Contents

A few words from your editor .................................................................................................................. 3
TAMERON CHAPPELL, CHARTERED OCCUPATIONAL PSYCHOLOGIST, HARVEY NASH

Committee update and news .................................................................................................................. 4

Pioneers and landmarks in intelligence testing:
5: Making intelligence testing more scientific:
Charles Spearman (1863–1945) ........................................................................................................ 6
DR. HUGH MCCREDIE, INDEPENDENT CHARTERED PSYCHOLOGIST

Neuroscience and Coaching .................................................................................................................. 9
DIANE NEWELL, MANAGING DIRECTOR OF OCM’S COACHING SERVICES, PRESENTING AT TPF 8TH MARCH EVENT.
Review by Sarah Perrott, Director of Cresco Consulting

Mental toughness – assessing how we think: it’s applications in developing performance, wellbeing and positive behaviours ............................................................................................................. 11
DOUG STRYCHARZCYK, CEO FOR AQR INTERNATIONAL, PRESENTING AT TPF 8TH MARCH EVENT.
Review by Victoria Hall, Talent Futures

New Frontiers 2017: Tales From the Frontier ....................................................................................... 17
- THE BIG FIVE AND DYSFUNCTIONALITY: DR HUGH MCCREDIE, INDEPENDENT PSYCHOLOGIST
- PERSON * ENVIRONMENT = BEHAVIOUR (ADAPTIVE AND MALADAPTIVE):
  DR. RAINER KURZ, MANAGING CONSULTANT AT CUBIKS GROUP
- NEW INSIGHTS INTO TEAM DEVELOPMENT: DR. NIGEL GUENOLE AT GOLDSMITHS, UNIVERSITY OF LONDON
- VALUES-BASED RECRUITMENT - CONSIDERATIONS AND CONSEQUENCES:
  JAMES BYWATER, DIRECTOR OF PRODUCT AND INNOVATION AT KORN FERRY

All four presented at the TPF 28th November event and are reviewed by Dr. Hugh McCormie, Independent Chartered Psychologist

Automating the Generation of 3D Rotational Items and Modelling Item Difficulty Using Item Response Theory .......................................................................................................................... 19
PRIZEWINNER OF THE 2017 PRACTITIONER EXCELLENCE AWARD
PHD WORK OF AIDEN LOE, UNIVERSITY OF CAMBRIDGE PSYCHOMETRICS CENTRE

Upcoming Events

2018 events – remaining dates

Thursday 17th May: Management of change in light of big data
We will be joined by Ben Thornton, Director of PwC HR Consulting Practice in an interactive session where he will share a number of recent PwC thought leadership papers on trends and Big Data.
Professor Peter Saville will also be speaking about the use and influence of Big Data in the afternoon session.

Tuesday 18th September: Leadership and development – The transition from middle to senior management

Wednesday 28th November: New Frontiers in Psychometrics

Wednesday 28th November: Mulled Wine and Mince Pies Networking following New Frontiers event
Welcome...

A few words from your editor...

TAMERON CHAPPELL
CHARTERED OCCUPATIONAL PSYCHOLOGIST, HARVEY NASH

Dear colleagues,

Welcome to the Spring edition of Psyche. This edition reviews the first of our 2018 workshops, the New Frontiers event from 2017 and showcases the winning poster from our Practitioner Excellence Award. We also have part 5 in the series Pioneers and Landmarks in Intelligence Testing from Dr. Hugh McCredie which does as ever, also appear in Assessment and Development Matters in a shortened form.

The content of our events for 2018 have been chosen to reflect the wishes of our membership (or the wishes of those that completed our surveys at least!) and we started the year with an event that is particularly interesting in light of the articles in the most recent edition of The Psychologist around ‘unlocking the social cure’ and how we need to look more comprehensively at improving mental health and wellbeing by considering social identity. Doug Strycharczyk talked about mental toughness and AQR’s assessment to measure mental toughness which looks at resilience through the more traditional lens of temperament. Diane Newell shared how she uses neuroscience in her coaching practice and referred, amongst others things to amygdala hijack and our chimp brains being in charge. If you’re curious about how the world might be if our chimp brains were in charge then read the theatre review in The Psychologist of Great Apes which is a play based on the Will Self novel where a man wakes from a night of revelry to discover that everyone is now a chimpanzee with their Chimp Brains making decisions. The protagonist is treated for being under the psychotic delusion that he’s a human.

Also in part on the darker side of human nature, our New Frontiers event last year chaired by Hugh McCredie saw presentations from Hugh himself, Dr. Rainer Kurz, Dr. Nigel Guenole and James Bywater and was as ever, one of our most popular events of the year. In the mince pies and mulled wine event following New Frontiers the prizewinner of our annual award to recognise outstanding use or development of psychometrics was Aiden Loe and many at the event commented on how ground-breaking his PhD work was. You can read for yourself how he is using AI to design systems based on cognitive models to generate psychological questions in his poster which is reproduced in this newsletter.

There are some more changes to our committee with Pavlos Stampoulidis joining us and taking on the coordination of our new blog. This brings an even more international feel to our committee with Pavlos being based in Athens and Adrian managing the LinkedIn group from Dubai. We are still hoping to find someone who can organize our New Frontier events as 2017 was Hugh’s last tour of duty in this regard. If any of you have an interest in joining the committee to organize our New Frontier events then you can speak to Hugh directly to gain insight into what was involved or to our Chairs Lynne or Trish. The event for this year is already being planned so you would have plenty of time to settle into the role!

Your committee needs you!

Tameron
The Psychometrics Forum

Committee update

The Blog
In the last edition we made a call for a Blog Manager as well as someone to take on organising the New Frontier events following from Hugh McCredie stepping down from the committee.

