

## From Shakespeare to Gogglebox – Is Psychometrics making progress?

Summary of a talk by Dr Roy Childs to  
the Association of Business Psychologists 15th September 2020

This talk challenges some of the basic assumptions made by mainstream psychometrics when applied to self-report questionnaires. Some key questions that it addresses are:

1. what is the real nature of the data?
2. are traditional methods of analysis either suitable or meaningful?
3. are reliability and validity being evaluated in the right way?
4. what is, or should be, the real purpose of self-report questionnaires?

Having worked as both a researcher and a practitioner as well as hosting seminars with Ray Cattell, running workshops with Will Schutz and starting his career in this field with Peter Saville, Roy claims to view this field from multiple perspectives. This has led him to recognise the great value of self-report psychometrics – but only if their nature is better understood so that their use is adapted to become more ‘fit for purpose’ and to be used within a more clearly defined ethical framework. This is essential if we are to avoid the ‘Google/Facebook’ syndrome of ‘sleep-walking into the internet’ with the current move towards AI to assess people.

Roy began his talk by inviting participants to self-identify as advocates, cynics or explorers – and the majority claimed to be either advocates and/or explorers. He then anticipated himself being cast as a cynic given his fundamental criticisms of the field and so was at pains to explain that well-constructed self-report questionnaires used well can add considerable value. However, his warning was that they were often misused, over-claimed and misunderstood.

He then explained why he had chosen the title of the talk. If watching Shakespeare is viewed as a pinnacle in literature and performance, then 400 years later, watching Gogglebox does not exemplify real progress. The parallel with psychometrics is whether the current direction in psychometrics represents progress or regression. In order to explore this issue, he suggested dividing it into questioning progress in terms of the methods of input, of output and with what is happening in the ‘black box’ in-between.

**Input Progress.** Roy was clear that there are clear advantages (and hence progress) in the collection of data digitally. These include the ease of collecting data (and hence analysing) as well as the increased variety of input stimuli (pictures, animations etc.). However, Roy questioned whether progress was being made in terms of **what** we should be measuring rather than how we measure it. Quoting Ray Cattell who said that the 'Big Five' were the factors that survived clumsy factor analysis, Roy made the point that the search for the basic 'ingredients' of personality had faltered on the altar of the Big Five. The curiosity to investigate and potentially discover a 'Periodic Table' of personality had been replaced by with lists of attractive labels that match workplace competencies that appeal to and which represent high-level labels (such as persuasiveness). Since almost any comprehensive set of such labels will produce a Big Five structure curiosity seems to have ended. The trend has been towards questionnaires with scales representing complex, multi-factorial constructs that appeal to practitioners without getting underneath the surface. He also suggested there was a lack of clear rationales for the choice of items which often had an unexplained mix of questions involving personal preferences versus behaviours versus attitudes versus motivations.

**Output Progress.** Once again, he acknowledged areas of clear progress. The speed of creating attractive reports and the emergence of more interactive styles of report are both to be welcomed. However, nice presentation with many pages of content do not necessarily equate to quality. Sometimes such reports look good but are overloaded with superfluous information designed for some general context which may or not be relevant or even accurate? Human judgement is being replaced by the 'expert in the machine' in many fields but again, this is not always progress and it can intimidate practitioners from challenging the algorithms from which they are derived.

**In between.** Roy also questioned whether the process between input and output – the analyses and algorithms – represented real progress. More and more data analysed with ever more sophisticated statistics is not always justified or done well. Similarly, he questioned whether our models of reliability and validity are really suitable for the nature of the data we are collecting? And simply because we have huge data sets, does this really justify producing 'Global Norms' – does mixing data from diverse backgrounds through multiple translations really provide meaningful comparisons or simply a meaningless average?

At the heart of Roy's criticisms is the misunderstanding and misuse of self-report numbers. To explain he made the distinction between soft and hard numbers using a truly remarkable example in which Will Schultz was invited by NASA to run a series of workshops to develop interpersonal "openness". Before attending the workshops, participants completed a survey which summarised their perceptions of the current level of 'openness' in their organisation. After a year of workshops participants were invited to re-take the survey. The shock was how the result was a drop in scores – the survey suggested they had become less open which was the opposite of the workshop objectives. However, more qualitative assessment revealed that this was not because participants had become more closed but rather that the concept of openness had become more widely understood. Participants realised that what they understood as openness prior to the workshops was a long way from where they wanted to be. The lower scores did, in fact, represent a move towards being more open but they were now using a different benchmark. The parallel in psychometrics is that, in assigning numbers to complex and personally defined constructs, we can never be sure that a rating of 4 is actually less than a rating of 5.

