing tests of literacy and reasoning, the number of healthy functioning young men with 'pathological' scores on this test was interesting. 2 For those interested: Eric Dowling (UNSW 1970) referred to these as 'Theoretical Terms (TT)', citing the two different types of TTs as Hypothetical Constructs (HC) and Intervening Variables (IV) - described by MacCorquodale and Meehl in 1948 - and the theories they were embedded in as essentialism and instrumentalism (Popper), respectively. Without going into detail, a test built upon IV terms cannot be interpreted in the same way as one using HCs. IVs are useful tools for predicting, as many tests of intelligence and vocational ability are; HCs are useful for locating something in space, as many neuropsychological and some personality (for example, Rorschach Inkblots) tests are. One can use the former IV/instrumentalist test to measure and predict a function, and the latter HC/essentialist type to infer brain damage, but not vice versa. 3 See GJI Meyer, et al, Psychological Testing and Psychological Assessment: A Review of Evidence and Issues. American Psychologist, 2001. 56(2): pp128-165. 4 TN Tombaugh, Seminar to Australian College of Clinical Psychologists. 2000. John James Hospital, Canberra ACT. 5 JN Butcher, El Megargee, Minnesota Multiphasic Personality Inventory-2 (MMPI-2), The University of Minnesota Press. 6 PT Costa Jr, RR McCrae, Neo Personality Inventory - Revised (NEO PI-R), PAR 1985, 1989, 1991. 7 IB Weiner, Principles of Rorschach Interpretation, Lawrence Erlbaum Associates Mahwah, NJ 1998. 8 JE Exner Jr, P Erdberg, The Rorschach, Advanced Interpretation, John Wiley & Sons, 2005.

Tom Sutton worked in psychiatric and general hospitals for 18 years, commencing private practice in 1988. He has been Chair of the ACT Psychologists' Registration Board, member of the Australian Army Psychology Corps and currently researches genetic and biomarker characteristics in recovery from traumatic brain injury.