We’re delighted to announce that Pavlos Stampoulidis has joined the committee and will be co-ordinating our new re-vamped blog which will be launched soon. Pavlos will be supported by Nina Muir who is not an official committee member as part of our constitution asserts that we maintain our independence from providers and distributors of psychometrics. Pavlos and Nina will be working together to collect and publish blog content from various writers. If you have a topic you would like to write about for the forum please get in touch with Pavlos or Nina.

Pavlos Stampoulidis,
TPF Blog Manager
Pavlos is a Chartered Psychologist with a strong research, statistical and IT background. Former SHL Greece R&D Manager, he is the founder of Psycholate, a software company offering design and development services for software platforms and web-based applications in the field of psychometric testing. Through his company he delivers solutions such as custom online tests and automated reporting tools to HR agencies and universities in US, Europe, China and New Zealand. Based in Athens, Pavlos is a regular traveller to the UK for business development purposes but also to attend TPF Events.

Nina Muir
Nina is an Occupational Psychology Consultant at Psysoft. Her areas of focus include psychometric test training, e-learning design and development, and bespoke assessment centre creation. Nina has previous experience in HR with various organisations within the financial, engineering and agricultural sectors. Nina holds a masters degree in Occupational Psychology from Birkbeck, University of London. She is CIPD qualified and holds a Certificate in Child Counselling Using The Arts from The Institute For Arts in Therapy & Education. Nina co-created the Brighton & Hove Psychology Network and is manager of the HR Users of Psychometrics LinkedIn group. Nina has recently joined The Psychometrics Forum.

In Memory of Bill Lubbock by Dr. Hugh McCredie
I was saddened to learn that a long-term member of TPF and The 16PF Users’ Group, Bill Lubbock, died recently. Bill was noted for his sense of humour as evidenced by his contributions to Psyche, and its predecessor, in the late 1990s and early 2000s. These contributions occasionally sailed perilously close to the boundaries of taste; see Edition 38, June 2005 by way of example. His humour apart, Bill was a very humane Christian believer who shared with me a moving story about how his one-time professor at UCL, the great Alec Rodger, believed that Bill’s prayers had eased his physical suffering.

According to Bill’s Linkedin profile, he first graduated in Physics from the University of Exeter in 1946 and in Psychology from UCL in 1951. As I understand it, he was a psychologist in the Royal Navy early in his professional career and was Head of HR at

At the end of Bill’s Linkedin profile, he commented: “Retired from active work, now fully occupied with NHS appointments. Am costing them a fortune and have already had a doctors query - should we still be treating someone this old!! Perhaps they have a point.”

In his later years, Bill was often accompanied to 16PF/TPF meetings by his daughter Jo Redbond, also a psychologist.

**Training Discounts**

We continue to negotiate training discounts with publishers and suppliers of psychometrics. Our 2018 discounts appear in alphabetical order in the table on the opposite and more details are available on our website.

<table>
<thead>
<tr>
<th>Assessment Qualification Training</th>
<th>Provider</th>
<th>Discount</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ-I 2.0 &amp; EQ 360</td>
<td>Psysoft</td>
<td>20%</td>
</tr>
<tr>
<td>Hogan Suite (HP1,MVP1,HDS,HDS Advanced)</td>
<td>PCL</td>
<td>20%</td>
</tr>
<tr>
<td>Implicitly (Unconscious bias test)</td>
<td>Hogrefe</td>
<td>20%</td>
</tr>
<tr>
<td>Leadership Climate and Personality Type Profile</td>
<td>JCA</td>
<td>See website</td>
</tr>
<tr>
<td>Leadership Judgement Indicator (LJI)</td>
<td>Hogrefe</td>
<td>20%</td>
</tr>
<tr>
<td>Managing Team Roles (MTR-I)</td>
<td>Team Focus</td>
<td>75%</td>
</tr>
<tr>
<td>Mental Toughness Questionnaire (MTQ48)</td>
<td>AQR</td>
<td>20%</td>
</tr>
<tr>
<td>NEO-PI-3</td>
<td>Hogrefe</td>
<td>20%</td>
</tr>
<tr>
<td>Pearman Personality Indicator</td>
<td>Psysoft</td>
<td>20%</td>
</tr>
<tr>
<td>Profile:Match360</td>
<td>PCL</td>
<td>20%</td>
</tr>
<tr>
<td>Risk Type Compass</td>
<td>PCL</td>
<td>20%</td>
</tr>
<tr>
<td>Team Management Profile (TMP)</td>
<td>TMSDI</td>
<td>25%</td>
</tr>
<tr>
<td>TUOA/TUOP</td>
<td>Hogrefe</td>
<td>20%</td>
</tr>
<tr>
<td>Type Pro Qualification</td>
<td>Type Pro</td>
<td>40%</td>
</tr>
</tbody>
</table>
Pioneers and landmarks in intelligence testing:

5. Making intelligence testing more scientific: Charles Spearman (1863–1945)

DR HUGH MCCREDIE
INDEPENDENT CHARTERED PSYCHOLOGIST

The most recent articles in this series reported on the use of intelligence tests to resolve specific social problems. Article 3 focussed on the identification of those with special educational needs whilst Article 4 was concerned with the assessment of individuals’ military potential in war-time. In contrast, the current piece will explore the nature of intelligence and its measurement. It will feature the work of Charles Spearman FRS, a former professional soldier turned academic at University College London (UCL) who was strongly influenced by the work of Francis Galton. In his turn, Spearman was a major influence on several mid-twentieth century psychometricians who will feature in later articles.