Roy then proceeded to illustrate four dilemmas underscoring the nature and accuracy of self-reports. These were issues to do with the nature of **language**, the ignoring of **context**, the concept of **change** and the disregard for aspects of the **scientific method**.

**Language.** The ambiguity that underlies the use of labels for high level constructs was illustrated using the example of 'I like change.' Most people will interpret that question in very personal (and different) terms. Nevertheless, statistical analyses can (and does) produce reasonable statistics using the traditional psychometric model which highlights how the statistics are an insufficient criterion for judging quality. Given the inherent ambiguity in language Roy suggested considering answers in self-report questionnaires as 'puzzles to be understood rather than answers to be interpreted'. This reflects a profound change in our expectations and hence usage of self-report questionnaires. Instead of thinking of them, as 'objective' measure designed to predict behaviour, they can be viewed as providing people with an opportunity for people to describe their experience in a structured and coherent way. In fact, at a fundamental level, the essence of the human condition is how we experience our lives and more emphasis should be placed on this as the window into a person's identity (note: Roy uses identity as the core concept that all the personality, values, role and interests questionnaires are illuminating). Hence psychometric questionnaires give people the opportunity to better understand themselves. They are the kaleidoscopes that produce a pattern from multiple pieces (the answers to many questions). They act as microphones that enable people to listen to themselves and to be heard by others.

To assist with addressing these issues, Team Focus has developed questionnaires that provide more varied ways of helping people to express themselves and to produce a more accurate picture. These include:

1. The Team Focus Resilience Scales Questionnaire (RSQ). This goes beyond asking for a 'Typical' rating on a Likert scale. It allows people to express how they feel 'Stretched' and 'Stressed' as well. This produces 3 scores which explore how people believe they change under different kinds of pressure. No matter where their initial 'Typical' rating was placed, we now have information about how they think they change. The degree and direction of this change provides a richer set of data which leads to more interesting and challenging reports.
2. The Team Focus Values-based Indicator of Motivation (VbIM). This takes a different approach by allowing people to rate and rank their values in such a way that three different rankings are produced. Differences in these rankings can then be used to question and challenge a person's self-narrative. It is interesting how, even from self-report, the potential contradictions can be particularly useful to go behind the obvious. People find it hard to articulate their true values and often fall back on some general 'socially desirable' constructs – until prompted to explore further and deeper.
3. Introducing the 'paired process'. Several of the Team Focus questionnaires (such as the TDI, VbIM, EIQ and RHA) accept the limitations of self-report and introduce feedback from a third party. The 'paired process' is an alternative to the better known 360-degree feedback and Roy claimed that it created a more intimate context within which self-exploration and identifying areas for development could flourish. Furthermore, the paired process changes the nature of the participant's engagement with the process. If well managed, people become more curious and less defensive which are pre-requisites for making good use of developmental feedback.

Clearly all these methods change the paradigm of using questionnaires from one which is viewed as 'external and objective' and hence '**done to**' people to one where we are using questionnaires **with** people. The first paradigm has been dominant and the increasing use of 'big data' from people's internet footprints follows on naturally. However, we are yet to properly address the ethical issues. Perhaps adopting the second paradigm would be not only more easily justified ethically but also more realistic?