Of all the luminaries whose work I have reviewed, Spearman’s is unquestionably the most difficult to grasp. This is mainly to do with his use of, what is now, outmoded terminology (e.g. test respondents or subjects are labelled ‘reagents’) and the unfamiliar statistical methods which are key to understanding his very important contribution. What follows is the essence of the latter as revealed in two of his offerings separated by a distance of 23 years.

‘General Intelligence’: Objectively determined and measured, (Spearman, C. (1904)).

This paper preceded the landmark contribution of Binet & Simon (1905) by a single year. Spearman reviewed the history of “mental tests” (p205) citing nearly 30 sources reporting research on measures ranging alphabetically, from accuracy in counting, to arabesque designs, association, aesthetic sentiment, arithmetical addition, attention, choice, combination, comprehending, copying accuracy, coup d’œil, discrimination of weight and size, erasure of letters, fallacy, general ability, hearing, imagination, intellectual acuteness, introspectively determined faculties, memory, memory span, mental images, metronome beats, moral sentiments, motor ability, muscular force, numbers, observation, perception, reaction-time, recognition, sensibility to pain, sensory discrimination, sight, suggestibility, tactile sensitivity, taste and will. (pp206-219). He concluded:

“There is scarcely one positive conclusion concerning the correlation between mental tests and independent practical estimates that has not been with equal force flatly contradicted.” (p219)

Spearman appears to attribute most of this discordance to the frequent absence of quantitative measures and a general failure of researchers to account for the idiosyncrasies of test circumstances and characteristics of the sample from which the data were obtained. He proposed the use of a statistic, ‘Probable Error’, involving the correlation value under consideration, the size of the sample and a constant based on a distributional assumption, to account for these intervening variables. In his own experiments, Spearman correlated tests of schoolchildren’s ability to detect variations of sound, light and weight stimuli, (as well as performance in languages, maths and music) with teachers’ estimates of general intelligence. As he reported:

“First, we must exactly determine the quantity of correlation actually
Spearman sought objective rather than subjective measures wherever possible and attempted to obtain support for his g factor by the use of the ‘tetrad equation’.

For Spearman, the key issue was how to reconcile the idea of intelligence as a unitary concept in the light of the wide variety of tests purporting to measure it. Reviewing 33 prominent journals he found the following numbers of reference to intellectual faculties: Sensory perception (15), Intellect (16), Memory (19), Imagination (11), Attention (10), Language (3), Movement (6), Various (16). He noted that Binet and Simon (1905) had resolved the issue by constituting Intelligence as the general level of ability calculated as, the average from a variety of tests. Spearman argued for correlation as a more scientific approach with its potential to recognise repetitions and omissions amongst the constituent tests. These were early days, before the full emergence of the statistical technique Factor Analysis inspired by Spearman’s work. However, he sought objective rather than subjective measures wherever possible and attempted to obtain support for his g factor by the use of the ‘tetrad equation’. The statistical methods employed are fully described in the Appendix to Spearman (1927), pp x-xii for the tetrad equation and for calculating the probable error of tetrad differences. In short, Spearman tabled the correlations amongst test scores and plotted these four at a time (hence tetrad) in a series of two by two matrices as exampled (p172) below:

<table>
<thead>
<tr>
<th>Abstract thought</th>
<th>Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reasoning</td>
<td>0.83</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.48</td>
</tr>
</tbody>
</table>

He then multiplied the correlations across the diagonals and deducted the latter from the former, thus: $0.83 \times 0.31 - 0.40 \times 0.48 = 0.065$
To support the existence of a general factor, Spearman was looking for an answer near zero. His data, in this instance, was subject to a probable error of 0.055, very close to 0.065. Remembering that to be valid he believed that the difference-to-probable error ratio would need to be 4 or 5 to 1, these data support the presence of a common factor. Applying his methods to the research data of others, Spearman found that tests which involved the "eduction" of a relationship, including tests of sensory discrimination, correlated well with \( g \). He reported:

"One conspicuous instance of educing such relations has been given in the Synonyms-Antonyms...reported by Otis and others...to conform with the criterion of \( g \) very exactly." (p173)

As may be recalled from the preceding article in this series on Pioneers, Otis was the main contributor to the US Army Alpha Test. A specimen item for this test is as follows:

If the two words of a pair mean the same or nearly the same, draw a line under same. If they mean the opposite or nearly the opposite, draw a line under opposite

**Item: Command----Obey**  **Answer:** Same Opposite

Of the tests which were not found to associate with \( g \), the most notable was that of simple retentive memory for facts. On the other hand, 'inventiveness or creativeness' did reflect the common factor as did 'psychological ability' defined as 'the understanding of the thoughts, feelings, and strivings of other people, as also the devising of effective behaviour towards them' (p195). He also reported that a test of 'reasoning from evidence' correlated 0.80 with \( g \). Spearman concluded:

"For the purpose of indicating the amount of \( g \) possessed by a person, any test will do just as well as any other, provided only that its correlation with \( g \) is equally high." (p197)

Stockton (1921) and others had suggested the superiority of tests containing abstract or symbolic items (e.g. language or mathematical concepts) over 'perceptual' stimuli (e.g. sensory or pictorial) in identifying individual differences but Spearman could find no statistical support. Spearman next turned to the issue of 'Group Factors' which "...have played...a baffling part in controversial writings" (p222); with particular regard to his Special Abilities (S). To resolve the issue as to whether any such factors existed between \( g \) and \( s \), he proposed further use of the tetrad difference statistic, but:

"Whereas before we were searching for all cases where the tetrad equation...is satisfied, we now want to discover all cases where it fails to be so....To find out whether any two abilities are linked together by overlap, specific correlation, or a group factor, there must be obtained two further abilities (reference abilities they may be called) so as to make up the tetrad. For this purpose, any two of the ordinary tests of \( g \) will serve." (p223)

He tried this out with two tests of reasoning from evidence, Inference and Likelihood, and found a tetrad difference of 0.018 compared with a probable error of 0.039, so concluded:

"The fact of both involving characteristically the relation of evidence has been insufficient to produce any such common element additional to \( g \)." (p224)

After applying the tetrad to a wide range of special abilities, faculties and forms of test material presentation, Spearman concluded:

"Group factors have been astonishingly rare... Among the exceptional cases the four most important have been in respect of...the logical, the mechanical, the psychological, and arithmetical abilities... in the few cases where the broad group factors and special abilities did make some appearance, there was commonly a suggestion of their being due to past experience rather than to native aptitude." (pp241-242)

The remainder of the book reported Spearman's attempts to use the tetrad device to explore the nature and boundaries of \( g \). He re-emphasises that \( g \) is about education rather than retention and, from this distinction, infers its association with 'mental energy'. He differentiates \( g \) from dispositional factors...
As an undergraduate at UCL, John C. Raven became friendly with Spearman and it seems highly likely that Spearman’s emphasis on the eductive nature of g and his finding that g was as readily measured by pictorial as by symbolic stimuli influenced the development of the renowned Raven’s Progressive Matrices tests.

"With...nationality and race...we found some indications of difference, especially in...g; but...very small...compared...within...the same race...the effects of heredity upon g are very large...To a more limited extent even s seems to be influenced in the same manner...differences of education between children of the same social status have but small influence upon g, however much they may have upon s...sex...showed itself here in the domain of cognition to have comparatively small effects." (p391)

As an undergraduate at UCL, John C. Raven became friendly with Spearman and it seems highly likely that Spearman’s emphasis on the eductive nature of g and his finding that g was as readily measured by pictorial as by symbolic stimuli influenced the development of the renowned Raven’s Progressive Matrices tests.

Neuroscience and coaching

PRESENTER: DIANE NEWELL, MD OF OCM’S COACHING SERVICES
REVIEWS BY SARAH PERROTT, DIRECTOR OF CRESCO CONSULTING

Diane Newell presented an engaging and thought-provoking session at The Psychometric Forum on 8th March talking about neuroscience and coaching. Firstly, she made clear her view that coaching is about change. She framed this deftly, by asking: ‘What? So What? and Now what?’. When taken in the context of neuroscience it is useful to know that there are different branches of neuroscience: behavioural, cognitive, cultural, social and systems. Diane made it plain that neuroscience is a very new science, and therefore, any research and information in this area should be carefully considered and interrogated and we should avoid ‘building castles on sand’, and be cautious about absolutes that might be proved inaccurate with future research. Neuroscience has certainly given us a better map of the brain. The Triune Brain Theory of the 1960s led to the emergence of terms such as the amygdala hijack and chimp brain. While to an extent this was and is true, it over-simplifies the functions by describing the lizard brain as our ‘autopilot’ and the mammal brain as our decision-maker, while the human brain reasons and rationalises. The mammal brain contains the limbic system, which is where our emotions, memories and habits lie.

Diane created a ‘thinking together’ environment during her session, asking the group for specific input and questions and her presentation consisted of a series of interesting areas where neuroscience could be giving us new insight into how our brain directly influences our thoughts, feelings and behaviours.

Unconscious Processing
Unconscious processing can affect our response to social threat, and our biases. Describing the role of Unconscious Processing – unconscious bias - Diane explained that
we see and encode experiences before we have even thought consciously about them. An example of this might be seeing a picture of a young black male – even before our conscious brain has a chance to make a ‘decision’ our unconscious brain might associate young black males with anger which would lead to a biased response. All this happens very fast and so Neuroscience tells us that before we have even thought about it explicitly we start from a particular position. In other words, when I see ‘people like me’ an encoded memory is activated and this then influences our following judgements.

Any input may trigger a biased response. However, visual input is the most researched, and within that modality the bias surrounding gender and skin colour has been investigated the most.

The Inter-relatedness of Thinking and Feeling
Our brains make meaning of our feelings. We may ‘think’ that we are feeling in response to reality but actually we have interpreted reality based on feelings. Thus our responses are often unconsciously triggered through associated feeling. Understanding the role of unconscious processing enables change. The key learning here, Diane pointed out, was that we should not ignore our feelings but interrogate them. Given that we know we have triggers that might lead to habitual responses, the coaching technique to use would be to practice acknowledging and identifying the feelings, to then be able to ‘separate them out’, and in separating them, thus be able to interrogate them.

Diane gave us a day-to-day example of this. When we are driving a car and see a traffic light showing orange, we automatically press the brake pedal before we think consciously about it. If there was a blue light showing at a traffic light, we might think ‘That’s unusual, what is it?’ But we would be less likely unconsciously to brake.

Cognitive Closure
Diane encouraged us to beware of ‘cognitive closure’. That is where we say to ourselves: ‘OK I understand that’. Cognitive closure is an aversion toward ambiguity and points to our desire to arrive at definite conclusions (sometimes irrationally). The need for closure introduces a bias into the judgment process. In other words when you are hell bent on being the one with all the answers, you are less likely to be open to the fact that you may be wrong.

Motivation and Changing Habits
We know that neuroplasticity tells us that we need to ‘use it or lose it’. Neural pathways form and die through use or disuse. Learning a new skill or behaviour, which involves forming new neural pathways and connections, requires open-minded focus and quite some effort. Conscious behaviour can become a low-effort habit if linked to reward over time. Old habits may have significant reward history and new habits that we are trying to learn need to compete with and overcome that.