**Context.** Professor Jerome Kagan in his book 'Psychology's Ghosts' identified a 'black hole' in psychology in his very first chapter which is called 'Missing Contexts'. Unfortunately, psychometrics has also fallen into this black hole. We all know that people have the ability to adapt their behaviour to circumstances – some better than others. However, all of us are limited in what we do by our history and context. Observing what we do only demonstrates what I have become rather than who I could become. Linked to the concept of flexibility and change Roy expressed concern about how most psychometric questionnaires ignored context but then intermingles it into people's answers by using phrases such as 'typically' or 'in the average of situations'. This is of critical importance, especially for those using psychometrics in a development context – and it also brings into question what we mean by personality. Is the psychometric concept of personality assuming or defining only those elements of people's character can't (or don't) change? What if we assume that some aspects of personality can change? This would challenge a core element of psychometric methodology – the evaluation of reliability and validity. Roy added the complication of how context and personality interact to produce behaviour which then challenged the concept of a 'work personality' which, by definition, is contextually defined? He argued that there had been a lack of rigorous thinking as exemplified by the OPQ which asks for work role behaviour but calls itself a personality questionnaire and by Belbin's Team Role questionnaire which claims to assess a preferred role but which assesses a much broader concept of what people do in a wide range of contexts and teams. To help disentangle these confused ideas Roy suggests using his '4-selves model' which suggests that we can be answering questionnaires from different parts of our identity – namely the 'Contextual Selves', the 'Identity Self' and the 'Ideal Self' plus remaining open to the fact that there are parts of ourselves that have not yet emerged or been tested which he called the 'Undiscovered Self'.

Given the potency of the 'Ideal Self' in a person's psyche it is, perhaps, surprising that psychometrics has not ventured into that terrain (with the exception of FIRO and the TDI). Roy advocated widening our current view of how people should answer questionnaires in order to map people's contextual, identity and ideal selves separately which not only provides a more realistic picture of why and how a person engages with the world around them but also acknowledges the way people can adapt and be flexible – at least to some degree.

**Change.** Without getting embroiled in the nature/nurture debate, Roy emphasised how the assumption that personality doesn't change becomes a self-fulfilling prophecy. Psychometric orthodoxy treats any change is called unreliability which drives out items and measures that are sensitive to change. However, by reframing self-report questionnaires as a way to help people articulate their perception and experience of themselves, we open the door to change being acceptable and reflecting either real change or, perhaps, a person's reinterpretation of their own perceptions. This is much closer to peoples lived experience and questionnaires become vehicles to help people develop their own narrative about themselves.

**Science.** Finally, the claim that psychometrics is the best example in psychology of science being operationalised needs challenging. It is based on the fact that psychometrics collects numbers. These numbers become the evidence that can be subjected to the scientific analysis. However, when we acknowledge that these are 'soft' numbers (as described earlier), current data analyses become more suspect. Alongside this we need to challenge the apparent 'gold standard' of psychometrics which is predictive validity. This is not as robust as many claim since:

1. there is the assumption that the predictor measurement and criterion are static. However, the reality is that both the predictor and criterion are dynamic – moveable and changing at least to some degree.
2. test manuals present correlations showing that scale scores predict certain criteria. Aside from the criterion problem (that criteria are hard to define and measure reliably) the approach taken in test manuals is to report correlations that demonstrate a level of prediction (statistically significant but often psychologically less meaningful). Then we remember that a basic scientific approach (as articulated by Karl Popper) is that science tries to **disprove** a theory. Very little is written about trying to disprove questionnaires.
3. Validation usually consists of identifying general trends in large data sets. However, this approach, whilst useful, aims to make broad generalisations and hence treats outliers as aberrations. However, as Todd Rose in his book 'The End of Average' demonstrates, this can sometimes obscure really important differences – especially when applied to individuals which is the most common use of self-report questionnaires.

In conclusion psychometrics is suffering from an historical legacy that has pervaded psychology. When we look at the contributions of such huge influences as Freud and Skinner, we see that, in spite of their very different perspectives, both took a reductive approach. Both look backwards to explain or justify the present. However, if we accept that a good part of what drives people involves the search for meaning, we change our focus to understanding a person's purpose which, in turn, makes us look forwards. What becomes important is what people are becoming rather than how they became. This is a more dynamic view of the human condition and was strongly advocated by Jung who, contrary to popular thinking, used his psychological type theory as a compass for what people are becoming rather than as the somewhat static model of what I cannot change.

Roy ended with a summary of the things he had covered as follows:

1. Numbers from self-report are 'soft' not 'hard'.
2. Words used have many different meanings for different people
3. Behaviour is not always an insight into the inner person
4. We need context to make sense of what people say and do
5. Some things may not change – but we are dealing with people's 'story' about themselves which can and does change
6. Used well, psychometrics provide stimulus for exploration.