Autonomy is a reward in itself. Choosing to adopt a new behaviour is more motivational than being told to change. It is also important to identify why a behaviour either is or is not valued. If a goal is linked to a person’s self image or self identity then it will have a higher value and therefore, a higher chance of being achieved. Diane said that there has to be a positive motivational reason for change. The question ‘Why do I want to do this?’ needs to be answered. Unless there is a strong reason for the change we won’t do it. If the change can be made part of one’s own identity, then we are more likely to achieve the change. Therefore the statement ‘I am a runner’ has more likelihood of producing the behaviour of going for a run than considering, ‘Shall I go for a run today or not?’.

Diane gave as a further example in the recent referendum on leaving the European Union. We were asked regarding Brexit, ‘Are you a remainer or a leaver?’ This spoke directly to people’s identity. The question was not ‘What are you going to vote? Leave or remain?’, but instead implied the deeper question, ‘What is your identity, do you identify as a leaver or as a remainer?’. Asking the question in this way asks us about our identity. A question posed in this way is therefore a threat to how we perceive our identity. When we feel threatened via this kind of social threat, this causes us to focus on avoidance...
of that threat. When social threat is increased, we are less adaptable, flexible and creative.

**Emotional Triggers**

Diane made the point that, given what we know about triggers, responses and habits, addressing behaviours that are triggered by one particular environment will not necessarily translate to changing the behaviours in a different environment. For example, if we were to feel anxious about giving presentations and we worked to reduce the anxiety that may be triggered when we were about to give a presentation then we shouldn’t expect that that work would also reduce our anxiety levels when we are in a queue at an airport waiting to board a train. The emotion may be the same (panic) leading to certain behaviours (racing heart, sweaty palms, panic attack) but the triggers would be different (about to give a presentation versus about to board a flight). Change happens when we train our brain to re-associate those feelings. For example, if I am starting a presentation I might think I feel panic, but if I can learn to re-associate that with excitement I can begin to create a new neural pathway. Thus we begin the process of change and experiencing life differently.

Neuroplasticity and habit formation answer some of the ‘so what?’ questions regarding change. If I make a decision to choose to change my behaviour, in my brain I will already start to be rewarded for that decision as that choice to change demonstrates autonomy and autonomy in itself is a reward. If I merely comply with an instruction from someone else then that creates less reward. As enticing as all this sounds, Diane ended by saying that with regard to neuroscience, coaching and psychometrics we should proceed with caution!

---

**Mental toughness – assessing how we think: applications in developing performance, wellbeing, aspirations and positive behaviours**

**PRESENTER: DOUG STRYCHARCZYK, CEO FOR AQR INTERNATIONAL**  
**REVIEWER: VICTORIA HALL OF TALENT FUTURES**

Doug Strycharczyk is happy to admit he is not a psychologist but he became enamoured with assessment, when working as the Head of HR at Castrol Oil. Seizing the opportunity to take a two-week course with Peter Saville to qualify him for his Level A and B, he became hooked on the power of psychometrics. Years later he hired and eventually partnered with Peter Clough of Hull University to develop the MTQPlus, a questionnaire to measure mental toughness. First published in 2002, it is designed to show how we think when responding to opportunity and challenge. With 25% of variation in performance determined by mental toughness, the MTQPlus aims to answer the question ‘Why we do what we do?’
“Mental Toughness is a personality trait which determines, in some part, how individuals perform when exposed to stressors, pressure and challenge... irrespective of the prevailing situation.” (Clough & Strycharczyk, 2011)

Put more simply, mental toughness is the ability to perform even when under pressure. Most of us in the room had taken the MTQPlus in preparation for the session, as it had been offered to session attendees, and we were interested to understand how it had been developed, what validity and reliability studies had been undertaken, and how best to interpret the results. The presentation focused more on the dimensions themselves and the work of the presenter when using the scale but the reliability and validity data are available in documentation from AQR.

Mentally tough individuals tend to be:
- Sociable and comfortable dealing with all types of people;
- Able to remain calm and relaxed in most circumstances - they are competitive or goal orientated in many situations and have lower anxiety levels than others.
- With a high sense of self-belief and an unshakeable faith that they control their own destiny, these individuals can remain relatively unaffected by competition or adversity.
- They can be enthusiastic about change and change even when the challenge is daunting. An individual with a low level of mental toughness is described as mentally sensitive (not as mentally weak).

There are 4 components of Mental Toughness: Control, Commitment, Challenge and Confidence as illustrated in the diagram shown below.

**Control**
Individuals who score high on the Control scale feel that they are in control of their work and of the environment in which they work. They are capable of exerting more influence on their working environment and are more confident about working in complex or multi-tasked situations. This means for example that, at one end of the scale, individuals are able to handle lots of things at the same time and at the other end, they may only be comfortable handling one thing at a time. Ongoing development of MTQ48 has enabled the identification of 2 subscales in the Control scale:
- **Control (Emotion)** – Individuals scoring highly on this scale are better
able to control their emotions and will manage what they show to others. They are able to keep their anxieties in check and are less likely to reveal their emotional state to other people.

- **Control (Life)** – Individuals scoring higher on this scale are more likely to believe that they have a significant degree of control over their lives. They feel that their plans will not be thwarted and that they can make a difference.

**Commitment**

Sometimes described as ‘stickability’, Commitment is the ability for an individual to carry out tasks successfully despite any problems or obstacles that arise whilst achieving the goal. Consequently, an individual who scores at the high end of the scale will be able to handle and achieve things to tough unyielding deadlines whereas, an individual at the lower end will need to be free from those kind of demands in order to achieve their goals.

There are two components to this scale:

- **Goal or target orientation.**
  Individuals scoring high on this scale appear to translate what they need to do into specific goals and targets which enable them to prioritise, plan and monitor several tasks at the same time.

- **Delivery (Completion).** Individuals scoring high on this scale appear to be prepared to do what it takes to deliver what has been promised (to themselves and to others).

**Challenge (or Change Orientation)**

Describes the extent to which individuals see change, setbacks and challenges as opportunities rather than as threats. Individuals who see them as opportunities will actively seek them out and will identify problems as ways for self-development. So for example, at one end of the scale we find those who thrive in continually changing environments and at the other end we find those who prefer to minimise their exposure to change and will strongly prefer to work in stable environments. The two components to this scale:

- **Preparedness to stretch oneself and push back boundaries.** This includes being prepared to take risks and seek out new experiences and challenges.

  In some cases, it will include creating those opportunities.

- **Openness to learning.** Being prepared to see all outcomes as learning opportunities – whatever the outcome, good or bad. Includes being disposed to repeat an experience even it was originally a failure, in order to apply what has been learned.

**Confidence**

Individuals who are high in confidence have the self-belief to successfully complete tasks, which may be considered too difficult by individuals with similar abilities but with lower confidence. Less confident individuals are also likely to be less persistent and may make more errors. For example, individuals at one end of the scale will be able to take setbacks (externally and self-generated) in their stride. They keep their heads when things go wrong and it may even strengthen their resolve to keep going. At the other end individuals will be unsettled by setbacks and will feel undermined by these. The 2 subscales to this scale:

- **Confidence (Abilities)** – Individuals scoring highly on this scale are more likely to believe that they are a truly worthwhile person. They are less dependent on external validation and tend to be more optimistic about life in general.

- **Confidence (Interpersonal)** – Individuals scoring highly on this scale tend to be more assertive. They are less likely to be intimidated in social settings and are more likely to push themselves forward in groups. They are also better able to cope with perceived difficult or awkward behaviours in others.

Doug asserted that Mental Toughness is linked to positive performance. The ‘Mentally Tough’ remain divergent in their thinking, rather than convergent in response to stress – simplifying the complexity and therefore maintaining their creativity.

The ‘Mentally Tough’ remain divergent in their thinking, rather than convergent in response to stress – simplifying the complexity and therefore maintaining their creativity.
Doug does much of his work in schools and he is finding that mental toughness is initially higher in year 7 school children but then plummets in the following years. In reviewing Doug’s 44 slides after the presentation, it seems that much of his focus in creating the MTQPlus is to help people become better adjusted, more resilient, more positive, less anxious, and to learn how to do this through visualisation, goal-setting and emotional control.

These are great lessons to be learned in school and as one who has built positive psychology into her practice, it is hard to argue with this noble endeavour. Clearly Doug has helped a number of people in this regard.

The audience asked many questions about the use and interpretation of the measure as well as the reliability and validity of its construction. A few of the Q&A’s:

Q: Are there any observed gender differences?
A: Doug and Peter have found in their studies that in senior management, mental toughness scores for females are much higher than they are for males.

Q: What are the reliability and validity values for the tool?
A: Reliability > 0.90; Validity > 0.30 – 0.42 but Doug was not able to provide any further detail on the specific type of reliability or validity but assured the audience that there was a wealth of evidence available.

Q: When I took the instrument I found I could jig the outcomes based on how I answered the questions. How do you address this?
A: The items need to be transparent and we don’t have a social desirability scale in there as they don’t work.

Q: Have you correlated scores with job performance?
A: This has been done and a number of 75% of performance due to mental toughness was shared but the details behind the study were not in the presentation material but were available from AQR directly.

Q: Is the measure used in assessment?
A: Yes. It’s short and it’s transparent so it is used alongside other tools but people with high metal toughness are happy with their results and those with low scores tend to feel relieved that they now have an explanation of what has been going wrong.

I had witnessed the psychometrics forum at its liveliest and while I felt uncomfortable for the presenter as he did not have materials which answered all of our questions, I was also grateful for being a part of the Psychometrics Forum, where colleagues can come together to be curious, to explore, to challenge, to debate. We never know what the outcome will be in our sessions, sometimes we pick up a psychometric that opens a new lens of understanding in our practice and we end up using it for years to come. Other times we hear from insightful individuals about their life’s work. And although this presentation may not have had all of the answers to the questions we asked as an audience, I can think of no better way to stay connected to my profession and my colleagues.

POST-EVENT INFORMATION
SHARE FROM PRESENTER
Additional Information on validity was shared by Doug Strychnarczyk after the event. You can also access the technical manual for the MTQ48 at the URL below which provides more detail about the measure and associated research.

https://tinyurl.com/y7zkufva

Mental Toughness Questionnaires Reliability
There are two forms of the Mental Toughness Questionnaire – the MTQPlus and the MTQ48. The MTQ48 is the original form developed in 2002 and sometimes called the 6-factor model. The MTQPlus was developed in 2017 from the MTQ48. MTQPlus is an 8 factor model which embraces the 6 factor model but adds 2 factors for two of the original factors – Commitment and Challenge. It has also developed the original Emotional Control factor.

Reliability
As items purport to measure the same underlying characteristic (e.g. interpersonal confidence), we should expect that if one answers with a mid-range score (3 out of
to one question from this subscale, they will likely answer another question from the same scale with similar mid-range score. The two most common assessments of internal consistency are Cronbach’s alpha ($\alpha$) and McDonald’s Omega ($\psi$). Scores on these metrics have a maximum of 1.00. That would be if there was a perfect correlation between every question within a subscale across a large group of participants.

Typically, scores greater than 0.70 are considered good but if scores are above 0.90, this might indicate that the questions are too closely related, meaning that some questions are redundant and serve only to unnecessarily lengthen the questionnaire or that the construct being measured is very narrow and therefore of less applied use.

The table above shows the most up to date analyses for both formats. Cronbach’s alpha assumes tau-equivalence. That is that each of the items in the scale equally make up the scale score. McDonald’s omega assumes congeneric models. That is, structural equation models whereby the latent variable (subscale score) does not have tau-equivalence (items contribute differently based on the amount of variance that they account for in the overall score).

**Comparative fit index**

The comparative fit index (CFI) analyzes the model fit by examining the discrepancy between the data and the hypothesized model, while adjusting for the issues of sample size inherent in the chi-squared test of model fit, and the normed fit index. CFI values range from 0 to 1, with larger values indicating better fit. A CFI value of .90 or larger is considered to indicate acceptable model fit.

This is not strictly an estimate of reliability but more a measure of model fit – so it is more about validity. This is derived from factor analysis or structural equation modelling. The end goal is to get a measure’s whole scale model fit with CFI > .90 and individual subscale reliabilities.

**MTQ48 reliabilities (2014)**

This study aimed to formally re-assess the factorial validity of the MTQ-48 using the MTQ-48 in a varied sample. 8207 participants (male=4019, female=3922)
and unspecified = 266, mean age = 37.00, SD = 12.09);

Overall sample subjected to CFA as a single factor, 4-factor and 6-factor model. ESEM analysis supported the 6-factor solution as the best, representing good model fit ($\chi^2$ (855) = 8269.1, CFI = .924, TLI = .924, SRMR = .019, RMSEA = .033, 90% CI [.032, .033]).

All latent factors significantly correlated to each other ($r = .61 -.90$, $p < .01$), with interrelationships between factors as this, it is evident that the less constrained ESEM provides a more appropriate method for assessing the model fit in this instance.

Overall, CFA loadings supported the 6-factor model, with loadings largely good to excellent on all subscales. Only emotional control demonstrated weaker loadings. Examining reliability, 5 of the 6-factors had acceptable internal consistency ($\alpha = .78 -.85$). Emotional control was less consistent ($\alpha = .65$). Composite reliability data was similar with 5 of the 6 factors having acceptable reliability (.71 - .80) and emotional control (.59). All samples revealed reasonably consistent results from the CFA and ESEM. Overall, the large-scale analysis of the structure of the MTQ48 supported its validity and it is recommended as a valid and reliable tool for future psychometric assessment.

### Construct validity

The most commonly research aspect is **factorial validity**, which uses sophisticated statistical methods such as confirmatory factor analysis or exploratory structural equation modelling. This determines the extent to which a statistical model is consistent with real data. For example, we propose a statistical model of saying that our interpersonal confidence subscale is determined by several items but that these items are positively associated, but not significant indicators of other subscales within our overall measure.

We then gather large data sets and test the congruence between our statistical model and our data.

There are various indices for how well this fits. The comparative fit index (CFI) is a commonly reported one and indicates the extent to which our data fits the model better than a null model, whereby it would fit it at all well. Therefore, a CFI > .90 indicates that our data fits our model much better than if it were invalid.

---

### Table 2: CFA and ESEM for 6-factor model with various samples

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\chi^2$</th>
<th>CFA CFI</th>
<th>CFA TLI</th>
<th>CFA SRMR</th>
<th>CFA RMSEA</th>
<th>ESEM CFI</th>
<th>ESEM TLI</th>
<th>ESEM SRMR</th>
<th>ESEM RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior management</td>
<td>11156.1</td>
<td>5121.3</td>
<td>.857</td>
<td>.939</td>
<td>.920</td>
<td>.044</td>
<td>.021</td>
<td>.046</td>
<td>.034</td>
</tr>
<tr>
<td>Lower and middle management</td>
<td>9463.9</td>
<td>2154.8</td>
<td>.840</td>
<td>.932</td>
<td>.830</td>
<td>.047</td>
<td>.023</td>
<td>.045</td>
<td>.032</td>
</tr>
<tr>
<td>Clerical and administrative</td>
<td>3451.4</td>
<td>1845.3</td>
<td>.823</td>
<td>.927</td>
<td>.812</td>
<td>.050</td>
<td>.026</td>
<td>.047</td>
<td>.034</td>
</tr>
<tr>
<td>Athletes</td>
<td>3133.1</td>
<td>1679.8</td>
<td>.779</td>
<td>.912</td>
<td>.766</td>
<td>.059</td>
<td>.029</td>
<td>.057</td>
<td>.040</td>
</tr>
<tr>
<td>Students</td>
<td>3832.9</td>
<td>1811.2</td>
<td>.827</td>
<td>.940</td>
<td>.816</td>
<td>.057</td>
<td>.024</td>
<td>.051</td>
<td>.034</td>
</tr>
<tr>
<td>Overall</td>
<td>20454.1</td>
<td>8508.0</td>
<td>.853</td>
<td>.942</td>
<td>.845</td>
<td>.045</td>
<td>.019</td>
<td>.046</td>
<td>.033</td>
</tr>
</tbody>
</table>

Note: $\chi^2$ = chi-square, CFI = comparative fit index, TLI = Tucker-Lewis index, SRMR = standardized root-mean-square residual, RMSEA = root mean square error of approximation.

CFA = confirmatory factor analysis, ESEM = exploratory structural equation modeling. Degrees of freedom (df) for CFA models = 1065. For ESEM model, df = 855.
**Criterion validity**

In this type of validity, our scale must predict observable behaviour. We set the behaviour as our criterion and if our theory is correct, our scale will predict the criterion. For example, pain tolerance is a useful criterion. We know that those who are more mentally tough perform better against this criterion as we would predict.

Criterion Validity is scored on a scale of 1 – 1.00. It is generally accepted that scores in excess of 0.25 are deemed acceptable or good. All scales on MTQ48 score in the region 0.30 – 0.42.

---

### Tales from the Frontier 2017

**REVIEW BY DR HUGH MCCREDIE, CHARTERED INDEPENDENT PSYCHOLOGIST**

As announced in the previous edition of Psyche, this was my tenth and final New Frontiers event, at least as chair of the proceedings.

I billed myself as the opening presenter in order to explore the relations between normal, dysfunctional, disordered and pathological measures of personality. This was intended to give participants some grounding in personality disorders and dysfunctions, preliminary to the presentation of Dr Rainer Kurz of Cubiks, speaking mainly from his personal researches.

Rainer’s theme was ‘Person × Environment = Behaviour (Adaptive & Maladaptive)’. He plotted the trajectory of cognitive ability scores of a female emerging from an abusive history. As a child of seven, she achieved a WISC Total Scale IQ at the 97th percentile; this was largely upheld by the WAIS Verbal Comprehension scale at the age of 25 but, remarkably, Working Memory was at the 14th percentile. This paradoxical profile was sustained at age 30. In the remainder of his presentation Rainer reviewed the single, dual, Big Five and 10 Aspects models of personality and how these related to measures of normal, dysfunctional and pathological functioning. Amongst the measures reviewed were Cubik’s own Personality and Preferences Inventory (PAPI).

Dr Nigel Guenole, Goldsmiths, University of London, entitled his presentation ‘Psychometrics and Human Capital: It’s time for a broader focus’. His central proposition was that organisations should focus on the stock of aggregate knowledge, skills, abilities, and other attributes (KSAs) and how these predict team, business-unit, and corporate outcomes. Personality traits fall into the domain of ‘other attributes’. Confirmatory factor analysis suggests that whilst the Big Five Factors may differentiate individual performance, team performance might be better predicted by two broader clusters (1) low Agreeableness plus Openness and (2) Extraversion plus Conscientiousness plus low Neuroticism.

My second, very brief, presentation was entitled ‘Relating implicit and explicit personality measures’. It reported how Conditional Reasoning (CR) based on a ‘cover’ task (like the created ‘dilemmas’ used in Situational Judgement Tests or SJT) can reveal reasoning biases reflecting implicit motives. It showed how self-reported traits channel the ways
in which such deep-seated motives are expressed in behaviour and life outcomes.

James Bywater of Korn Ferry, presented on ‘Values based recruitment: Considerations and consequences’. He postulated that effective performance was likely to stem from a combination of (1) competencies and (2) the alignment of values with those of the organisation. In situations where competencies were readily available (e.g. with candidates for care work), a measure of values could be a cost-effective means of selecting the most suitable. James outlined the Talent Q Values SJT and reported that the data yielded small (circa 0.25), but significant, correlations with care worker performance. The Talent Q values reflect enduring beliefs, personal objectives, preferences and behaviours.
Automating the Generation of 3D Rotational Items and Modelling Item Difficulty using Item Response Theory

1. Automatic Item Generation (AIG)
   - The widespread use of smartphones and the Internet is making it easier for test-takers to compromise test security and share tests with other prospective test-takers.
   - The process of item development is expensive and time consuming (Rudner, 2010). Typically, about 40% of the items do not pass the empirical evaluation stage (Haladyna 2012).
   - Advances in computer processing power opens up the use of computer-adaptive testing more readily in real testing situations. With AIG techniques, it enables the possibility of producing large item banks with predictable psychometric properties in a cost-effective and efficient manner.
   - The purpose of this research is twofold. First, to generate 3D rotational items using a generator. Second, to evaluate the psychometric quality of these items using the IRT (Rasch) model.

2. Data Collection and Procedure
   - 21 items were automatically generated using the AIG R package (Loe, 2017).
   - Respondents have to identify the correct figure by rotating each of the 5 available response figures until they find the one that matches the display figure.
   - The test was distributed online via the psychometrics centre testing website. 85% of the test-takers were from the UK / US.
   - 101 participants completed the test and 98 provided demographics information. 3 did not respond.
   - 70 males and 28 females completed the test.
   - The average age was 32.4 years old (SD=12.2), ranging from 18 to 69 years.

3. Data Analysis
   - All analyses were conducted using the R programming language.
   - Item Response Theory was used to mathematically model the responses as a function of the latent ability (Embreton & Reise, 2013).
   - Specifically, the Rasch model was employed to estimate the item difficulty parameter. Under the Rasch model, the discrimination parameter is fixed to 1.
   - The Wald test (median split) was used to evaluate item misfit (Glas & Verhelst, 1995).
   - The internal consistency of the test was measured using the Person Separation Reliability index. Values greater than 0.7 indicate adequate reliability (Andrich, 1982).

4. Results
   - In the Wald test, item 10 and item 12 were identified as misfits and were subsequently removed.
   - The item difficulty estimates for the remaining 19 items ranged from -1.50 (item 7) to 1.37 (item 5).
   - The test information was highest at 4.04 around the average ability, with a downward trend towards both ends of the continuum.
   - The Person Separation Reliability index was 0.81, well above the cut-off criteria.
   - The items mapped onto the ability continuum adequately.

5. Discussion and Research Contributions
   - While the creation of the generator is still in its early stages, this research provides evidence that it is possible to automatically generate large quantities of items efficiently.
   - Furthermore, the item properties adhered to rigorous psychometric models using Item Response Theory.
   - Follow up research should hypothesise specific cognitive rules that influence the difficulty of 3D rotational items.
   - Researchers are now one step closer to producing pre-calibrated items that need not be piloted. This will reduce item exposure rate and increase test security for recruitment and selection purposes.

Key References